**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2022 Т. 14 № 2 С. 473–495**
DOI: 10.20537/2076-7633-2022-14-2-473-495

Ки&М

МАТЕМАТИЧЕСКИЕ ОСНОВЫ И ЧИСЛЕННЫЕ МЕТОДЫ МОДЕЛИРОВАНИЯ

УДК: 519.85

Субградиентные методы для задач негладкой оптимизации с некоторой релаксацией условия острого минимума

С. С. Аблаев[1,2,a], Д. В. Макаренко[1,b], Ф. С. Стонякин[1,2,c], М. С. Алкуса[1,3,d], И. В. Баран[2,e]

[1]Московский физико-технический институт,
Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., 9
[2]Крымский федеральный университет им. В. И. Вернадского,
Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, 4
[3]Национальный исследовательский университет «Высшая школа экономики»,
Россия, 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: [a] seydamet.ablaev@yandex.ru, [b] devjiu@gmail.com, [c] fedyor@mail.ru,
[d] mohammad.alkousa@phystech.edu, [e] matemain@mail.ru

*Получено 12.02.2022, после доработки — 30.03.2022.
Принято к публикации 05.04.2022.*

Задачи негладкой оптимизации нередко возникают во многих приложениях. Вопросы разработки эффективных вычислительных процедур для негладких задач в пространствах больших размерностей весьма актуальны. В таких случаях разумно применять методы первого порядка (субградиентные методы), однако в достаточно общих ситуациях они приводят к невысоким скоростным гарантиям. Одним из подходов к этой проблеме может являться выделение подкласса негладких задач, допускающих относительно оптимистичные результаты о скорости сходимости в пространствах больших размерностей. К примеру, одним из вариантов дополнительных предположений может послужить условие острого минимума, предложенное в конце 1960-х годов Б. Т. Поляком. В случае доступности информации о минимальном значении функции для липшицевых задач с острым минимумом известен субградиентный метод с шагом Б. Т. Поляка, который гарантирует линейную скорость сходимости по аргументу. Такой подход позволил покрыть ряд важных прикладных задач (например, задача проектирования точки на выпуклый компакт или задача отыскания общей точки системы выпуклых множеств). Однако как условие доступности минимального значения функции, так и само условие острого минимума выглядят довольно ограничительными. В этой связи в настоящей работе предлагается обобщенное условие острого минимума, аналогичное известному понятию неточного оракула. Предложенный подход позволяет расширить класс применимости субградиентных методов с шагом Б. Т. Поляка на ситуации неточной информации о значении минимума, а также неизвестной константы Липшица целевой функции. Более того, использование в теоретической оценке качества выдаваемого методом решения локальных аналогов глобальных характеристик целевой функции позволяет применять результаты такого типа и к более широким классам задач. Показана возможность применения предложенного подхода к сильно выпуклым негладким задачам и выполнено экспериментальное сравнение с известным оптимальным субградиентным методом на таком классе задач. Более того, получены результаты о применимости предложенной методики для некоторых типов задач с релаксациями выпуклости: недавно предложенное понятие слабой β-квазивыпуклости и обычной квазивыпуклости. Исследовано обобщение описанной методики на ситуацию с предположением о доступности на итерациях δ-субградиента целевой функции вместо обычного субградиента. Для одного из рассмотренных методов найдены условия, при которых на практике можно отказаться от проектирования итеративной последовательности на допустимое множество поставленной задачи.

Ключевые слова: субградиентный метод, острый минимум, квазивыпуклая функция, слабо β-квазивыпуклая функция, липшицева функция, δ-субградиент

Исследования по обоснованию теорем 5, 6 и 7 выполнены за счет гранта Российского научного фонда и города Москвы № 22-21-20065 (https://rscf.ru/project/22-21-20065/).

COMPUTER RESEARCH AND MODELING
2022 VOL. 14 NO. 2 P. 473–495
DOI: 10.20537/2076-7633-2022-14-2-473-495

MATHEMATICAL MODELING AND NUMERICAL SIMULATION

UDC: 519.85

Subgradient methods for non-smooth optimization problems with some relaxation of sharp minimum

S. S. Ablaev[1,2,a], D. V. Makarenko[1,b], F. S. Stonyakin[1,2,c], M. S. Alkousa[1,3,d], I. V. Baran[2,e]

[1]Moscow Institute of Physics and Technology,
9 Institutskiy per., Dolgoprudny, Moscow region, 141701, Russia
[2]V. I. Vernadsky Crimean Federal University,
4 Academician Vernadsky Avenue, Simferopol, Republic of Crimea, 295007, Russia
[3]HSE University,
20 Myasnitskaya st., Moscow, 101000, Russia

E-mail: [a] seydamet.ablaev@yandex.ru, [b] devjiu@gmail.com, [c] fedyor@mail.ru,
[d] mohammad.alkousa@phystech.edu, [e] matemain@mail.ru

Received 12.02.2022, after completion — 30.03.2022.
Accepted for publication 05.04.2022.

Non-smooth optimization often arises in many applied problems. The issues of developing efficient computational procedures for such problems in high-dimensional spaces are very topical. First-order methods (subgradient methods) are well applicable here, but in fairly general situations they lead to low speed guarantees for large-scale problems. One of the approaches to this type of problem can be to identify a subclass of non-smooth problems that allow relatively optimistic results on the rate of convergence. For example, one of the options for additional assumptions can be the condition of a sharp minimum, proposed in the late 1960s by B. T. Polyak. In the case of the availability of information about the minimal value of the function for Lipschitz-continuous problems with a sharp minimum, it turned out to be possible to propose a subgradient method with a Polyak step-size, which guarantees a linear rate of convergence in the argument. This approach made it possible to cover a number of important applied problems (for example, the problem of projecting onto a convex compact set). However, both the condition of the availability of the minimal value of the function and the condition of a sharp minimum itself look rather restrictive. In this regard, in this paper, we propose a generalized condition for a sharp minimum, somewhat similar to the inexact oracle proposed recently by Devolder – Glineur – Nesterov. The proposed approach makes it possible to extend the class of applicability of subgradient methods with the Polyak step-size, to the situation of inexact information about the value of the minimum, as well as the unknown Lipschitz constant of the objective function. Moreover, the use of local analogs of the global characteristics of the objective function makes it possible to apply the results of this type to wider classes of problems. We show the possibility of applying the proposed approach to strongly convex non-smooth problems, also, we make an experimental comparison with the known optimal subgradient method for such a class of problems. Moreover, there were obtained some results connected to the applicability of the proposed technique to some types of problems with convexity relaxations: the recently proposed notion of weak β-quasi-convexity and ordinary quasi-convexity. Also in the paper, we study a generalization of the described technique to the situation with the assumption that the δ-subgradient of the objective function is available instead of the usual subgradient. For one of the considered methods, conditions are found under which, in practice, it is possible to escape the projection of the considered iterative sequence onto the feasible set of the problem.

Keywords: subgradient method, sharp minimum, quasi-convex function, weak β-quasi-convex function, Lipschitz-continuous function, δ-subgradient

Citation: *Computer Research and Modeling*, 2022, vol. 14, no. 2, pp. 473–495 (Russian).

This work in Theorems 5, 6, and 7 was supported by Russian Science Foundation and Moscow city (project 22-21-20065, https://rscf.ru/project/22-21-20065/).

1. Введение

Негладкие оптимизационные задачи часто возникают в самых разных приложениях. Однако известные методы субградиентного типа для таких задач приводят к пессимистичным теоретическим оценкам скорости сходимости в пространствах больших размерностей. Одним из подходов к этой проблеме может служить выделение специального класса задач с условием острого минимума [Поляк, 1969; Polyak, 1987]. Говорят, что f удовлетворяет условию острого минимума, если

$$f(x) - f(x_*) \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2 \quad \forall x \in Q \tag{1}$$

для некоторого фиксированного $\alpha > 0$ и $f(x_*) = f^* = \min\limits_{x \in Q} f(x)$ для всякого $x_* \in X_*$, где Q — выпуклое и замкнутое подмножество $\mathbb{R}^n$, X_* — компакт и $\|\cdot\|_2$ — евклидова норма. При таком допущении удается предложить субградиентный метод с гарантией линейной скорости сходимости в случае доступности информации о точном значении f^* [Поляк, 1969] без использования в теоретических оценках скорости сходимости параметра размерности пространства. Условие острого минимума верно, например, для задачи проектирования точки на выпуклый компакт. Однако требование доступности f^* довольно ограничительно. В этой связи в настоящей работе рассмотрено некоторое обобщение условия острого минимума

$$f(x) - \overline{f} \geqslant \alpha \min_{x_* \in X_*} \|x - x_*\|_2 - \Delta, \tag{2}$$

где $\overline{f}$ — это некоторое приближение минимального значения функции f^*, причем $\overline{f} \geqslant f^*$. Оказалось, что такое обобщение позволяет несколько расширить класс применимости субградиентных методов для задач с острым минимумом и шагом Б. Т. Поляка. Например, оно может покрыть постановку задачи с неточной информацией о f^*. При этом в работе выводятся оценки качества выдаваемого решения субградиентным методом с «неточным» аналогом шага Б. Т. Поляка на задачи с неизвестной константой Липшица целевой функции. Такой подход, связанный с использованием в теоретических результатах локальных аналогов глобальных характеристик целевой функции (в данном случае константа Липшица f), позволяет применять полученные результаты и к более широким классам задач с необязательно липшицевыми целевыми функциями.

В настоящей статье показана возможность применения предложенного подхода к сильно выпуклым негладким задачам (вообще говоря, не обладающим острым минимумом) и выполнено экспериментальное сравнение с одним из известных оптимальных субградиентных методов [Поляк, 1969] на таком классе задач. Более того, получены результаты о применимости предложенной методики для некоторых типов задач с релаксациями выпуклости: недавно предложенное условие слабой β-квазивыпуклости [Hardt, Ma, Recht, 2018] или обычное условие квазивыпуклости (унимодальности). Также исследовано обобщение указанной методики на ситуацию с предположением о доступности на итерациях δ-субградиента целевой функции вместо обычного субградиента. Для одного из рассмотренных методов найдены условия, при которых на практике можно отказаться от проектирования на допустимое множество поставленной задачи.

Итак, в работе рассмотрен класс задач

$$f(x) \to \min_{x \in Q}, \tag{3}$$

где f — непрерывная и выпуклая, или квазивыпуклая (унимодальная), или слабо β-квазивыпуклая при некотором $\beta \in (0; 1]$ функция (понятие слабой β-квазивыпуклости введено недавно в [Hardt, Ma, Recht, 2018]). Для задач вида (3) с предположением типа (2) в настоящей работе исследованы некоторые вариации субградиентного метода с шагом Б. Т. Поляка.

