КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ И МОДЕЛИРОВАНИЕ 2024 Т. 16 № 7 С. 1765–1778
DOI: 10.20537/2076-7633-2024-16-7-1765-1778

Ки&М

СПЕЦИАЛЬНЫЙ ВЫПУСК

УДК: 519.8

Субградиентные методы для слабо выпуклых задач с острым минимумом в случае неточной информации о функции или субградиенте

Ф. С. Стонякин[1,2,3,a], Е. А. Лушко[1,b], И. Д. Третьяк[1,c], С. С. Аблаев[1,2,d]

[1]Московский физико-технический институт,
Россия, 141701, Московская обл., г. Долгопрудный, Институтский пер., д. 9
[2]Крымский федеральный университет им. В. И. Вернадского,
Россия, 295007, Республика Крым, г. Симферополь, проспект академика Вернадского, д. 4
[3]Университет Иннополис,
Россия, 420500, г. Иннополис, ул. Университетская, д. 1

E-mail: [a] fedyor@mail.ru, [b] lushko.ea@phystech.edu, [c] tretiak.id@phystech.edu, [d] seydamet.ablaev@yandex.ru

Получено 30.10.2024.
Принято к публикации 25.11.2024

Проблема разработки эффективных численных методов для невыпуклых (в том числе негладких) задач довольно актуальна в связи с широкой распространенностью таких задач в приложениях. Работа посвящена субградиентным методам для задач минимизации липшицевых μ-слабо выпуклых функций, причем не обязательно гладких. Хорошо известно, что для пространств большой размерности субградиентные методы имеют невысокие скоростные гарантии даже на классе выпуклых функций. При этом, если выделить подкласс функций, удовлетворяющих условию острого минимума, а также использовать шаг Поляка, можно гарантировать линейную скорость сходимости субградиентного метода. Однако возможны ситуации, когда значения функции или субградиента численному методу доступны лишь с некоторой погрешностью. В таком случае оценка качества выдаваемого этим численным методом приближенного решения может зависеть от величины погрешности. В настоящей статье для субградиентного метода с шагом Поляка исследованы ситуации, когда на итерациях используется неточная информация о значении целевой функции или субградиента. Доказано, что при определенном выборе начальной точки субградиентный метод с аналогом шага Поляка сходится со скоростью геометрической прогрессии на классе μ-слабо выпуклых функций с острым минимумом в случае аддитивной неточности в значениях субградиента. В случае когда как значение функции, так и значение ее субградиента в текущей точке известны с погрешностью, показана сходимость в некоторую окрестность множества точных решений и получены оценки качества выдаваемого решения субградиентным методом с соответствующим аналогом шага Поляка. Также в статье предложен субградиентный метод с клиппированным шагом и получена оценка качества выдаваемого им решения на классе μ-слабо выпуклых функций с острым минимумом. Проведены численные эксперименты для задачи восстановления матрицы малого ранга. Они показали, что эффективность исследуемых алгоритмов может не зависеть от точности локализации начального приближения внутри требуемой области, а неточность в значениях функции и субградиента может влиять на количество итераций, необходимых для достижения приемлемого качества решения, но почти не влияет на само качество решения.

Ключевые слова: субградиентный метод, адаптивный метод, шаг Поляка, слабо выпуклые функции, острый минимум, неточный субградиент

Работа выполнена при поддержке гранта Российского научного фонда № 24-21-00210 (https://rscf.ru/project/24-21-00210/).

COMPUTER RESEARCH AND MODELING
2024 VOL. 16 NO. 7 P. 1765–1778
DOI: 10.20537/2076-7633-2024-16-7-1765-1778

SPECIAL ISSUE

UDC: 519.8

Subgradient methods for weakly convex problems with a sharp minimum in the case of inexact information about the function or subgradient

F. S. Stonyakin[1,2,3,a], E. A. Lushko[1,b], I. D. Tretyak[1,c], S. S. Ablaev[1,2,d]

[1]Moscow Institute of Physics and Technology,
9 Institutskiy per., Dolgoprudny, Moscow region, 141701, Russia
[2]V. I. Vernadsky Crimean Federal University,
4 Academician Vernadsky ave., Simferopol, Republic of Crimea, 295007, Russia
[3]Innopolis University,
1 Universitetskaya st., Innopolis, 420500, Russia

E-mail: [a] fedyor@mail.ru, [b] lushko.ea@phystech.edu, [c] tretiak.id@phystech.edu, [d] seydamet.ablaev@yandex.ru

Received 30.10.2024.
Accepted for publication 25.11.2024

The problem of developing efficient numerical methods for non-convex (including non-smooth) problems is relevant due to their widespread use of such problems in applications. This paper is devoted to subgradient methods for minimizing Lipschitz μ-weakly convex functions, which are not necessarily smooth. It is well known that subgradient methods have low convergence rates in high-dimensional spaces even for convex functions. However, if we consider a subclass of functions that satisfies sharp minimum condition and also use the Polyak step, we can guarantee a linear convergence rate of the subgradient method. In some cases, the values of the function or it's subgradient may be available to the numerical method with some error. The accuracy of the solution provided by the numerical method depends on the magnitude of this error. In this paper, we investigate the behavior of the subgradient method with a Polyak step when inaccurate information about the objective function value or subgradient is used in iterations. We prove that with a specific choice of starting point, the subgradient method with some analogue of the Polyak step-size converges at a geometric progression rate on a class of μ-weakly convex functions with a sharp minimum, provided that there is additive inaccuracy in the subgradient values. In the case when both the value of the function and the value of its subgradient at the current point are known with error, convergence to some neighborhood of the set of exact solutions is shown and the quality estimates of the output solution by the subgradient method with the corresponding analogue of the Polyak step are obtained. The article also proposes a subgradient method with a clipped step, and an assessment of the quality of the solution obtained by this method for the class of μ-weakly convex functions with a sharp minimum is presented. Numerical experiments were conducted for the problem of low-rank matrix recovery. They showed that the efficiency of the studied algorithms may not depend on the accuracy of localization of the initial approximation within the required region, and the inaccuracy in the values of the function and subgradient may affect the number of iterations required to achieve an acceptable quality of the solution, but has almost no effect on the quality of the solution itself.