Данная статья состоит из введения, заключения и шести основных разделов. В § 2 вводится аналог острого минимума (понятие Δ-острого минимума) и рассматриваются примеры задач, для которых это условие заведомо выполнено (примеры 1–3). При этом может не выполняться условие обычного острого минимума. В § 3 исследован адаптивный субградиентный метод на классе задач слабо β-квазивыпуклых функций, допускающих Δ-острый минимум. В § 4 рассмотрен частично адаптивный субградиентный метод на классе квазивыпуклых функций в случае острого минимума, а также в случае Δ-острого минимума. Получены оценки скорости сходимости рассматриваемых методов (теоремы 1–4), также исследована возможность использования субградиентного метода для квазивыпуклых задач без необходимости проектирования на допустимое множество. Для этого получена оценка удаления траектории субградиентного метода от начальной точки (теорема 5). В § 5 исследуется аналог предложенного в разделе 4 метода на классе выпуклых липшицевых функций с использованием на итерациях δ-субградиентов вместо обычных субградиентов. Отметим, что использование δ-субградиентов потенциально может позволить сэкономить вычислительные затраты метода за счет отказа от требования доступности точного значения субградиента целевой функции в текущей точке. При этом доказано, что в оценке качества выдаваемого решения не накапливаются величины, соответствующие $\delta > 0$ (теорема 7). В § 6 описаны вычислительные эксперименты по демонстрации эффективности и сравнению предложенной в § 3 методики для задач с Δ-острым минимумом с работой известного субградиентного метода [Julien, Schmidt, Bach, 2012] для некоторых сильно выпуклых задач (выбор этого типа задач для сравнения обусловлен тем, что для них известны применимые на практике оценки качества решения по аргументу). Попутно также выведена адаптивная оценка качества выдаваемого субградиентным методом [Julien, Schmidt, Bach, 2012] решения. В § 7 на классе сильно выпуклых задач описана оценка качества решения субградиентного метода [Julien, Schmidt, Bach, 2012] с использованием адаптивно подбираемых параметров и для задачи о наименьшем покрывающем шаре описаны результаты экспериментов по сравнению эффективности предложенных в настоящей работе субградиентных методов с шагом типа Б. Т. Поляка и субградиентного метода типа [Julien, Schmidt, Bach, 2012].

2. О некоторой релаксации понятия острого минимума

В данной работе мы рассмотрим обобщенное условие острого минимума (2), где параметры $\alpha > 0$ и $\Delta > 0$ фиксированы и задано число $\overline{f}$ (в частности, его можно интерпретировать как приближение минимума f^*). Заметим, что условие (2) можно рассматривать и как обобщение острого минимума, в некотором смысле аналогичное неточному оракулу [Devolder, Glineur, Nesterov, 2014]. Приведем несколько примеров.

Пример 1. Начнем с известной задачи о нахождении общей точки системы выпуклых компактов $\{\Omega_i\}_{i\in I=\overline{1,m}}$ в пространстве $\mathbb{R}^n$, которая может иметь смысл нахождения положения какой-нибудь изучаемой системы, которая соответствует ряду заявленных требований. Можно решать такую задачу методом последовательного проектирования, выбирая $x_0 \in \Omega_1$, а далее в качестве $x_i \in \Omega_{i+1}$ при $i = \overline{0, m-1}$ точку, ближайшую к x_i. При этом каждая из m подзадач будет сводиться к задаче минимизации функции вида $f_i(x) = \min\limits_{y\in\Omega_{i+1}} \|x-y\|_2$, которая удовлетворяет условию Липшица ($M = 1$) с острым минимумом ($\alpha = 1$) и начальной точкой $x = x_i$ и $f_i^* = 0$ при условии, что пересечение множеств $\{\Omega_i\}_{i\in I}$ не пусто. В таком случае субградиентный метод

$$x_{k+1} = x_k - h_k \nabla f(x_k)$$

с шагом Б. Т. Поляка [Поляк, 1969]

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

будет для каждой подзадачи выполнять проектирование точки x_i на множество Ω_{i+1} за 1 шаг.

Можно рассмотреть ситуацию, когда поставленная задача неразрешима и пересечение множеств $\{\Omega_i\}_{i\in I}$ пусто, но есть достаточно близкая ко всем множествам (условно подходящая) точка (удаленная не более чем на $\Delta > 0$). Тогда вполне логично рассматривать обобщение задачи об отыскании общей точки системы множеств с целью найти точку, достаточно близкую ко всем множествам. В этом случае можно также применять описанную выше схему. Но оптимальные значения f_i^* в этом случае уже не известны. Тогда можно выбрать $\overline{f}_i \geqslant \Delta$ (на практике можно рассматривать и ситуацию $\overline{f}_i \leqslant \Delta$) и применять уже предлагаемые ниже в настоящей статье вариации субградиентных методов с заменой f_i^* на $\overline{f}_i$ и полученные теоретические результаты с учетом условия Δ-острого минимума. Отметим выполненные эксперименты с реализацией подхода указанного типа (см. замечание 3 ниже).

Приведем еще несколько примеров.

Пример 2. Пусть

$$f(x) = \|x\|_2 + \gamma\|x - c\|_2^2$$

при $x \in \mathbb{R}^n$ для некоторого фиксированного вектора $c \in \mathbb{R}^n$ малой нормы. Тогда $0 < f^* \leqslant f(0) = \gamma\|c\|_2^2$. Ясно, что при $\|x\|_2 > \gamma\|c\|_2^2$ вектор x заведомо не совпадает с x_*, т. е. x_* лежит в шаре с центром в 0 радиусом $\gamma\|c\|_2^2$. В этом случае

$$f(x) \geqslant \|x - x_*\|_2 - \|x_*\|_2 + \gamma\|x - c\|_2^2,$$

и тогда

$$f(x) - f^* \geqslant \|x - x_*\|_2 - \|x_*\|_2 + \gamma\|x - c\|_2^2 - \gamma\|c\|_2^2,$$

т. е.

$$f(x) - f^* \geqslant f(x) - \gamma\|c\|_2^2 \geqslant \|x - x_*\|_2 - \|x_*\|_2 - \gamma\|c\|_2^2 \geqslant \|x - x_*\|_2 - 2\gamma\|c\|_2^2,$$

и можно применять описанную далее методику для $\overline{f} = \gamma\|c\|_2^2$, $\alpha = 1$ и $\Delta = 2\gamma\|c\|_2^2$.

С другой стороны, можно учитывать 2γ-сильную выпуклость и применять к задаче минимизации f, например, субградиентный метод [Julien, Schmidt, Bach, 2012]. Эксперименты для такого примера приведены далее в § 7.

Пример 3. В качестве примера задачи, для которой верно обобщенное условие острого минимума, рассмотрим условие **слабого острого минимума вида**

$$f(x) - f^* \geqslant \mu\|x - x_*\|_2^p \quad \forall x \in Q \tag{4}$$

при $p \in (1; \infty)$. При этом обычное условие острого минимума, вообще говоря, не выполняется.

В частности, при $p = 2$ неравенство (4) верно для 2μ-сильно выпуклых функций.

Если $\|x - x_*\|_2 \geqslant \varepsilon$, то

$$f(x) - f^* \geqslant \mu\varepsilon^{p-1}\|x - x_*\|_2.$$

В противном случае (т. е. $\|x - x_*\|_2 < \varepsilon$) имеем

$$\mu\|x - x_*\|_2^p > \mu\|x - x_*\|_2 - \mu\varepsilon,$$

откуда

$$f(x) - f^* > \mu\varepsilon^{p-1}\|x - x_*\|_2 - \mu\varepsilon,$$

т. е. выполнено условие обобщенного острого минимума (2) при

$$\alpha = \mu\varepsilon^{p-1}, \quad \Delta = \mu\varepsilon, \quad \overline{f} = f^*.$$

Более того, можно рассмотреть случай неточной информации о f^* для функции f, удовлетворяющей (4), т. е. когда известно лишь $\overline{f}$ такое, что $\overline{f} - f^* \leqslant \widetilde{\Delta}$. В этом случае условие обобщенного острого минимума (2) будет верно при

$$\alpha = \mu\varepsilon^{p-1} \quad \text{и} \quad \Delta = \mu\varepsilon + \widetilde{\Delta}.$$

Ясно, что список примеров задач можно продолжить (например, отправляясь от материалов § 3 статьи [Поляк, 1969]).

3. Адаптивный субградиентный метод для слабо β-квазивыпуклых задач с Δ-острым минимумом

Пусть $Q \subseteq \mathbb{R}^n$ — выпуклое замкнутое множество и $f\colon Q \to \mathbb{R}$ — слабо β-квазивыпуклая функция [Hardt, Ma, Recht, 2018] для некоторого фиксированного параметра $\beta \in (0; 1]$. Напомним [Hardt, Ma, Recht, 2018], что f называется *слабо β-квазивыпуклой* относительно точки минимума x_* задачи (3) на множестве Q, если для произвольного $x \in Q$ выполнено неравенство

$$f(x_*) \geqslant f(x) + \frac{1}{\beta}\langle \nabla f(x), x_* - x\rangle, \tag{5}$$

где $\nabla f(x)$ — произвольный субградиент f в точке x. Под субградиентом мы здесь и всюду далее понимаем элемент субдифференциала Кларка f в точке x и предполагаем его существование. Если f дифференцируема в точке x, то под ∇f понимаем обычный градиент. Это вполне естественно для липшицевых функций (для существования субдифференциала Кларка в точке достаточно локальной липшицевости f в окрестности этой точки). Если функция f выпуклая, то субдифференциал Кларка совпадает с обычным субдифференциалом в смысле выпуклого анализа, и в таком случае условие слабой β-квазивыпуклости (5) верно при $\beta = 1$. Ясно, что если неравенство (5) верно для некоторого $\beta = \beta_0 \in (0; 1]$, то оно верно и при $\beta \in (0; \beta_0]$. Примеры функций, для которых возможно проверить свойство слабой β-квазивыпуклости и оценить параметр β, приведены в [Hinder, Sidford, Sohoni, 2020]. В частности, это верно для невыпуклой функции $f(x) = |x|(1 - e^{-|x|})$ при $\beta = 1$ [Hardt, Ma, Recht, 2018], которая имеет Δ-острый минимум при достаточно малом $\Delta > 0$. Отметим также, что субградиент f может быть нулевым только в точке минимума: равенство $\nabla f(x) = 0$ влечет $f(x) \leqslant f(x_*)$, т. е. $f(x) = f^*$.

Будем рассматривать задачи вида (3), где Q — выпуклое замкнутое подмножество $\mathbb{R}^n$, $f^* = f(x_*) = \min\limits_{x \in Q} f(x)$, а f — слабо β-квазивыпуклая функция при некотором $\beta \in (0; 1]$. Мы отправляемся от модификации субградиентного метода с шагом Б. Т. Поляка для задач минимизации слабо β-квазивыпуклой функции [Стонякин, Аблаев, Баран, 2021]

$$x_{k+1} = Pr_Q\{x_k - h_k \nabla f(x_k)\}, \tag{6}$$

где Pr_Q — оператор проектирования на множество Q и $\|\nabla f(x_k)\|_2 \neq 0$ при $k \geqslant 0$.

Предположим, что верно обобщенное условие острого минимума (2), где параметры $\alpha > 0$ и $\Delta > 0$ фиксированы и задано число $\overline{f} \geqslant 0$, $\overline{f} \geqslant f^*$ (его можно рассматривать как приближение минимума). Исследуем метод (6) с шагом

$$h_k = \frac{\beta(f(x_k) - \overline{f})}{\|\nabla f(x_k)\|_2^2}. \tag{7}$$

Справедлива следующая теорема.

Теорема 1. *Пусть f — слабо β-квазивыпуклая функция и для задачи* (3) *с условием* (2) *используется метод* (6) *с шагом $h_k = \frac{\beta(f(x_k)-\overline{f})}{\|\nabla f(x_k)\|_2^2}$. Пусть также $\forall i \geqslant 0$ верно $\alpha^2\beta^2 \leqslant 2\|\nabla f(x_i)\|_2^2$. Тогда верно неравенство*

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \leqslant \prod_{i=0}^{k}\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_i)\|_2^2}\right)\min_{x_*\in X_*}\|x_0-x_*\|_2^2 + \sum_{i=0}^{k-1}\prod_{j=i+1}^{k}\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_j)\|_2^2}\right)\Delta_i + \Delta_k, \quad (8)$$

где $\Delta_k = \frac{\Delta^2}{2\|\nabla f(x_k)\|_2^2}$ для всякого $k \geqslant 0$.

Доказательство. Имеем следующие неравенства (см. [Поляк, 1969], соотношения (3) из доказательства теоремы 1):

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

Это означает, что

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \leqslant \min_{x_*\in X_*}\|x_k-x_*\|_2^2 + h_k^2\|\nabla f(x_k)\|_2^2 - 2\beta^2 h_k(f(x_k)-\overline{f}) =$$
$$= \min_{x_*\in X_*}\|x_k-x_*\|_2^2 - \frac{\beta^2(f(x_k)-\overline{f})^2}{\|\nabla f(x_k)\|_2^2}.$$

С учетом условия (2) получим

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

Пусть $\Delta_k = \frac{\Delta^2}{2\|\nabla f(x_k)\|_2^2}$. Тогда

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \leqslant \min_{x_*\in X_*}\|x_k-x_*\|_2^2\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_k)\|_2^2}\right) + \Delta_k \leqslant$$
$$\leqslant \left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_k)\|_2^2}\right)\left(\min_{x_*\in X_*}\|x_{k-1}-x_*\|_2^2\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_{k-1})\|_2^2}\right)+\Delta_{k-1}\right)+\Delta_k \leqslant$$
$$\leqslant \ldots \leqslant \prod_{i=0}^{k}\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_i)\|_2^2}\right)\min_{x_*\in X_*}\|x_0-x_*\|_2^2 + \sum_{i=0}^{k-1}\prod_{j=i+1}^{k}\left(1-\frac{\alpha^2\beta^2}{2\|\nabla f(x_j)\|_2^2}\right)\Delta_i + \Delta_k.$$

□

Однако некоторая проблема предыдущего подхода заключается в потенциально возможных малых знаменателях дробей из (8), равных квадратам норм градиентов. Тогда накопление в оценке качества решения величин, соответствующих параметру Δ, потенциально может оказаться существенным. Поэтому рассмотрим на классе M-липшицевых функций также метод (6) с шагом

$$h_k = \frac{\beta(f(x_k)-\overline{f})}{M^2}. \quad (9)$$

Тогда верны неравенства

$$\begin{aligned}
\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 &\leqslant \min_{x_*\in X_*}\|x_k-x_*\|_2^2\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)+\frac{\Delta^2}{M^2}\leqslant\\
&\leqslant \min_{x_*\in X_*}\|x_{k-1}-x_*\|_2^2\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^2+\frac{\Delta^2}{M^2}\left(1+1-\frac{\alpha^2\beta^2}{2M^2}\right)\leqslant\\
&\leqslant \min_{x_*\in X_*}\|x_{k-2}-x_*\|_2^2\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^3+\frac{\Delta^2}{M^2}\left(1+1-\frac{\alpha^2\beta^2}{2M^2}+\left(1-\frac{\alpha^2}{2M^2}\right)^2\right)\leqslant\ldots\leqslant\\
&\leqslant\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2+\frac{\Delta^2}{M^2}\sum_{i=0}^{k}\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^i\leqslant\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2+\frac{2\Delta^2}{\alpha^2\beta^2}.
\end{aligned}$$

Таким образом, верна

Теорема 2. *Пусть f — M-липшицева слабо β-квазивыпуклая функция и для задачи* (3) *с условием* (2) *используется метод* (6) *с шагом $h_k=\frac{\beta(f(x_k)-\overline{f})}{M^2}$, причем $\alpha^2\beta^2\leqslant 2M^2$. Тогда для всякого $k\geqslant 0$ верно неравенство*

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2\leqslant\left(1-\frac{\alpha^2\beta^2}{2M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2+\frac{2\Delta^2}{\alpha^2\beta^2}. \tag{10}$$

Отметим, то в задаче из примера 1 для каждой функции $\alpha=1$ и $M=1$. В этом случае использование адаптивного метода и соответствующие оценки теряют смысл и мы можем использовать теорему 2, в которой четко описан уровень влияния значения параметра Δ на качество выдаваемого методом решения.

Следствие 1. *Если в условиях предыдущей теоремы положить $\Delta=0$* (*обычный острый минимум*)*, то метод* (6) *сходится со скоростью геометрической прогрессии*:

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2\leqslant\left(1-\frac{\alpha^2\beta^2}{M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2. \tag{11}$$

Доказательство. Действительно,

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

С учетом условия (1) предыдущее неравенство для всякого $k\geqslant 0$ принимает вид

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

Откуда

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

□

4. Субградиентный метод для квазивыпуклых задач: исследование поведения траектории и возможность отказа от операции проектирования

Здесь мы рассмотрим субградиентный метод для достаточно известного обобщения выпуклости. При этом указанная общность потребовала уже информацию о константе Липшица M в случае, если целевая функция удовлетворяет условию Липшица. Рассмотрим задачу вида (6), где Q — выпуклое замкнутое подмножество $\mathbb{R}^n$, $f^* = f(x_*) = \min\limits_{x \in Q} f(x)$, а функция f квазивыпукла, т. е.

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

Предположим, что f удовлетворяет условию Липшица с константой $M > 0$:

$$|f(x) - f(y)| \leqslant M\|x - y\|_2 \quad \forall x,\ y \in Q. \tag{12}$$

Введем (следуя [Нестеров, 1989; Нестеров, 2010]) вспомогательную величину

$$\upsilon_f(x,\ x_*) := \begin{cases} \left\langle \dfrac{\nabla f(x)}{\|\nabla f(x)\|_2},\ x - x_* \right\rangle, & \text{если } x \neq x_*; \\ 0 & \text{при } x = x_*. \end{cases} \tag{13}$$

Если $\nabla f(x) = 0$ при $x \neq x_*$, то вместо $\nabla f(x)$ можно использовать ненулевой вектор нормали $\mathcal{D}f(x)$ ко множеству уровня функции f в точке x. Но для упрощения изложения далее сделаем допущение, что $\nabla f(x) \neq 0$ при $x \neq x_*$. Справедлива [Нестеров, 2010] следующая

Лемма 1. *Если f — квазивыпуклая и M-липшицева функция, то для всякого $x \in Q$ верно неравенство*

$$f(x) - f(x_*) \leqslant M\nu_f(x,\ x_*). \tag{14}$$

Рассмотрим метод ($k = 0,\ 1,\ 2,\ \ldots$)

$$x_{k+1} = Pr_Q\{x_k - h_k \nabla f(x_k)\}, \quad \text{где} \quad h_k = \frac{f(x_k) - f(x_*)}{M\|\nabla f(x_k)\|_2}, \tag{15}$$

где M удовлетворяет (14).

Теорема 3 (см. [Стонякин, Аблаев, Баран, 2021]). *Пусть верно* (14) *при некотором* $0 < M < +\infty$ *и f имеет острый минимум. Тогда метод* (15) *сходится с линейной скоростью*:

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2}{M^2}\right)^{k+1} \min_{x_* \in X_*} \|x_0 - x_*\|_2^2. \tag{16}$$

Некоторым недостатком алгоритма (15) по сравнению с алгоритмом (6) может считаться требование знать M из неравенства леммы 1 при выполнении алгоритма. Однако эта же особенность позволяет предложить вариацию субградиентного метода с Δ-обобщением острого минимума вида (2). Справедлива следующая теорема.

Теорема 4. *Пусть M-липшицева функция f имеет Δ-острый минимум при некотором $\Delta > 0$ и известно $\overline{f} \geqslant f^*$. Тогда для метода $x_{k+1} = Pr_Q\{x_k - h_k \nabla f(x_k)\}$ с шагом вида*

$$h_k = \frac{f(x_k) - \overline{f}}{M\|\nabla f(x_k)\|_2} \tag{17}$$

справедливо неравенство

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \min_{x_* \in X_*} \|x_0 - x_*\|_2^2 \left(1 - \frac{\alpha^2}{2M^2}\right)^{k+1} + \frac{2\Delta^2}{\alpha^2}. \tag{18}$$

Таким образом, наличие параметра Δ (связанного с рассматриваемым обобщением острого минимума) приводит к дополнительному слагаемому вида $O(\Delta)$ в оценке (18). По-видимому, в случае малых значений $\|\nabla f(x_k)\|_2$ для алгоритма (6) такой вывод сделать уже нельзя и можно довольствоваться лишь оценкой (8).

Отметим, что на классе выпуклых функций неравенства (10) и (18) совпадают. Однако функция f в указанных неравенствах удовлетворяет разным предположениям об аналоге выпуклости. На практике сходимость метода с шагом (17) может оказаться лучше, чем у метода с шагом (9), поскольку возможна ситуация, когда $\|\nabla f(x_k)\|_2 < M$ при некоторых k, что приводит к увеличению длины шага (17).