Keywords: subgradient method, adaptive method, Polyak step, weakly concave functions, sharp minimum, inaccurate subgradient

Citation: *Computer Research and Modeling*, 2024, vol. 16, no. 7, pp. 1765–1778 (Russian).
This work was supported by Russian Science Foundation, project 24-21-00210 (https://rscf.ru/project/24-21-00210/).

1. Введение

Задачи невыпуклой оптимизации возникают во многих приложениях. Вопросы разработки эффективных вычислительных процедур для невывыпуклых оптимизационных задач очень актуальны. С точки зрения приложений особый интерес представляет класс μ-слабо выпуклых функций. Напомним, что функция $f(x)$ называется μ-слабо выпуклой при $\mu \geqslant 0$, если функция $f(x) + \frac{\mu}{2}\|x\|_2^2$ является выпуклой. Слабо выпуклыми являются, например, задача восстановления фазы, задача восстановления малоранговых матриц в случае неевклидовой метрики, также слабо выпуклые функции встречаются в задачах анализа данных и машинного обучения [Davis, Drusvyatskiy, Paquette, 2020; Li et al., 2020].

Известно [Davis et al., 2018], что методы субградиентного типа сходятся со скоростью геометрической прогрессии на классе μ-слабо выпуклых функций с острым минимумом. Напомним, что функция $f(x)\colon Q \to \mathbb{R}$ имеет острый минимум в точке x_*, если $\forall x \in Q$ справедливо неравенство

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

где $\alpha > 0$, $Q \subseteq \mathbb{R}^n$, $f^* = f(x_*)$ — минимальное значение функции в искомой точке минимума, X_* — множество точек минимума, норма $\|\cdot\|_2$ евклидова.

На итерациях субградиентного метода с шагом Поляка используются значения функции и субградиента в текущих точках процесса. Однако возможна ситуация, когда значения функции или субградиента известны лишь приближенно. Настоящая работа посвящена развитию результатов работ [Davis et al., 2018; Стонякин и др., 2023], в которых получены оценки скорости сходимости субградиентных методов с вариациями шага Поляка на классе μ-слабо выпуклых функций с острым минимумом, с акцентом на случай неточной информации о функции или субградиенте. Таким образом, в статье исследуется ситуация, когда значения функции или субградиента, используемые на итерациях метода, известны с погрешностью.

Вопросы влияния неточности доступной информации на вычислительные гарантии оптимизационных методов достаточно актуальны [Stonyakin, Kuruzov, Polyak, 2023]. Известно, что неускоренные методы, по сравнению с ускоренными методами градиентного типа для класса выпуклых гладких функций, являются более устойчивыми к погрешностям, поскольку быстрые методы накапливают погрешность и из-за этого теряют асимптотическую сходимость. Для μ-сильно выпуклых достаточно гладких задач известна нижняя оценка точности решения задачи, которой можно достичь градиентным методом в случае использования Δ-неточного градиента: $f(x) - f^* = O\left(\frac{\Delta^2}{\mu}\right)$.

Будем рассматривать задачу минимизации вида

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

где Q — выпуклое замкнутое подмножество $\mathbb{R}^n$, $f\colon Q \to \mathbb{R}$ — μ-слабо выпуклая функция, удовлетворяющая условию острого минимума. Также предположим, что функция f липшицева, с константой $M > 0$, то есть для всяких $x, y \in Q$ справедливо неравенство

$$|f(x) - f(y)| \leqslant M\|x - y\|_2.$$

Как известно, слабо выпуклые функции локально липшицевы, поэтому в качестве субградиента можно использовать произвольный элемент субдифференциала Кларка:

$$\partial_{\mathrm{Cl}} f(x) = \operatorname{conv}\left\{v \in \mathbb{R}^n \mid \exists\{x_k\}\colon x_k \in T(f),\ x_k \to x,\ \nabla f(x_k) \to v\right\},$$

где $\operatorname{conv}\{\cdot\}$ обозначает выпуклую оболочку множества, $T(f)$ — множество точек дифференцируемости функции $f(x)$. Будем считать, что если функция $f(x)$ дифференцируема в точке x, то $\nabla f(x)$ — это ее градиент. Если функция $f(x)$ недифференцируема в точке x, то под субградиентом $\nabla f(x)$ будем понимать произвольный элемент субдифференциала Кларка функции f в точке x.

Напомним [Davis et al., 2018], что для μ-слабо выпуклой функции f справедливо субградиентное неравенство:

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

Работа состоит из введения, заключения и четырех основных разделов. В § 2 доказана линейная скорость сходимости субградиентного метода с аналогом шага Поляка для слабо выпуклых функций с острым минимумом в случае аддитивной неточности в значениях субградиента. В § 3, на классе слабо выпуклых функций с острым минимумом, исследован случай, когда значения и функции, и субградиента содержат аддитивную неточность. Для этой ситуации показано, что субградиентный метод с шагом Поляка сходится в некоторую окрестность множества точных решений, а также получена оценка качества выдаваемого методом решения. В § 4 предложен субградиентный метод с клиппированным шагом и получена оценка качества выдаваемого методом решения на выделенном классе задач. В § 5 проведены численные эксперименты, демонстрирующие работоспособность предложенных методов.

2. Субградиентный метод с шагом типа Поляка в случае аддитивной неточности значения субградиента

В данном разделе исследуется субградиентный метод с шагом Поляка для задачи минимизации μ-слабо выпуклой функции с острым минимумом при наличии аддитивной неточности в значениях субградиента.