Замечание 1. Оказывается [Стонякин, Аблаев, Баран, 2021], что рассмотренный в этом разделе алгоритм (15) универсален в том смысле, что при некоторых условиях можно выписать теоретический результат о сходимости (15) со скоростью геометрической прогрессии для задач минимизации квазивыпуклых субдифференцируемых (т. е. локально липшицевых) гёльдеровых функций с острым минимумом и заранее известным точным значением f^*. Если f удовлетворяет условию Гёльдера

$$|f(x) - f(y)| \leqslant M_\nu \|x - y\|_2^\nu \quad \forall x, y \in Q \tag{19}$$

при некотором фиксированном $\nu \in [0; 1]$ и $0 \leqslant M_\nu \leqslant +\infty$, то с учетом (1) и (19) получаем, что для всякого $x \in Q$

$$\alpha \min_{x_* \in X_*} \|x - x_*\|_2 \leqslant f(x) - f^* \leqslant M_\nu \min_{x_* \in X_*} \|x - x_*\|_2^\nu.$$

Поэтому

$$\min_{x_* \in X_*} \|x - x_*\|_2^{1-\nu} \leqslant \frac{M_\nu}{\alpha},$$

откуда при $\nu < 1$ верно неравенство

$$\min_{x_* \in X_*} \|x - x_*\|_2 \leqslant \left(\frac{M_\nu}{\alpha}\right)^{1/(1-\nu)}.$$

Поэтому при $0 \leqslant \nu < 1$ можно локализовать допустимую область Q, заменив ее на пересечение исходного множества Q с евклидовым шаром с центром в точке x_0 и радиусом $\left(\frac{M_\nu}{\alpha}\right)^{1/(1-\nu)}$. Если при этом $a > a_0 > 0$ не является достаточно малым, то для подходящей $M_\nu > 0$ верно неравенство

$$M_\nu a^\nu \leqslant \frac{M_\nu^{2/(1+\nu)}}{2} \frac{a^2}{\delta^{(1-\nu)/(1+\nu)}} + \frac{\delta}{2} \quad \forall \delta > 0.$$

Такое неравенство позволит оценить

$$M = \max\left\{M_\nu, \frac{M_\nu^{2/(1+\nu)}}{2}\right\}$$

при достаточно большом $\upsilon_f(x, x_*)$ для некоторого $x_* \in X_*$ и применять метод (15), если требования к искомой точности не очень высоки. В этом случае теорема 3 позволит гарантировать достижение сопоставимой с a_0 точности решения по аргументу за линейное время. Отметим, что описываемые рассуждения основаны на работе [Stonyakin et al., 2020], где впервые была замечена возможность переноса оптимальной на классе выпуклых липшицевых задач оценки сложности субградиентного метода на классе задач минимизации квазивыпуклых гёльдеровых субдифференцируемых функций в случае не очень высокой точности $\varepsilon \geqslant a_0 > 0$ приближенного решения. При этом для достаточно высокой точности $\varepsilon < a_0$ такой вопрос тривиален ввиду локальной липшицевости f.

Рассмотрим еще такую ситуацию. С точки зрения вычислительной сложности операция проектирования на допустимое множество может оказаться довольно затратной. Поэтому возникает вопрос об исследовании субградиентного метода без операций проектирования на допустимое множество на итерациях. Для этого необходимо описать удаление траекторий метода от начальной точки. Рассмотрим аналог метода (15) без операции проектирования x_k на множество Q и исследуем удаление траектории $\{x_k\}$ от начальной точки x_0:

$$x_{k+1} = x_k - h_k \nabla f(x_k), \quad \text{где} \quad h_k = \frac{f(x_k) - f^*}{M\|\nabla f(x_k)\|_2}. \tag{20}$$

Это позволит, в частности, определить условия на Q, которые позволят применить метод (15) для задач с острым минимумом без дополнительной операции проектирования x_k на Q. Если допустить, что f — M-липшицева при $M > 0$, то

$$f(x_k) - f^* \leqslant M\|x_k - x_*\|_2$$

для ближайшего к x_k точного решения x_*: $f(x_*) = f^*$.

Оценим $\|x_N - x_0\|_2$. Из теоремы 3 для всякого $k \geqslant 0$ имеем

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

Поэтому

$$\|x_{k+1} - x_k\|_2 = \frac{f(x_k) - f^*}{M} \leqslant \frac{1}{M}\left(1 - \frac{\alpha^2}{M^2}\right)^{k/2} \min_{x_* \in X_*} \|x_0 - x_*\|_2,$$

откуда имеем

$$\|x_N - x_0\|_2 \leqslant \sum_{k=0}^{N-1} \|x_{k+1} - x_k\|_2 \leqslant \frac{1}{M} \sum_{k=0}^{N-1} \left(1 - \frac{\alpha^2}{M^2}\right)^{k/2} \min_{x_* \in X_*} \|x_0 - x_*\|_2^2 \leqslant$$

$$\leqslant \frac{1}{M\left(1 - \sqrt{1 - \frac{\alpha^2}{M^2}}\right)} \|x_0 - x_*\|_2 = \frac{M + \sqrt{M^2 - \alpha^2}}{\alpha^2} \|x_0 - x_*\|_2.$$

Таким образом, имеет место следующая

Теорема 5. *Если f удовлетворяет условию Липшица с постоянной $M > 0$ и условию острого минимума* (1) *с постоянной $\alpha > 0$, то для метода* (20) *справедливо неравенство*

$$\|x_N - x_0\|_2 \leqslant \frac{M + \sqrt{M^2 - \alpha^2}}{\alpha^2} \|x_0 - x_*\|_2.$$

Таким образом, если допустимое множество Q содержит шар с центром в точке x_0 радиусом $\frac{M + \sqrt{M^2 - \alpha^2}}{\alpha^2}\|x_0 - x_*\|_2$, то можно применять метод (20) без дополнительных операций проектирования x_{k+1} ($k \geqslant 0$) на множество Q.

5. Субградиентный метод для задач с Δ-острым минимумом с использованием на итерациях неточных δ-субградиентов

Теперь допустим, что нам недоступна информация о точном значении субградиента ∇f, используемая на итерациях алгоритма (6). Будем считать, что доступен лишь δ-субградиент f в текущей точке x, т. е.

$$f(y) - f(x) \geqslant \langle \nabla_\delta f(x),\, y - x\rangle - \delta \quad \forall y \in Q \tag{21}$$

при некотором фиксированном $\delta > 0$ (если $\delta = 0$, то $\nabla_\delta f(x) = \nabla f(x)$).

Например, для функции $f(x) = \max\limits_{z\in P} \varphi(x, z)$, где φ выпукла по x и непрерывна по $z \in P$ (P — некоторое компактное множество), $\nabla_\delta f(x)$ есть обычный субградиент по переменной x функции $\varphi(x, z_\delta)$ такой, что

$$\max_{z\in P} \varphi(x, z) - \varphi(x, z_\delta) \leqslant \delta.$$

В качестве примера задачи, где можно использовать вместо обычного субградиента δ-субградиент, рассмотрим задачу проектирования точки на выпуклый компакт X_*. Если точка X далеко от множества X_*, то при нахождении субградиента функции расстояния от точки до множества можно пренебречь точностью направления единичного вектора субградиента и использовать δ-субградиент вместо обычного субградиента. Отметим, что такой подход реализован далее экспериментально (см. замечание 2).

Перейдем к теоретическому обоснованию аналога метода с использованием δ-субградиентов. По аналогии с (13) введем вспомогательную величину

$$v_f^\delta(x, x_*) := \left\langle \frac{\nabla_\delta f(x)}{\|\nabla_\delta f(x)\|_2},\, x - x_* \right\rangle$$

при $\nabla_\delta f(x) \neq 0$ и $v_f^\delta(x, x_*) = 0$ при $\nabla_\delta f(x) = 0$.

В предположении о том, что функция удовлетворяет условию Липшица с константой M, рассмотрим метод (6) с шагом

$$h_k = \frac{f(x_k) - f^* - \delta}{M\|\nabla_\delta f(x_k)\|_2}. \tag{22}$$

Тогда справедлив следующий результат о качестве точки выхода метода (6).

Теорема 6. *Пусть f — квазивыпуклая M-липшицева функция с острым минимумом* (1)*, для которой в каждой точке $x \in Q$ можно вычислить δ-субградиент* (21) *и для задачи $f(x) \to \min\limits_{x\in Q}$ используется алгоритм* (6) *с шагом* (22)*, а также $\alpha^2 \leqslant 2M^2$. Тогда верно неравенство*

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

Доказательство. Справедливы следующие неравенства (см. доказательство теоремы 1):

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

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

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

по лемме 1.1 из [Стонякин, 2020] имеем

$$\leqslant \min_{x_*\in X_*} \|x_k - x_*\|_2^2 - \left(\frac{f(x_k) - f^* - \delta}{M}\right)^2.$$

С учетом условия острого минимума (1)

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \leqslant \min_{x_*\in X_*}\|x_k-x_*\|_2^2 - \frac{\alpha^2}{M^2}\min_{x_*\in X_*}\|x_k-x_*\|_2^2 - \frac{2\alpha\delta}{M^2}\min_{x_*\in X_*}\|x_k-x_*\|_2 - \frac{\delta^2}{M^2}.$$

Заметим, что ввиду неравенства $2xy \leqslant x^2+y^2$ (верного для любых x, y)

$$\frac{2\alpha\delta}{M^2}\min_{x_*\in X_*}\|x_k-x_*\|_2 \leqslant \frac{\alpha^2}{2M^2}\min_{x_*\in X_*}\|x_k-x_*\|_2^2 + \frac{2\delta^2}{M^2}.$$

Тогда

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

Далее, имеем

$$\min_{x_*\in X_*}\|x_{k+1}-x_*\|_2^2 \leqslant \left(1-\frac{\alpha^2}{2M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2 + \left(\sum_{i=0}^{k}\left(1-\frac{\alpha^2}{2M^2}\right)^i\right)\frac{\delta^2}{M^2} =$$

$$= \left(1-\frac{\alpha^2}{2M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0-x_*\|_2^2 + \frac{2\delta^2}{\alpha^2}. \quad (23)$$

□

Теперь рассмотрим метод (6) с шагом

$$h_k = \frac{f(x_k)-\overline{f}-\delta}{M\|\nabla_\delta f(x_k)\|_2} \quad (24)$$

в предположении, что f удовлетворяет условию Δ-острого минимума (2). Тогда

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

и с учетом Δ-острого минимума имеем

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

Поэтому по аналогии с (23)

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

Таким образом, верна

Теорема 7. *Пусть f — выпуклая M-липшицева функция* ($M>0$)*, удовлетворяющая условию* (2)*, такая, что в каждой точке доступен δ-субградиент f и для задачи минимизации f на множестве Q используется метод* (6) *с шагом* (24)*, а также $\alpha^2 \leqslant 2M^2$. Тогда верно неравенство*

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

Как видим, условие Δ-острого минимума, а также использование на итерациях рассматриваемого метода δ-субградиентов влекут за собой дополнительное слагаемое $\frac{2(\Delta+\delta)^2}{\alpha^2}$ в оценке качества выдаваемого указанным методом решения.

6. Адаптивная оценка скорости сходимости одного субградиентного метода для сильно выпуклых задач и вычислительные эксперименты для задачи о наименьшем покрывающем шаре

Теперь перейдем к экспериментальному сравнению работы предложенных в § 3 подходов для задач с Δ-острым минимумом с работой известного субградиентного метода [Julien, Schmidt, Bach, 2012] для некоторых сильно выпуклых задач. Выбор класса сильно выпуклых задач и метода [Julien, Schmidt, Bach, 2012] для сравнения обусловлен известными и применимыми на практике теоретическими оценками качества приближенного решения по аргументу. Отметим адаптивный аналог [Stonyakin et al., 2021] теоретической оценки качества выдаваемого решения для субградиентного метода [Julien, Schmidt, Bach, 2012]. Напомним, что в работе рассматриваются задачи вида

$$f(x) \to \min_{x \in Q}, \tag{25}$$

где Q — выпуклое замкнутое подмножество $\mathbb{R}^n$. Для субградиентного метода вида

$$x_{k+1} := Pr_Q\{x_k - h_k \nabla f(x_k)\}, \quad \text{где} \quad h_k = \frac{2}{\mu(k+1)}, \tag{26}$$

известна следующая оценка скорости сходимости [Julien, Schmidt, Bach, 2012]:

$$f(\widehat{x}) - f(x_*) \leqslant \frac{2M^2}{\mu(N+1)} \quad \text{при} \quad \widehat{x} = \sum_{k=1}^{N} \frac{2k}{N(N+1)} x_k, \tag{27}$$

где M — константа Липщица целевой функции f.

Оказывается, что данную оценку можно несколько улучшить на классе сильно выпуклых задач [Stonyakin et al., 2021]. Для доказательства понадобится следующая вспомогательная лемма.

Лемма 2. *Если x_k и x_{k+1} удовлетворяют* (26)*, то для произвольного $x \in Q$ верно неравенство*

$$h_k\langle \nabla f(x_k),\, x_k - x\rangle \leqslant \frac{h_k^2\|\nabla f(x_k)\|_2^2}{2} + \frac{1}{2}\|x - x_k\|_2^2 - \frac{1}{2}\|x - x_{k+1}\|_2^2.$$

Доказательство. Непосредственно проверяются следующие неравенства:

$$\begin{aligned} h_k\langle \nabla f(x_k),\, x_k - x\rangle &\leqslant h_k\langle \nabla f(x_k),\, x_k - x_{k+1}\rangle + \frac{1}{2}\|x - x_k\|_2^2 - \frac{1}{2}\|x - x_{k+1}\|_2^2 - \frac{1}{2}\|x_{k+1} - x_k\|_2^2 \leqslant \\ &\leqslant h_k\|\nabla f(x_k)\|_2\|x_{k+1} - x_k\|_2 + \frac{1}{2}\|x - x_k\|_2^2 - \frac{1}{2}\|x - x_{k+1}\|_2^2 - \frac{1}{2}\|x_{k+1} - x_k\|_2^2 \leqslant \\ &\leqslant \frac{h_k^2\|\nabla f(x_k)\|_2^2}{2} + \frac{1}{2}\|x - x_k\|_2^2 - \frac{1}{2}\|x - x_{k+1}\|_2^2. \end{aligned}$$

□

Согласно лемме 2 получим, что для всяких $k \geqslant 0$ и $x \in Q$

$$\langle \nabla f(x_k),\, x_k - x\rangle \leqslant \frac{h_k\|\nabla f(x_k)\|_2^2}{2} + \frac{1}{2h_k}\|x - x_k\|_2^2 - \frac{1}{2h_k}\|x - x_{k+1}\|_2^2.$$

Далее, с учетом μ-сильной выпуклости f

$$f(x_k) - f(x) + \frac{\mu}{2}\|x_k - x\|_2^2 \leqslant \langle \nabla f(x_k),\, x_k - x\rangle \quad \forall x \in Q,$$

откуда при всяком $k \geqslant 0$

$$2kf(x_k) - 2kf(x) + k\mu\|x - x_k\|_2^2 \leqslant \frac{2k\|\nabla f(x_k)\|_2^2}{\mu(k+1)} + k(k+1)\frac{\mu}{2}\|x - x_k\|_2^2 - k(k+1)\frac{\mu}{2}\|x - x_{k+1}\|_2^2.$$

Пусть алгоритм (26) отработал N шагов. Тогда можно просуммировать неравенства по k от 1 до N и учесть, что $\frac{k}{k+1} \leqslant 1$:

$$\sum_{k=1}^{N} 2k\left(f(x_k) - f(x) + \frac{\mu}{2}\|x_k - x\|_2^2\right) \leqslant \frac{2N\|\nabla f(x_k)\|_2^2}{\mu},$$

откуда с учетом $2(1 + 2 + \ldots + N) = N(N+1)$ получаем

$$\sum_{k=1}^{N} \frac{2k}{N(N+1)}\left(f(x_k) - f(x) + \frac{\mu}{2}\|x_k - x\|_2^2\right) \leqslant \frac{2\|\nabla f(x_k)\|_2^2}{\mu(N+1)}.$$

Ввиду выпуклости f теперь верно неравенство

$$f(\widehat{x}) - f(x) \leqslant \frac{2}{\mu N(N+1)} \sum_{k=1}^{N} \frac{k\|\nabla f(x_k)\|_2^2}{k+1} \leqslant \varepsilon$$

после $N = O\left(\frac{M^2}{\mu\varepsilon}\right)$ итераций алгоритма (26). Поэтому справедлива следующая

Теорема 8. *Пусть f — μ-сильно выпуклая функция. Тогда после N итераций алгоритма*:

$$x_{k+1} := Pr_Q\{x_k - h_k\nabla f(x_k)\}, \quad \text{где} \quad h_k = \frac{2}{\mu(k+1)},$$

будет верно неравенство

$$f(\widehat{x}) - f(x_*) \leqslant \frac{2}{\mu N(N+1)} \sum_{k=1}^{N} \frac{k\|\nabla f(x_k)\|_2^2}{k+1}, \tag{28}$$

где

$$\widehat{x} = \sum_{k=1}^{N} \frac{2k}{N(N+1)} x_k.$$

Если f еще и M-липшицева при $M > 0$, то

$$f(\widehat{x}) - f(x) \leqslant \varepsilon$$

после $N = O\left(\frac{M^2}{\mu\varepsilon}\right)$ итераций алгоритма (26).

Отметим, что если x_* — точное решение задачи минимизации f, то можно получить оценку скорости сходимости по аргументу вида

$$\|\widehat{x} - x_*\|_2 \leqslant \frac{4}{\mu N(N+1)} \sum_{k=1}^{N} \frac{k\|\nabla f(x_k)\|_2^2}{k+1} \leqslant \frac{4M^2}{\mu(N+1)}. \tag{29}$$

Полученный в теореме 8 результат применим и в случаях, когда константа Липщица (M) бесконечна или ее значение сложно оценить. Более того, данный подход может быть распространен на важные прикладные задачи, среди которых задача бинарной классификации методом опорных векторов (SVM) [Julien, Schmidt, Bach, 2012]. По аналогии с работой [Julien, Schmidt, Bach, 2012] можно применять стохастический вариант зеркального спуска (26). Также отметим, что данные рассуждения и метод могут быть обобщены на класс вариационных неравенств, лагранжевых и седловых задач [Stonyakin et al., 2021].

Для сравнения скорости сходимости метода [Julien, Schmidt, Bach, 2012] и полученной в теореме 8 оценки с предложенными в настоящей работе (§ 3) вариациями субградиентных методов для задач с Δ-острым минимумом проведены численные эксперименты для задачи о наименьшем покрытии точек шаром для 2-сильно выпуклой функции

$$f(x) := \max_{x \in Q} \left\{ \|x - a_0\|_2^2,\ \|x - a_1\|_2^2,\ \ldots,\ \|x - a_m\|_2^2 \right\}, \tag{30}$$

а также для не сильно выпуклой (но выпуклой) функции

$$f(x) := \max_{x \in Q} \left\{ \|x - a_0\|_2,\ \|x - a_1\|_2,\ \ldots,\ \|x - a_m\|_2 \right\}. \tag{31}$$

Начнем с иллюстрации преимуществ адаптивной оценки метода [Julien, Schmidt, Bach, 2012] из теоремы 8. Будем рассматривать множество Q, которое равно евклидову шару с центром в 0. Начальная точка выбиралась случайно, но внутри Q. На рис. 1 и 2 ниже показаны поведение и характер убывания для оригинальной оценки (27) (сплошная линия), адаптивной оценки (28) (штрихпунктирная линия) и непосредственно невязки по функции и по аргументу соответственно (штриховая линия). На рис. 1 показано поведение глобальной оценки, адаптивной и невязки по функции и аргументу в случае ограниченного Q ($R = 5$). Случай, когда глобальной оценкой воспользоваться не получится, показан на рис. 2. В этом случае мы работаем на $Q = \mathbb{R}^n$, потому наблюдаем соотношение между адаптивной оценкой и непосредственно невязкой. Данные графики наглядно демонстрируют, насколько более точной может оказаться адаптивная оценка (28) для задачи (30).

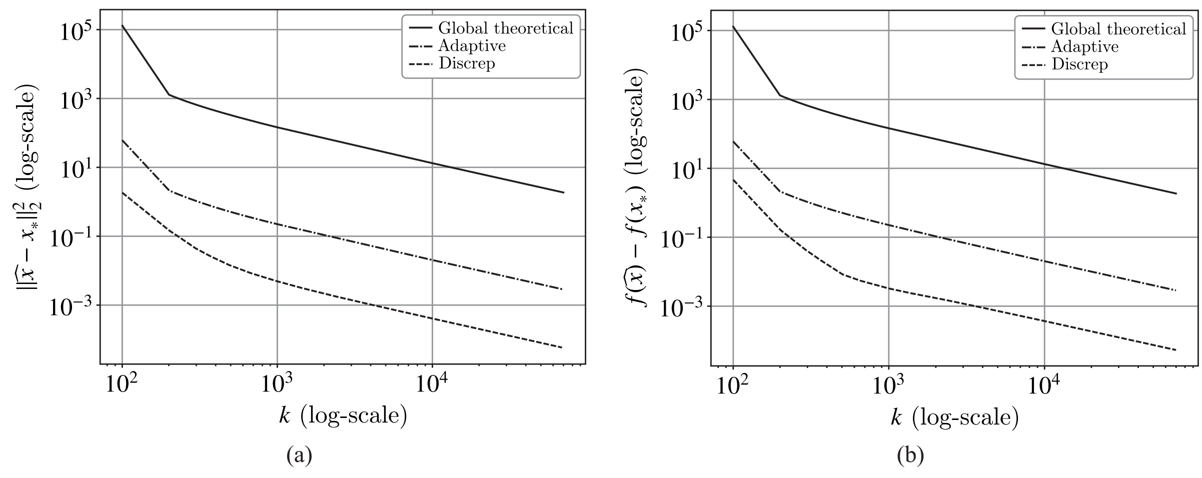


Рис. 1. Результаты решения задачи минимизации (30), где $n = 1000$, $r = 5$ и шар Q радиусом 4

Теперь перейдем к выпуклой постановке (31) с целью исследования эффективности предложенных в § 3 субградиентных методов с Δ-острым минимумом. К существующему набору точек, представленных для покрытия, с известным значением центра добавим дополнительную точку, которая находится вне исходного шара достаточно близко к границе (удалена не более чем на $\Delta > 0$). Данный подход позволяет оценить «приближенное» значение минимума $\overline{f}$, что позволит применить разработанные выше варианты субградиентного метода с Δ-острым минимумом. При этом новое значение минимума останется внутри исходной сферы. Поскольку оптимальное