Пусть в каждой (запрашиваемой на итерациях метода) точке x доступен аддитивно неточный субградиент $\widetilde{\nabla} f(x)$: $\|\widetilde{\nabla} f(x) - \nabla f(x)\|_2 \leqslant \Delta$ для некоторого достаточно малого $\Delta > 0$.

Для решения задачи (1) будем использовать субградиентный метод вида

$$x_{k+1} := \mathrm{Proj}_Q(x_k - h_k \widetilde{\nabla} f(x_k)) \tag{3}$$

со следующей вариацией шага Поляка:

$$h_k = \frac{f(x_k) - f^*}{\|\widetilde{\nabla} f(x_k)\|_2^2}. \tag{4}$$

Всюду далее будем считать, что $\widetilde{\nabla} f(x_k) \neq 0$ при $k \geqslant 0$ и $f(x_k) > f^*$. Если же существует k, при котором $\widetilde{\nabla} f(x_k) = 0$, то для этого k решение задачи (1) найдено. Аналогично при $f(x_k) = f^*$.

Теорема 1. *Пусть f — μ-слабо выпуклая липшицева (с константой $M > 0$) функция, удовлетворяющая условию острого минимума, а начальная точка x_0 такова, что $\min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha}{2\mu}$, а также справедливо неравенство $\alpha > 4\Delta$. Тогда субградиентный метод* (3) *с шагом* (4) *сходится со скоростью геометрической прогрессии*:

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2 - 4\Delta\alpha}{2M^2}\right)^{k+1} \min_{x_* \in X_*} \|x_0 - x_*\|_2^2. \tag{5}$$

Доказательство. Докажем теорему методом математической индукции. Базис индукции очевиден, так как $\min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha}{2\mu}$. Предположим, что справедливо неравенство

$$\min_{x_* \in X_*} \|x_k - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2 - 4\Delta\alpha}{2M^2}\right) \min_{x_* \in X_*} \|x_{k-1} - x_*\|_2^2.$$

Отметим, что из предположения индукции вытекает неравенство $\min_{x_* \in X_*} \|x_k - x_*\|_2 \leqslant \frac{\alpha}{2\mu}$. Покажем, что индукционный переход обеспечивает справедливость неравенства (5).

В случае неточного субградиента для слабо выпуклых функций из (2) следует следующее неравенство:

$$f^* - f(x_k) \geqslant \langle \nabla f(x_k),\ x_* - x_k \rangle - \frac{\mu}{2}\|x_k - x_*\|_2^2 \geqslant \langle \widetilde{\nabla} f(x_k),\ x_* - x_k \rangle - \Delta \|x_* - x_k\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2.$$

Следовательно, имеем неравенство

$$\langle \widetilde{\nabla} f(x_k),\ x_k - x_* \rangle \geqslant f(x_k) - f^* - \Delta \|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2. \tag{6}$$

Как известно, для всякого $k \geqslant 0$ и ближайшего к x_k точного решения $x_* \in X_*$ справедливо соотношение

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

Используя неравенство (6), получим оценку невязки по аргументу:

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 + h_k^2 \|\widetilde{\nabla} f(x_k)\|_2^2 - 2h_k \left(f(x_k) - f^* - \Delta \|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2 \right).$$

После преобразований имеем следующее неравенство:

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - \frac{\left(f(x_k) - f^*\right)\left(f(x_k) - f^* - 2\Delta \|x_k - x_*\|_2 - \mu \|x_k - x_*\|_2^2\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2}.$$

Для выполнения $\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2$ достаточно, чтобы было верно

$$f(x_k) - f^* - 2\Delta \|x_k - x_*\|_2 - \mu \|x_k - x_*\|_2^2 \geqslant 0.$$

Используя условие острого минимума и неравенство $\|x_k - x_*\|_2 \leqslant \frac{\alpha}{2\mu}$, а также предположение $\alpha > 4\Delta$, получим

$$f(x_k) - f^* - 2\Delta \|x_k - x_*\|_2 - \mu \|x_k - x_*\|_2^2 \geqslant \|x_k - x_*\|_2 \left(\frac{\alpha}{2} - 2\Delta \right) \geqslant 0.$$

Итак, после k итераций субградиентного метода с шагом Поляка справедливо неравенство

$$\|x_{k+1} - x_*\|_2^2 \leqslant \prod_{i=0}^{k} \left(1 - \frac{\frac{\alpha^2}{2} - 2\Delta\alpha}{\|\widetilde{\nabla} f(x_i)\|_2^2} \right) \|x_0 - x_*\|_2^2.$$

В силу липшицевости целевой функции нормы субградиентов равномерно ограничены, т. е. $\forall i = \overline{0, k}$ верно, что $\|\widetilde{\nabla} f(x_i)\|_2 \leqslant M$, следовательно,

$$1 - \frac{\frac{\alpha^2}{2} - 2\Delta\alpha}{\|\widetilde{\nabla} f(x_i)\|_2^2} \leqslant 1 - \frac{\frac{\alpha^2}{2} - 2\Delta\alpha}{M^2}.$$

Таким образом, после перехода по индукции справедливо неравенство

$$\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\alpha^2 - 4\Delta\alpha}{2M^2} \right)^{k+1} \|x_0 - x_*\|_2^2,$$

что указывает на сходимость со скоростью геометрической прогрессии. □

3. Субградиентный метод с шагом Поляка в случае аддитивной неточности значения целевой функции и субградиента

В этом разделе рассмотрим ситуацию, когда значения и целевой функции, и субградиента известны лишь приближенно.

Пусть в каждой (запрашиваемой на итерациях метода) точке x доступны аддитивно неточный субградиент $\widetilde{\nabla} f(x_k)$ $\left(\text{т. е. } \|\widetilde{\nabla} f(x_k) - \nabla f(x_k)\|_2 \leqslant \Delta\right)$ и аддитивно неточное значение функции $\widetilde{f}(x_k)$ $\left(\text{т. е. } |\widetilde{f}(x_k) - f(x_k)| \leqslant \delta\right)$ для некоторых достаточно малых $\Delta > 0$ и $\delta > 0$ соответственно.

Всюду далее будем считать, что $\delta \ll \alpha$.

Рассмотрим следующий вариант шага Поляка для случая неточной информации о функции и субградиенте:

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

Теорема 2. *Пусть f — μ-слабо выпуклая липшицева (с константой $M > 0$) функция, удовлетворяющая условию острого минимума, начальная точка x_0 такова, что $\min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$. Если известно, что $\alpha > \frac{8}{3}\Delta$, а также для произвольного x_k при $k \geqslant 0$ верно, что $f(x_k) \geqslant \widetilde{f}(x_k) > f^*$, то для метода* (3) *с шагом* (7) *справедливо неравенство $\min\limits_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \min\limits_{x_* \in X_*} \|x_k - x_*\|_2^2$, а также верна альтернатива*:

1) *либо*

$$\min_{x_* \in X_*} \|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\alpha\Delta}{\|\widetilde{\nabla} f(x_k)\|_2^2}\right) \min_{x_* \in X_*} \|x_k - x_*\|_2^2 + \frac{\frac{\alpha\delta}{4\mu}\left(\frac{3\alpha}{4} - 2\Delta\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2}; \tag{8}$$

2) *либо* $\min\limits_{x_* \in X_*} \|x_k - x_*\|_2 \leqslant \frac{\delta}{\alpha}$.

Доказательство. Докажем теорему методом математической индукции. Согласно предположению теоремы, $\min\limits_{x_* \in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$. Предположим, что утверждение теоремы верно при $k = i - 1$, и докажем его для $k = i$.

Аналог субградиентного неравенства для данного случая:

$$f^* - f(x_k) \geqslant \langle \nabla f(x_k), x_* - x_k \rangle - \frac{\mu}{2}\|x_k - x_*\|_2^2 \geqslant \langle \widetilde{\nabla} f(x_k), x_* - x_k \rangle - \Delta\|x_* - x_k\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2.$$

Следовательно, верно неравенство

$$\langle \widetilde{\nabla} f(x_k), x_k - x_* \rangle \geqslant f(x_k) - f^* - \Delta\|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2.$$

Для всякого $k \geqslant 0$ и ближайшего к x_k точного решения x_* имеют место неравенства

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - 2h_k\langle \widetilde{\nabla} f(x_k), x_k - x_* \rangle + h_k^2\|\widetilde{\nabla} f(x_k)\|_2^2 \leqslant$$
$$\leqslant \|x_k - x_*\|_2^2 + \frac{\left(\widetilde{f}(x_k) - f^*\right)^2}{\|\widetilde{\nabla} f(x_k)\|_2^2} - 2\frac{\widetilde{f}(x_k) - f^*}{\|\widetilde{\nabla} f(x_k)\|_2^2} \cdot \left(f(x_k) - f^* - \Delta\|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2\right).$$

После преобразований имеем следующее неравенство:

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - \frac{\left(\widetilde{f}(x_k) - f^*\right)\left(2f(x_k) - f^* - \widetilde{f}(x_k) - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2}.$$

Для выполнения $\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2$ достаточно потребовать, чтобы

$$\left(\widetilde{f}(x_k) - f^*\right)\left(2f(x_k) - f^* - \widetilde{f}(x_k) - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2\right) \geqslant 0. \tag{9}$$

Ясно, что

$$\widetilde{f}(x_k) - f^* \geqslant f(x_k) - f^* - \delta \geqslant \alpha\|x_k - x_*\|_2 - \delta.$$

Рассмотрим ситуацию $\|x_k - x_*\|_2 > \frac{\delta}{\alpha}$. В противном случае приемлемая точность решения уже достигнута. С учетом этого далее имеем

$$\begin{aligned}2f(x_k) - f^* - \widetilde{f}(x_k) - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 \geqslant f(x_k) - f^* - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 \geqslant \\ \geqslant \alpha\|x_k - x_*\|_2 - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 \geqslant \alpha\|x_k - x_*\|_2 - 2\Delta\|x_k - x_*\|_2 - \frac{\alpha}{4}\|x_k - x_*\|_2 = \\ = \left(\frac{3\alpha}{4} - 2\Delta\right)\|x_k - x_*\|_2.\end{aligned}$$

В силу неравенства $\alpha > \frac{8}{3}\Delta$ имеем, что $\frac{3\alpha}{4} - 2\Delta > 0$, откуда следует, что (9) неотрицательно. Следовательно, $\min\limits_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \min\limits_{x_*\in X_*}\|x_k - x_*\|_2^2$.

Справедливы неравенства

$$\begin{aligned}\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - \frac{(\alpha\|x_k - x_*\|_2 - \delta)\left(\frac{3\alpha}{4} - 2\Delta\right)\|x_k - x_*\|_2}{\|\widetilde{\nabla}f(x_k)\|_2^2} = \\ = \|x_k - x_*\|_2^2 - \frac{\left(\frac{3\alpha^2}{4} - 2\alpha\Delta\right)\|x_k - x_*\|_2^2 - \left(\frac{3\alpha\delta}{4} - 2\delta\Delta\right)\|x_k - x_*\|_2}{\|\widetilde{\nabla}f(x_k)\|_2^2} \leqslant \\ \leqslant \|x_k - x_*\|_2^2 - \frac{\left(\frac{3\alpha^2}{4} - 2\alpha\Delta\right)\|x_k - x_*\|_2^2 - \left(\frac{3\alpha\delta}{4} - 2\delta\Delta\right)\frac{\alpha}{4\mu}}{\|\widetilde{\nabla}f(x_k)\|_2^2}.\end{aligned}$$

После преобразований имеем

$$\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\alpha\Delta}{\|\widetilde{\nabla}f(x_k)\|_2^2}\right)\|x_k - x_*\|_2^2 + \frac{\frac{\alpha\delta}{4\mu}\left(\frac{3\alpha}{4} - 2\Delta\right)}{\|\widetilde{\nabla}f(x_k)\|_2^2}.$$

□

Теперь рассмотрим ситуацию, когда доступное методу приближенное значение целевой функции больше точного ее значения.