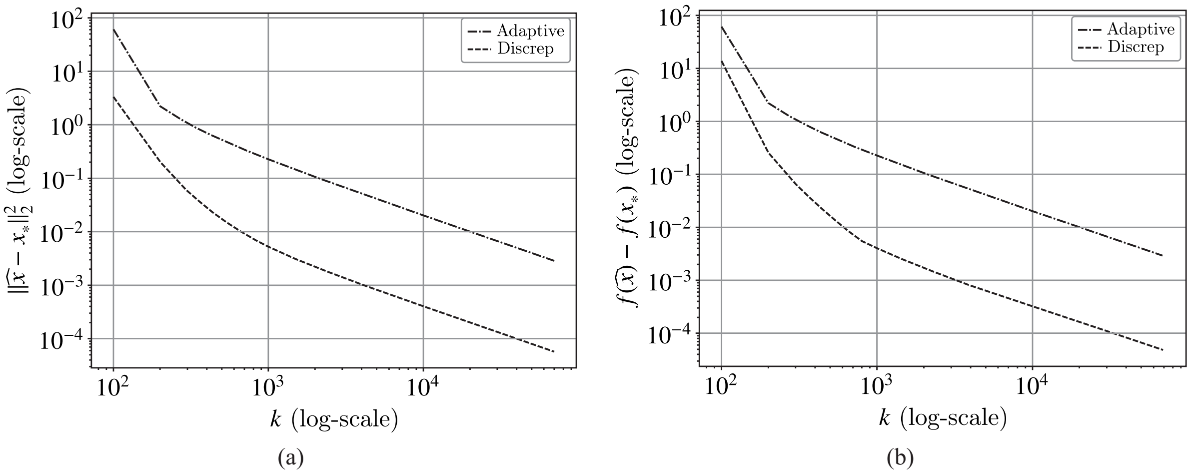


Рис. 2. Результаты решения задачи минимизации (30), где $n = 1000$, $r = 5$ и $Q = \mathbb{R}^n$

значение функции — это радиус искомого шара, покрывающего все точки, а x_* всегда будет расположена внутри него, то для всякого x верно неравенство $f(x) \geqslant \|x - x_*\|_2$. Рассмотрим целевую функцию вида

$$f(x) := \alpha \max_{x \in Q}\{\|x - a_0\|_2, \|x - a_1\|_2, \ldots, \|x - a_m\|_2\}. \tag{32}$$

Тогда значение Δ можно оценить из (2): $f(x) - \overline{f} \geqslant \alpha\|x - x_*\|_2 - \Delta$, $\Delta \geqslant \overline{f}$.

Отметим, что данная постановка значительно влияет на величину теоретической оценки качества решения (8) для метода (6). Наиболее значимый вклад в оценку (8) дает последнее слагаемое $\frac{\Delta^2}{2\|\nabla f(x_k)\|_2^2}$, причем $\Delta \sim \overline{f} \sim \alpha\|\overline{x} - a\|_2$ и $\|\nabla f(x_k)\|_2 = \alpha$. Поэтому последнее слагаемое пропорционально радиусу шара, соответствующему «приближенному» решению. Это и подтверждается экспериментально. Для сравнения ниже на рис. 3 и 4 приведены результаты работы для того же набора входных точек, которые необходимо покрыть в обоих постановках — (32) и (30). Начальная точка также одна и та же. Сравниваются методы (6) и (26). Первый из этих методов обеспечивает сходимость буквально за 10 итераций к «приближенному» решению с заданной точностью и даже позволяет эту точность повысить. Второй же метод достигает схожих (с геометрической точки зрения) результатов за значительно большее количество итераций, однако он позволяет повышать точность приближенного решения на дальнейших итерациях.

Подтверждение данного теоретического наблюдения хорошо иллюстрируется на рис. 3 и 4. На рис. 3 показано поведение субградиентного спуска, использующего Δ-острый минимум (теорема 4), а именно быстрая сходимость к «приближенному» решению. Штрихпунктирная линия соответствует оценке (2), а штриховая — невязке по функции и аргументу. На рис. 4 показано поведение метода для той же задачи, но с использованием сильно выпуклого целевого функционала (теорема 8). Скорость убывания уже не столь высокая, но точность получаемого решения в итоге выше. Сплошная линия — это глобальная оценка (27), штрихпунктирная — адаптивная (28), а штриховая — невязка по функции и аргументу.

Тем не менее сравнение с известным точным решением x_*, а также график динамики значения целевой функции показывают, что за малое число шагов (значительно меньшее, чем для метода (26)) реализация метода (6) приводит к неплохому качеству приближенного решения. При этом, однако, для метода (6) после достижения такого уровня дальнейшее повышение качества выходной точки, в отличие от метода (26), уже не наблюдается.

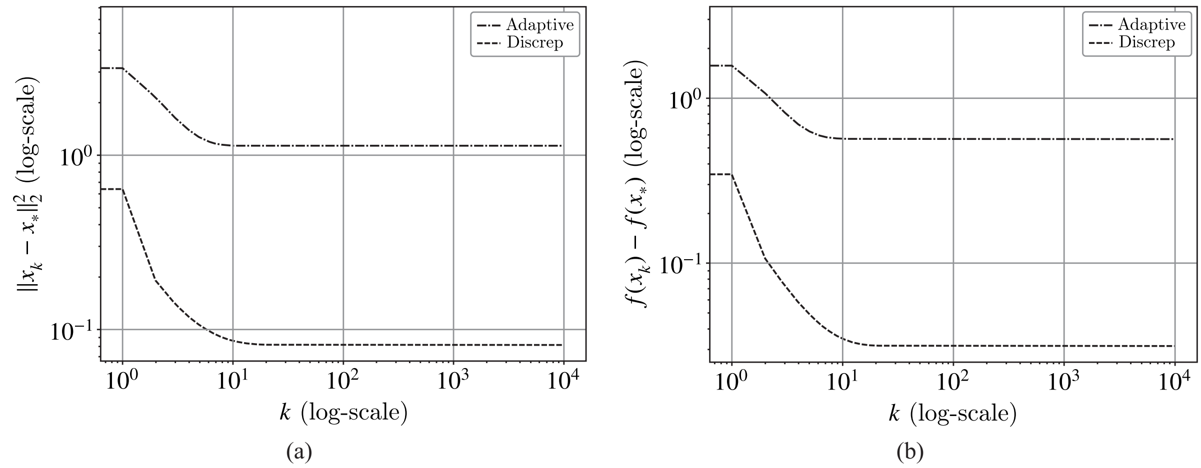


Рис. 3. Результаты решения задачи минимизации (32), где $n = 1000$, $r = 0{,}7525$, $\alpha = 0{,}6$

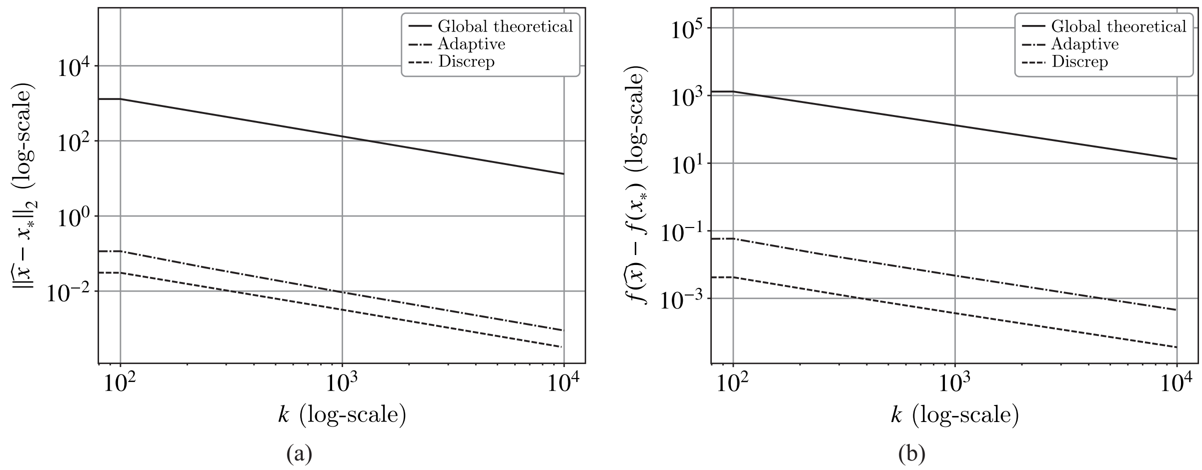


Рис. 4. Результаты решения задачи минимизации (30), где $n = 1000$, $r = 0{,}7525$

7. Другие вычислительные эксперименты для рассматриваемых методов с шагом типа Б. Т. Поляка

Для демонстрации эффективности предлагаемого алгоритма (6) с шагом (7) (адаптивный вариант) и шагом (9) (неадаптивный вариант) были также проведены некоторые численные эксперименты для сильно выпуклых задач и сравнены результаты их работы с предложенным в [Julien, Schmidt, Bach, 2012] субградиентным методом с шагом вида

$$h_k = \frac{2}{\mu(k+1)} \quad \forall k \geqslant 0, \tag{33}$$

где μ — параметр сильной выпуклости целевой функции. Для алгоритма (33) из [Julien, Schmidt, Bach, 2012] известна следующая оценка качества выходной точки после N итераций:

$$\|\widehat{x} - x_*\|_2^2 \leqslant \frac{4M^2}{\mu^2(N+1)} \quad \forall k \geqslant 0. \tag{34}$$

Отметим, что вместо $\widehat{x}$ можно выбрать $x_{\min}$: $f(x_{\min}) = \min\limits_{k=\overline{1,N}} f(x_k)$. Если значения $f(x_k)$ убывают с ростом k, то в неравенстве (34) можно $\widehat{x}$ заменить на x_N.

Эксперименты проводились для множества Q, равного шару с центром в $0 \in \mathbb{R}^n$ и радиусом R, т. е. $Q = B_R(0) = \{x \in \mathbb{R}^n\colon \|x\|_2 \leqslant R\}$. При этом рассмотрены три примера.

Пример 4. Пусть целевая функция имеет вид

$$f(x) = \|x\|_2 + 2\gamma\|x\|_2^2, \quad \gamma > 0. \tag{35}$$

Такая функция M-липшицева с константой $M = 1 + 2\gamma R$ и 4γ-сильно выпукла ($\beta = 1$ в (7) и (9)). Также отметим, что f из (35) удовлетворяет обобщенному условию острого минимума (2) при $\alpha = 1$, $\Delta = 0$, $\overline{f} = f^* = 0 \in \mathbb{R}$ и $x_* = 0 \in Q$.