Теорема 3. *Пусть f — μ-слабо выпуклая липшицева (с константой $M > 0$) функция, удовлетворяющая условию острого минимума, пусть также справедливо неравенство $\alpha > \frac{8}{3}\Delta$. Если для произвольного x_k при $k \geqslant 0$ верно, что $f(x_k) < \widetilde{f}(x_k)$ и $\min\limits_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$, то для метода (3) с шагом (7) справедливо неравенство $\min\limits_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \min\limits_{x_*\in X_*}\|x_k - x_*\|_2^2$, а также верна альтернатива:*

1) *либо*

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\alpha\Delta}{\|\widetilde{\nabla}f(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\frac{\alpha^2\delta}{2\mu}}{\|\widetilde{\nabla}f(x_k)\|_2^2}; \tag{10}$$

2) *либо* $\min\limits_{x_*\in X_*}\|x_k - x_*\|_2 < \frac{8\delta}{3\alpha - 8\Delta}$.

Доказательство. Докажем теорему методом математической индукции. Согласно предположению теоремы, $\min\limits_{x_*\in X_*} \|x_0 - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$. Предположим, что утверждение теоремы верно при $k = i - 1$, и докажем его для $k = i$.

Имеем следующие неравенства:

$$f^* - f(x_k) \geqslant \langle \nabla f(x_k),\, x_* - x_k\rangle - \frac{\mu}{2}\|x_k - x_*\|_2^2 \geqslant \langle \widetilde{\nabla} f(x_k), x_* - x_k\rangle - \Delta \cdot \|x_* - x_k\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2,$$

$$\langle \widetilde{\nabla} f(x_k),\, x_k - x_*\rangle \geqslant f(x_k) - f^* - \Delta\|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2 \geqslant$$
$$\geqslant \widetilde{f}(x_k) - f^* - \Delta \cdot \|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2 - \delta.$$

Для всякого $k \geqslant 0$ и ближайшего к x_k точного решения x_* справедливы неравенства

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

$$\leqslant \|x_k - x_*\|_2^2 + \frac{\left(\widetilde{f}(x_k) - f^*\right)^2}{\|\widetilde{\nabla} f(x_k)\|_2^2} - 2 \cdot \frac{(\widetilde{f}(x_k) - f^*)}{\|\widetilde{\nabla} f(x_k)\|_2^2} \cdot \left(\widetilde{f}(x_k) - f^* - \Delta\|x_k - x_*\|_2 - \frac{\mu}{2}\|x_k - x_*\|_2^2 - \delta\right) \leqslant$$

$$\leqslant \|x_k - x_*\|_2^2 - \frac{\left(\widetilde{f}(x_k) - f^*\right)\left(\widetilde{f}(x_k) - f^* - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 - 2\delta\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2}.$$

Для выполнения $\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2$ достаточно потребовать, что

$$\left(\widetilde{f}(x_k) - f^*\right)\left(\widetilde{f}(x_k) - f^* - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 - 2\delta\right) \geqslant 0.$$

Ясно, что $\widetilde{f}(x_k) - f^* > f(x_k) - f^* \geqslant \alpha\|x_k - x_*\|_2 > 0$. Осталось показать справедливость неравенства

$$\widetilde{f}(x_k) - f^* - 2\Delta\|x_k - x_*\|_2 - \mu\|x_k - x_*\|_2^2 - 2\delta \geqslant 0. \tag{11}$$

Применив условие острого минимума и неравенство $\|x_k - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$, получим

$$\alpha\|x_k - x_*\|_2 - 2\delta - 2\Delta\|x_k - x_*\|_2 - \frac{\alpha}{4}\|x_k - x_*\|_2 \geqslant 0. \tag{12}$$

Если справедливо неравенство (12), то справедливо и (11).

Оценим значения параметров, при которых справедливо (12). После преобразований получим

$$\|x_k - x_*\|_2 \geqslant \frac{8\delta}{3\alpha - 8\Delta}.$$

Тогда справедливо либо неравенство (12), либо

$$\|x_k - x_*\|_2 < \frac{8\delta}{3\alpha - 8\Delta},$$

а это означает, что выполнен п. 2 доказываемой теоремы.

Справедливы соотношения

$$\|x_{k+1} - x_*\|_2^2 \leqslant \|x_k - x_*\|_2^2 - \frac{\alpha\|x_k - x_*\|_2\left(\left(\frac{3\alpha}{4} - 2\Delta\right)\|x_k - x_*\|_2 - 2\delta\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2} =$$

$$= \|x_k - x_*\|_2^2 - \frac{\left(\frac{3\alpha^2}{4} - 2\alpha\Delta\right)\|x_k - x_*\|_2^2 - 2\alpha\delta\|x_k - x_*\|_2}{\|\widetilde{\nabla} f(x_k)\|_2^2}.$$

Итак, имеем оценку

$$\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\alpha\Delta}{\|\widetilde{\nabla} f(x_k)\|_2^2}\right)\|x_k - x_*\|_2^2 + \frac{\frac{\alpha^2\delta}{2\mu}}{\|\widetilde{\nabla} f(x_k)\|_2^2}.$$

Теорема доказана. □

4. Вариант субградиентного метода с клиппированным шагом типа Поляка

Мы показали, что на очередном шаге субградиентного метода (3) справедливо либо одно из неравенств (8), (10):

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\alpha\Delta}{\|\widetilde{\nabla} f(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\frac{\alpha\delta}{4\mu}\left(\frac{3\alpha}{4} - 2\Delta\right)}{\|\widetilde{\nabla} f(x_k)\|_2^2},$$

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{\frac{3\alpha^2}{4} - 2\Delta\alpha}{\|\widetilde{\nabla} f(x_k)\|_2^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\frac{\alpha^2\delta}{2\mu}}{\|\widetilde{\nabla} f(x_k)\|_2^2},$$

либо неравенство

$$\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \max\left\{\frac{\delta}{\alpha}, \frac{8\delta}{3\alpha - 8\Delta}\right\}.$$

Теперь покажем, как изменятся оценки скорости сходимости субградиентного метода (3) из теорем 2 и 3 в случае, когда функция f липшицева, с константой M, и используется аналог шага Поляка вида (предполагаем, что $\widetilde{f}(x_k) > f(x_k)$ для всякого x_k, иначе x_k — искомое приближенное решение поставленной задачи)

$$h_k = \frac{\widetilde{f}(x_k) - f^*}{M^2}. \tag{13}$$

Утверждение 1. *Если в условиях теоремы* 2 *дополнительно предположить, что функция* $f(x)$ *липшицева с константой* M *и для метода* (3) *используется шаг* (13), *то неравенство* (8) *принимает вид*

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\alpha\delta(3\alpha - 8\Delta)}{16\mu M^2}.$$

Утверждение 2. *Если в условиях теоремы* 3 *дополнительно предположить, что функция* $f(x)$ *липшицева с константой* M *и для метода* (3) *используется шаг* (13), *то неравенство* (10) *принимает вид*

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\alpha^2\delta}{2\mu M^2}.$$

Будем регулировать шаг h_k субградиентного метода (3) по норме субградиента целевой функции в текущей точке.

Алгоритм 1. Субградиентный метод с клиппированным шагом

Require: $x_0\colon \min\limits_{x_*\in X_*}\|x_0 - x_*\|_2 \leqslant \frac{\alpha}{4\mu}$, $\delta > 0$, $M > 0$, $\Delta > 0$.

$I := \varnothing$, $J := \varnothing$.

if $\|\widetilde{\nabla} f(x_k)\|_2 \leqslant M\sqrt[4]{\delta}$ **then**

$$x_{k+1} := \mathrm{Proj}_Q\left(x_k - \frac{\widetilde{f}(x_k) - f^*}{M^2}\widetilde{\nabla} f(x_k)\right),$$

  $k \to I$

else

  if $\|\widetilde{\nabla} f(x_k)\|_2 > M\sqrt[4]{\delta}$ **then**

$$x_{k+1} := \mathrm{Proj}_Q\left(x_k - \frac{\widetilde{f}(x_k) - f^*}{\|\widetilde{\nabla} f(x_k)\|_2^2}\widetilde{\nabla} f(x_k)\right).$$

    $k \to J$

  end if

end if

Теорема 4. *Пусть f — μ-слабо выпуклая липшицева (с константой $M > 0$) функция, удовлетворяющая условию острого минимума, известно, что $\alpha > \frac{8}{3}\Delta$, а начальная точка x_0 такова, что*

$$\min_{x_*\in X_*}\|x_0 - x_*\|_2 \leqslant \frac{\alpha}{4\mu}.$$

Тогда для алгоритма 1 *справедлива альтернатива*:

1) *либо*

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + C(\delta, \mu),$$

где $C(\delta, \mu) = \frac{2\alpha\sqrt{\delta}}{\mu(3\alpha - 8\Delta)}$;

2) *либо*

$$\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \max\left\{\frac{\delta}{\alpha}, \frac{8\delta}{3\alpha - 8\Delta}\right\}.$$

Доказательство. Если $\|\widetilde{\nabla} f(x_k)\|_2 \leqslant M\sqrt[4]{\delta}$, то либо

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\alpha^2\delta}{2\mu M^2}, \tag{14}$$

либо

$$\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \max\left\{\frac{\delta}{\alpha}, \frac{8\delta}{3\alpha - 8\Delta}\right\}.$$

Если $\|\widetilde{\nabla} f(x_k)\|_2 > M\sqrt[4]{\delta}$, то либо

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)\min_{x_*\in X_*}\|x_k - x_*\|_2^2 + \frac{\alpha^2\sqrt{\delta}}{2\mu M^2}, \tag{15}$$

либо

$$\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \max\left\{\frac{\delta}{\alpha},\ \frac{8\delta}{3\alpha - 8\Delta}\right\}.$$

По рекурсии из неравенств (14) и (15) получим соответственно

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + \frac{\alpha^2\delta}{2\mu M^2}\sum_{i=0}^{k}\left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^i,$$

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + \frac{\alpha^2\sqrt{\delta}}{2\mu M^2}\sum_{i=0}^{k}\left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^i.$$

Вычислим сумму бесконечно убывающей геометрической прогрессии:

$$\sum_{i=0}^{k}\left(1 - \frac{3\alpha^2 - 8\Delta\alpha}{4M^2}\right)^i = \frac{4M^2}{3\alpha^2 - 8\Delta\alpha}.$$

Следовательно, справедливы неравенства

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + \frac{2\alpha\delta}{\mu(3\alpha - 8\Delta)},$$

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + \frac{2\alpha\sqrt{\delta}}{\mu(3\alpha - 8\Delta)}.$$

Данные неравенства указывают на сходимость со скоростью геометрической прогрессии в $O\left(\frac{\sqrt{\delta}}{\mu}\right)$-окрестность множества точек минимума.