Пример 5. В данном примере (см. пример 2) выберем негладкую функцию вида

$$f(x) = \|x\|_2 + 2\gamma\|x - c\|_2^2, \tag{36}$$

которая является M-липшицевой с константой $M = 1 + 2\gamma(R + \|c\|_2)$ и 4γ-сильно выпуклой. При этом функция (36) удовлетворяет обобщенному условию острого минимума (2) при $\alpha = 1$, $\Delta = 2\gamma\|c\|_2^2$ и $\overline{f} = \gamma\|c\|_2^2$.

Сравниваемые алгоритмы были запущены для $n = 1000$ и $n = 10\,000$ со стартовой точкой $x_0 = \left(\frac{R}{\sqrt{n}}, \ldots, \frac{R}{\sqrt{n}}\right)$ и с разными значениями R и γ. Результаты сравнения представлены на рис. 5 и 6 для примера 4 и на рис. 7 для примера 5. *На этих рисунках жирная, пунктирная и штриховая кривые указывают на поведение адаптивной и неадаптивной версий алгоритма* (6)*, а также субградиентного метода* [Julien, Schmidt, Bach, 2012] *соответственно. Для примера* 4 *мы отображаем динамику значений* $f(x_k) - f_* = f(x_k)$ *и* $\|x_k - x_*\|_2^2 = \|x_k\|_2^2$ *в зависимости от номера итерации k. Для примера* 5*, ввиду того что известно лишь* $\overline{f}$*, мы отображаем* $f(x_k) - \overline{f}$*, а также теоретические оценки* (8)*,* (10)*,* (34)*, а также левую часть* (29) *для иллюстрации качества решения, достигнутого сравниваемыми алгоритмами.*

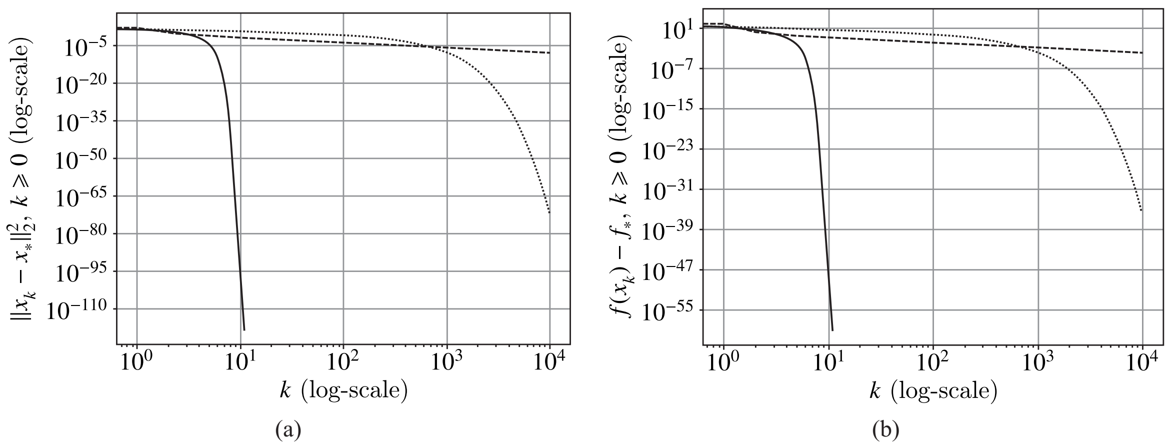
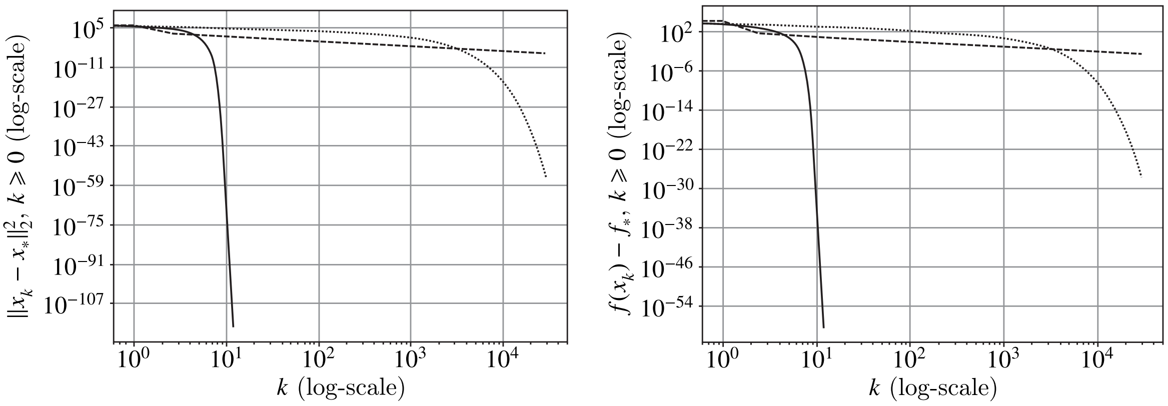


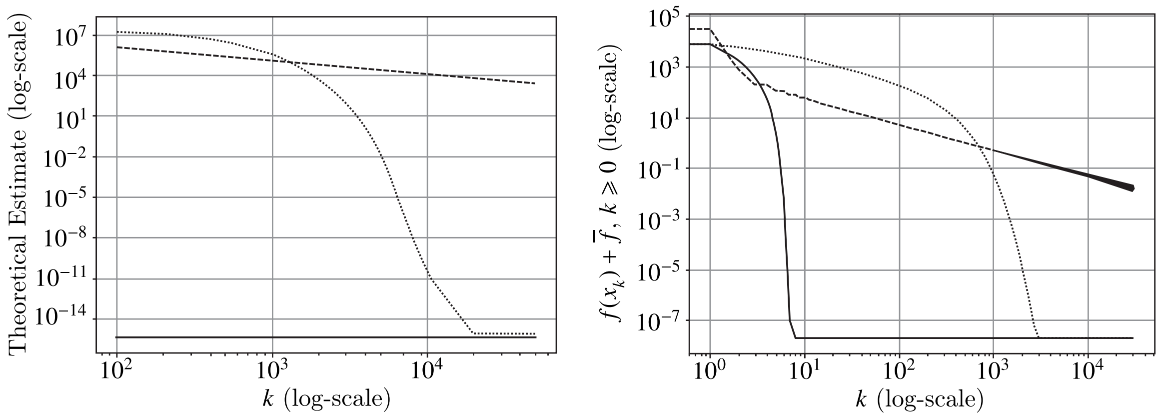
Рис. 5. Результаты для примера 4, где $n = 10\,000$, $R = 10$ и $\gamma = 0{,}5$

Полученные результаты для примеров 4 и 5 показывают хорошую эффективность предложенных в настоящей статье вариаций субградиентного метода с шагом Б. Т. Поляка. Как адаптивная, так и неадаптивная версии алгоритма (6) работают лучше по сравнению с предложенным в [Julien, Schmidt, Bach, 2012] субградиентным методом. Кроме того, мы видим, что адаптивная версия является лучшей с точки зрения скорости сходимости как по функции, так и по аргументу. Кроме того, предлагаемые в работе вариации алгоритма (6) с шагом Б. Т. Поляка (как в адаптивном, так и в неадаптивном случае) приводят к лучшему качеству точки выхода как



(a) (b)

Рис. 6. Результаты для примера 4, где $n = 10\,000$, $R = 1000$ и $\gamma = 0{,}01$



(a) (b)

Рис. 7. Результаты для примера 5, где $n = 1000$, $R = 5000$, $\gamma = 0{,}001$ и $\|c\|_2 = 0{,}01$

с точки зрения теоретических оценок (8) и (10), так и с точки зрения практической работы методов (см. рис. 7), если нет высоких требований к точности. При этом для вариации алгоритма (6) необходима информация о приближенном значении целевой функции $\overline{f}$ (работа методов может позволить получить соответствующее качество в точке выхода и «по аргументу»).

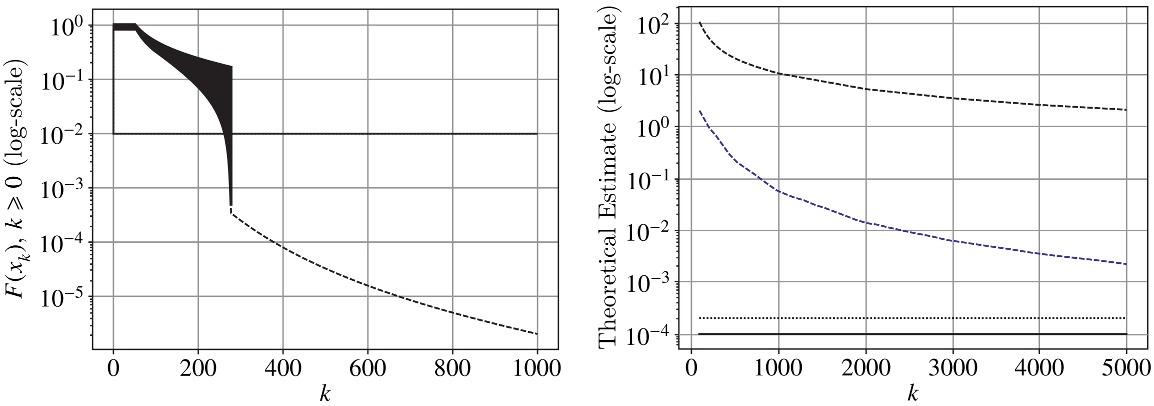
Также мы можем убедиться в эффективности предложенных в § 3 настоящей работы подходов в случае Δ-острого минимума на следующем примере.

Пример 6. Пусть целевая функция имеет следующий вид:

$$F(x) = \min_{A \in S}\{\|x - A\|_2\} + \gamma\|x\|_2^2, \tag{37}$$

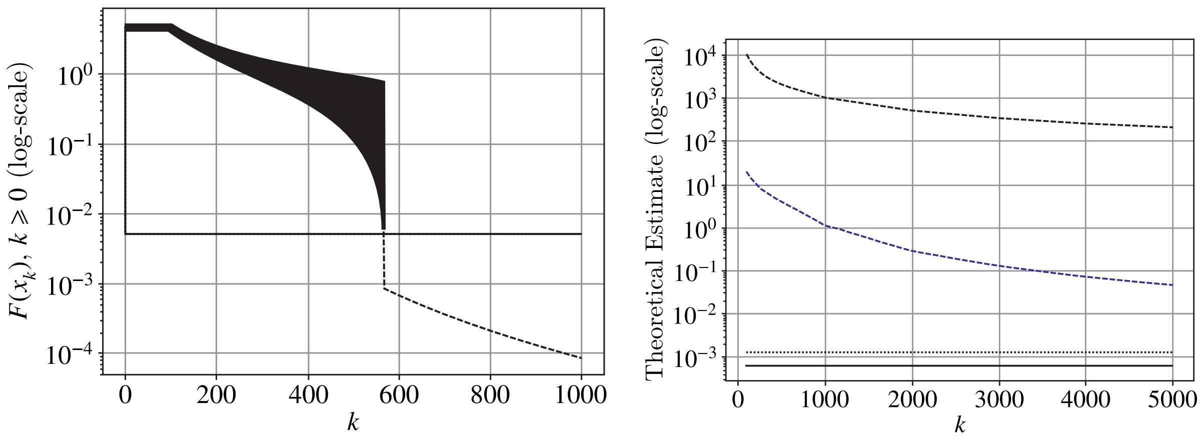
где S — выпуклая оболочка двух шаров с одним радиусом r, так что $S \subset Q$, функция F является M-липшицевой с константой $M = 1 + 2\gamma R$ и 2γ-сильно выпуклой (считаем, что Q — евклидов шар радиусом $R > 0$). При этом функция (37) удовлетворяет обобщенному условию острого минимума (2) при $\alpha = 1$, $\Delta = \gamma R^2$ и $\overline{F} = \gamma R^2$.