Учитывая, что при малых δ верно неравенство $\delta < \sqrt{\delta}$, имеем, что либо

$$\min_{x_*\in X_*}\|x_{k+1} - x_*\|_2^2 \leqslant \left(1 - \frac{3\alpha^2 - 8\alpha\Delta}{4M^2}\right)^{k+1}\min_{x_*\in X_*}\|x_0 - x_*\|_2^2 + \frac{2\alpha\sqrt{\delta}}{\mu(3\alpha - 8\Delta)},$$

либо

$$\min_{x_*\in X_*}\|x_k - x_*\|_2 \leqslant \max\left\{\frac{\delta}{\alpha},\ \frac{8\delta}{3\alpha - 8\Delta}\right\}.$$

□

5. Результаты вычислительных экспериментов

В этом разделе опишем выполненные численные эксперименты для иллюстрации сходимости субградиентного метода с использованием неточной информации о целевой функции и субградиенте на примере задачи восстановления малоранговой матрицы. Все расчеты проводились с использованием 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.

В работе [Li et al., 2020] рассматривается задача восстановления малоранговой матрицы $X^* \in \mathbb{R}^{m\times n}$ по линейным измерениям, которые могут быть доступны с искажениями (вбросами). Модель измерений имеет вид

$$y = \mathcal{A}(X^*) + s,$$

где $y \in \mathbb{R}^d$ — вектор измерений, $s \in \mathbb{R}^d$ — вектор шума, имеющий значения произвольной величины в $\Omega \subset \{1, \ldots, d\}$ позициях и нули — в остальных, $\mathcal{A}\colon \mathbb{R}^{m\times n} \to \mathbb{R}^d$ — линейный оператор, сожержащий информацию о результах измерений. Долю выбросов будем обозначать как $p = \frac{|\Omega|}{d}$. Если матрица X^* имеет ранг r, то она допускает скелетное разложение $X^* = UV^T$, где $U \in \mathbb{R}^{m\times r}$, $V \in \mathbb{R}^{n\times r}$.

В общем случае в рассматриваемой работе предлагается поставить задачу в виде

$$\min_{U\in\mathbb{R}^{m\times r},\, V\in\mathbb{R}^{n\times r}} \frac{1}{d}\left\|y - \mathcal{A}\left(UV^T\right)\right\|_1 + \lambda\left\|U^TU - V^TV\right\|_F = \min_{U\in\mathbb{R}^{m\times r},\, V\in\mathbb{R}^{n\times r}} g(U, V),$$

где $\lambda > 0$ — параметр регуляризации, добавленный для учета двойственности в понимании факторов. Такая постановка позволяет сделать оптимизационную задачу устойчивой к рассматриваемому шуму почти до $\frac{1}{2}$ доли выбросов. В случае симметричной положительно полуопределенной матрицы $X^* \in \mathbb{R}^{n\times n}$ задача формулируется несколько проще:

$$\min_{U\in\mathbb{R}^{n\times r}} \frac{1}{d}\left\|y - \mathcal{A}\left(UU^T\right)\right\|_1 = \min_{U\in\mathbb{R}^{n\times r}} f(U).$$

Известно [Li et al., 2020], что данная задача в случае $\frac{l_1}{l_2}$-RIP-свойства оператора $\mathcal{A}$ [Li et al., 2020] будет μ-слабо выпуклой, с α-острым минимумом. Например, для операторов измерений, задающихся случайными гауссовыми матрицами $\mathcal{A}(X^*) = \left[\left\langle A_1, X^*\right\rangle, \ldots, \left\langle A_d, X^*\right\rangle\right]$, где $\langle\,,\,\rangle$ — фробениусово скалярное произведение, а элементы матриц A_k являются независимыми одинаково распределенными стандартными гауссовыми величинами, оператор $\mathcal{A}$ будет обладать этим свойством с высокой вероятностью. Кроме того, любой глобальный минимум будет приводить к точному восстановлению матрицы X^* даже в случае, когда доля выбросов близка к $\frac{1}{2}$. Необходимая точность инициализации U^0, V^0 или U^0 в симметричном положительно полуопределенном случае может быть достигнута за счет применения спектрального метода [Li et al., 2020].

Представим результаты реализации рассмотренных в настоящей статье алгоритмов для данной задачи в случае, когда оператор измерений $\mathcal{A}$ задается случайными матрицами так, как описано выше. В качестве матрицы X^* мы выбираем симметричную положительно полуопределенную $X^* = U^*U^{*T}$, где $U^* \in \mathbb{R}^{n\times r}$ с элементами $U^*_{ij} \sim N(0, 1)$. Вектор шума генерируем для различных долей выбросов p, выбирая dp позиций и генерируя в них случайное число, имеющее гауссово распределение с дисперсией 100 и нулевым математическим ожиданием.

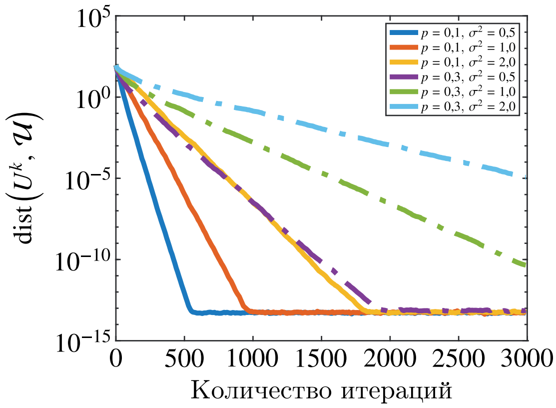


Рис. 1. Зависимость расстояния до минимума на k-м шаге метода для различных значений относительной доли выбросов p и дисперсии шума в субградиенте σ^2

На рис. 1 проиллюстрирована зависимость расстояния до ближайшей точки минимума $\mathrm{dist}\left(U^k, \mathcal{U}\right) = \min\limits_{U\in\mathcal{U}}\left\|U^k - U\right\|_F$ от количества итераций для шага Поляка в случае аддитивной неточности в субградиенте