Результаты для примера 6 представлены на рис. 8 и 9 для различных значений γ, R и r, где штриховая черная линия указывает на оценку (34), а синяя — на левую часть (29). Также показан характер поведения значений функции при работе сравниваемых методов. Подобное поведение функции при использовании [Julien, Schmidt, Bach, 2012] связано с негладкостью рассматриваемой задачи, в оценках (28) использование усреднения позволяет убрать такой эффект. На графиках справа из рис. 8 и 9 линии, которые расположены внизу правых графиков, соответствуют теоретическим оценкам для методов с шагом типа Б. Т. Поляка в § 3. Оценке (8) соответствует самая «нижняя» горизонтальная линия, а неадаптивной оценке (10) — параллельная ей пунктирная линия.



(a) (b)

Рис. 8. Результаты для примера 6, где $n = 1000$, $\gamma = 0{,}01$, $R = 1$ и $r = 0{,}1$



(a) (b)

Рис. 9. Результаты для примера 6, где $n = 1000$, $\gamma = 0{,}001$, $R = 5$ и $r = 0{,}5$

Из рис. 8 и 9 можно сделать такой вывод: что алгоритм (33) из [Julien, Schmidt, Bach, 2012] приводит к лучшей динамике значений минимизируемой целевой функции F, но мы можем наблюдать эффективность предлагаемой в настоящей статье методики (§ 3) в плане теоретической оценки с использованием Δ-острого минимума.

Замечание 2. Экспериментально проверялась также работа предлагаемых алгоритмов для целевой функции вида

$$f(x) = \min_{A \in S}\{\|X - A\|_2\} \quad \forall X \in Q = B_R(0), \tag{38}$$

где S — выпуклая оболочка m шаров с одним радиусом r, с центрами O_i, $i = 1, \ldots, m$, недалеко друг от друга, причем $S \subset Q$. При этом функция (38) 1-липшицева и удовлетворяет условию острого минимума (1) при $\alpha = 1$ и $f^* = 0$. При этом существенно, что направление спуска вычислялось неточно путем случайного выбора одного из центров O_i, то есть на итерациях использовались неточные δ-субградиенты. Адаптивность метода тут несущественна, поскольку норма δ-субградиента целевой функции равна 1 в точке, отличной от точного решения. Оказалось, что после нескольких итераций (в среднем 5 итераций) методы останавливаются и выдают точное решение (38).

Замечание 3. Экспериментально проверялась также работа предложенной методики для примера 1. Поскольку в данном случае известно, что $M = 1$, то нет смысла отдельно выделять адаптивный шаг. Расчеты выполнены при $\Delta = 0{,}5$, $n = 1000$ для $m = 20$ шаров Ω_i ($i = 1, \ldots, m$) радиусом $r = 10$. Центры шаров O_i различны и подобраны так, что расстояние от всех O_i до нуля $0 \in \mathbb{R}^n$ равно $\Delta + r$. По итогам проведенных расчетов за малое количество итераций (около 20) удалось найти лучшую (по сравнению со стартовой) точку, удаленную от всех шаров не более чем на 0,407, что несколько лучше Δ. Стоит отметить, что для данного примера была использована вариация шага Б. Т. Поляка (9) с выбором параметров для подзадач $\overline{f}_i \leqslant \Delta$ без требования $\overline{f}_i \geqslant f_i^*$, хотя теоретические результаты настоящей статьи получены при допущении $\overline{f} \geqslant f^*$.

8. Заключение

В статье рассмотрены некоторые подходы к методам для задач негладкой минимизации с оценками качества выдаваемого решения, инвариантными по размерности пространства, что потенциально интересно для задач в пространствах больших размерностей. Большая часть работы посвящена новым результатам по методам с «неточными» аналогами шага Б. Т. Поляка (7), (9) и (17), а также «неточным» аналогом условия острого минимума (2). Такой подход дает хорошие оценки скорости сходимости при достаточно малом $\Delta \geqslant 0$ как для выпуклых, так и для некоторых типов задач с релаксациями выпуклости. Приведено несколько примеров постановок задач, не обладающих обычным острым минимумом, но для которых может быть выполнено его обобщение (2). Тем самым предлагаемый подход несколько расширяет класс применимости методов с шагом Б. Т. Поляка и обычным острым минимумом [Поляк, 1969]. Более того, полученные в настоящей работе адаптивные и частично адаптивные оценки качества приближенного решения, выдаваемого предлагаемыми субградиентными методами, потенциально приводят к возможности использовать такие методы и полученные для них теоретические оценки для нелипшицевых задач (например, для минимизации локально липшицевых функций, удовлетворяющих условию Гёльдера) или для задач с неизвестной константой Липшица M. Однако в достаточно общей ситуации можно говорить лишь о применимости без гарантий оптимальности предложенного подхода, связанного с оценкой (8), что в общей ситуации представляется довольно ограничительным. Тем не менее мы приводим пример вычислительных экспериментов[1] для сильно выпуклой задачи с острым минимумом, где данная методика работает лучше по сравнению с известным методом [Julien, Schmidt, Bach, 2012]. Попутно также выведена оценка качества приближенного решения, выдаваемого субградиентным методом [Julien, Schmidt, Bach, 2012], с адаптивно подбираемыми параметрами. Выполнено экспериментальное сравнение указанных подходов для задачи о наименьшем покрывающем шаре. Оно показало, что метод с шагом Б. Т. Поляка может за малое число шагов приводить к относительно неплохому уровню точности для некоторых вариантов датасета (набора точек) в силу того, что учитывает геометрию задачи. Но при этом даже при специальных допущениях удается показать лишь условие Δ-острого минимума при некотором $\Delta > 0$, что не позволяет гарантировать глобальной сходимости метода со скоростью геометрической прогрессии. В плане возможных задач на будущее можно было бы отметить актуальность разработки рандомизированных вариантов исследованных в статье методов, а также более детальное исследование возможности замены предположения $\overline{f} \geqslant f^*$ менее жесткими

[1] Все эксперименты были реализованы на Python 3.4 и пакете numpy при помощи стандартного представления чисел в формате np.float64.

требованиями для теоретических оценок и практической реализации методов. Естественный интерес представляет исследование методики рестартов адаптивных субградиентных методов [Стонякин, 2020; Stonyakin et al., 2020] с использованием условия острого минимума.

Список литературы (References)

Нестеров Ю. Е. Эффективные методы нелинейного программирования. — М.: Радио и связь, 1989. — 301 с.

Nesterov Yu. E. Effektivnye metody nelineinogo programmirovaniya [Efficient Nonlinear Programming Methods]. — Moscow: Radio i Svyaz' Publ., 1989. — 301 p. (in Russian).

Нестеров Ю. Е. Методы выпуклой оптимизации. — М.: МЦНМО. 2010. — 281 с.

Nesterov Yu. E. Metody vypukloi optimizatsii [Convex optimization methods]. — Moscow: Publishing House MCNMO, 2010. — 281 p. (in Russian).

Поляк Б. Т. Минимизация негладких функционалов // Журн. вычисл. математики и мат. физики. — 1969. — Т. 9, № 3. — С. 509–521. — DOI: 10.1016/0041-5553(69)90061-5

Polyak B. T. Minimizatsiya negladkikh funktsionalov [Minimization of Nonsmooth Functionals] // Zhurn. vychisl. matematiki i mat. fiziki. — 1969. — Vol. 9, No. 3. — P. 509–521. — DOI: 10.1016/0041-5553(69)90061-5 (in Russian).

Поляк Б. Т. Введение в оптимизацию. — М.: Наука, 1983. — 384 с.

Polyak B. T. Introduction to optimization. — New York: Optimization software Inc., Publications Division, 1987. — xxvii+438 p. (Russ. ed.: *Polyak B. T.* Vvedenie v optimizatsiyu. — Moscow: Nauka, 1983. — 384 p.)

Стонякин Ф. С. Адаптивные зеркальные спуски для задач выпуклого программирования с использованием δ-субградиентов // arXiv preprint. — 2020. — https://arxiv.org/pdf/2012.12856

Stonyakin F. S. Adaptivnye zerkal'nye spuski dlya zadach vypuklogo programmirovaniya s ispol'zovaniem δ-subgradientov [Adaptive Mirror Descent Methods for Convex Programming Problems with *delta*-subgradients] // arXiv preprint. — 2020. — https://arxiv.org/pdf/2012.12856 (in Russian).

Стонякин Ф. С., Аблаев С. С., Баран И. В. Адаптивные методы градиентного типа для задач минимизации с относительной точностью и острым минимумом // Труды ИММ УрО РАН. — 2021. — Т. 27, № 4. — С. 175–188. — DOI: 10.21538/0134-4889-2021-27-4-175-188

Stonyakin F. S., Ablaev S. S., Baran I. V. Adaptivnye metody gradientnogo tipa dlya zadach minimizatsii s otnositel'noi tochnost'yu i ostrym minimumom [Adaptive gradient-type methods for optimization problems with relative error and sharp minimum] // Trudy IMM UrO RAN. — 2021. — Vol. 27, No. 4. — P. 175–188. — DOI: 10.21538/0134-4889-2021-27-4-175-188 (in Russian).

Devolder O., Glineur F., Nesterov Yu. First-order methods of smooth convex optimization with inexact oracle // Math. Programming. — 2014. — Vol. 146, No. 1. — P. 37–75. — DOI: 10.1007/s10107-013-0677-5

Hardt M., Ma T., Recht B. Gradient descent learns linear dynamical systems // J. Mach. Learn. Res. — 2018. — Vol. 19, No. 29. — 44 p.

Hinder O., Sidford A., Sohoni N. S. Near-optimal methods for minimizing star-convex functions and beyond // Proceedings of Machine Learning Research. — 2020. — Vol. 125. — P. 1894–1938.

Julien S. L., Schmidt M., Bach F. A simpler approach to obtaining an $O(1/t)$ convergence rate for the projected stochastic subgradient method // arXiv preprint. — 2012. — https://arxiv.org/pdf/1212.2002

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

Stonyakin F. S., Titov A. A., Makarenko D. V., Alkousa M. S. Some Methods for Relatively Strongly Monotone Variational Inequalities // arXiv preprint. — 2021. — https://arxiv.org/pdf/2109.03314