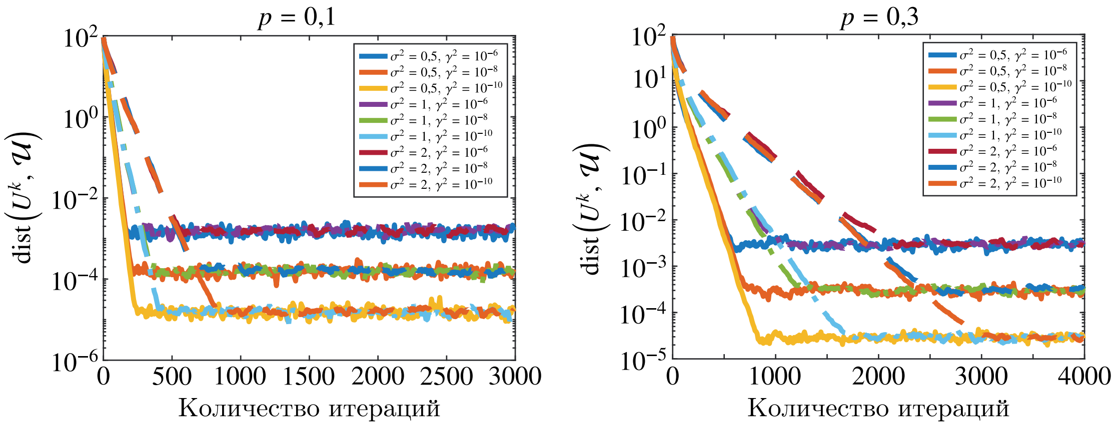
$$\widetilde{\nabla} f(U) = \nabla f(U) + W, \quad \nabla f(U) \in \partial f(U),$$

где $W \in \mathbb{R}^{n \times r}$ — матрицы шума в субградиенте с элементами $W_{ij} \sim N\left(0, \sigma^2\right)$. Приводятся графики сходимости для различного значения дисперсии $\sigma^2 \in \{0{,}5,\ 1{,}0,\ 1{,}5,\ 2{,}0\}$. Доля выбросов в векторе измерений — $p \in \{0{,}1,\ 0{,}3\}$. Размерность задачи — $n = 50$, целевой ранг — $r = 5$, количество измерений — $d = 5nr$. Скорость сходимости для всех экспериментов остается линейной. Наличие шума со все большей дисперсией требует лишь большего количества итераций за счет уменьшения коэффициента геометрической прогрессии.

В эксперименте с аддитивным шумом не только для субградиента, но и для значения целевой функции (рис. 2) выбирается значение дисперсии $\gamma^2 \in \left\{10^{-6},\ 10^{-8},\ 10^{-10}\right\}$, дисперсия шума в субградиенте варьируется аналогично, доля выбросов фиксируется равной $p = 0{,}1$ для верхнего графика и $p = 0{,}3$ — для нижнего,

$$\widetilde{f}(U) = f(U) + \delta, \quad \delta \sim N\left(0, \gamma^2\right). \tag{16}$$

В данном случае можно наблюдать сходимость решения в стационарное состояние, расстояние от которого до минимума зависит от величины аддитивного шума в значении целевой функции. Наблюдения показали, что точность этого стационарного состояния не зависит от точности локализации начального приближения внутри требуемой области (как это предполагалось в теоретических результатах о качестве выдаваемого решения). Наличие неточностей в измерениях шума в субградиенте влияет на требуемое число итераций (коэффициент линейной сходимости), но почти не влияет на стационарное состояние (достигаемого траекторией метода).



(а) (б)

Рис. 2. Зависимость расстояния до минимума на k-м шаге метода для различных значений относительной доли выбросов p и дисперсии шума в субградиенте σ^2 и значений функции γ^2

6. Заключение

В работе получены новые результаты о сходимости субградиентных методов с шагом Поляка на классе слабо выпуклых функций с острым минимумом. Детально исследована ситуация, когда значения функции или ее субградиента, используемые на итерациях метода, известны лишь приближенно.

Во-первых, доказано, что в случае, когда значения субградиента известны лишь приближенно, при некоторых ограничениях на выбор начальной точки, субградиентный метод с аналогом шага Поляка сходится со скоростью геометрической прогрессии.

Во-вторых, в случае, когда значения и функции, и субградиента известны с погрешностью, показана сходимость субградиентного метода с вариацией шага Поляка в некоторую окрестность множества точных решений и получены оценки качества выдаваемого субградиентным методом решения. В этой ситуации важно, какое из условий выполнено: $f(x_k) < \widetilde{f}(x_k)$ или $f(x_k) \geqslant \widetilde{f}(x_k)$. В работе исследованы оба случая.

В-третьих, предложен субградиентный метод с клиппированным шагом Поляка, показана его сходимость в некоторую окрестность множества точных решений и получена оценка качества выдаваемого методом решения.

Наконец, проведены численные эксперименты для задачи восстановления малоранговой матрицы, которые показали, что эффективность исследуемых алгоритмов может не зависеть от точности локализации начального приближения внутри требуемой области, а неточность в значениях функции и субградиента может влиять на количество итераций, но почти не влиять при этом на качество решения выдаваемого методом решения.

Список литературы (References)

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

Dudov S. I., Osipcev M. A. Harakterizaciya resheniya zadach sil'no-slabo vypuklogo programmirovaniya [Characterization of the solution for the problems of strong and weak convex programming] // Sbornik: Mathematics. — 2021. — Vol. 212, No. 6. — P. 43–72.

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

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

Davis D., Drusvyatskiy D., MacPhee K. J., 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.

Li X., Zhu Z., Man-Cho So A., Vidal R. Nonconvex robust low-rank matrix recovery // SIAM Journal of Optimization. — 2020. — Vol. 30, No. 1. — P. 660–686.

Stonyakin F., Kuruzov I., Polyak B. Stopping rules for gradient methods for non-convex problems with additive noise in gradient // Journal of Optimization Theory and Applications. — 2023. — Vol. 198. — P. 531—551.

Zhang J., He T., Sra S., Jadbabaie A. Why gradient clipping accelerates training: A theoretical justification for adaptivity // arXiv preprint. — 2019. — https://arxiv.org/pdf/1905.11881