ТРУДЫ ИНСТИТУТА МАТЕМАТИКИ И МЕХАНИКИ УрО РАН

Том 26 № 3 2020

УДК 519.85

О НЕКОТОРЫХ АЛГОРИТМАХ ДЛЯ УСЛОВНЫХ ЗАДАЧ ОПТИМИЗАЦИИ С ОТНОСИТЕЛЬНОЙ ТОЧНОСТЬЮ ПО ЦЕЛЕВОМУ ФУНКЦИОНАЛУ[1]

Ф. С. Стонякин, И. В. Баран

В настоящей работе получены оценки скорости сходимости некоторых субградиентных методов для задачи минимизации негладкого выпуклого липшицева однородного функционала с относительной точностью по целевому функционалу при наличии функциональных ограничений. К таким задачам предлагается применять аналоги известных субградиентных схем с переключениями. Это позволяет рассматривать и некоторые классы не обязательно выпуклых функционалов ограничений. Получена оценка скорости сходимости адаптивного зеркального спуска с переключениями на классе слабо α-квазивыпуклых целевых функционалов и функционалов ограничений. Обоснована оценка скорости сходимости предложенного субградиентного метода с переключениями с относительной точностью по целевому функционалу для задач минимизации выпуклого однородного целевого функционала со слабо α-квазивыпуклым функционалом ограничения. Рассмотрен также метод для задач минимизации выпуклого однородного липшицева функционала с унимодальным липшицевым функционалом ограничения и выведена оценка его скорости сходимости. Доказанные оценки скорости сходимости указывают на оптимальность предложенных алгоритмических процедур с точки зрения теории нижних оракульных оценок.

Ключевые слова: относительная точность, выпуклый однородный функционал, слабо α-квазивыпуклый функционал, зеркальный спуск, липшицев функционал, унимодальный функционал.

F. S. Stonyakin, I. V. Baran. On some algorithms for constrained optimization problems with relative accuracy with respect to the objective functional.

Convergence rate estimates are derived for some subgradient methods for the problem of minimization of a nonsmooth convex Lipschitz homogeneous functional with relative accuracy with respect to the objective functional under functional constraints. It is proposed to apply analogs of known switching subgradient schemes to such problems, which allows us to consider some classes of nonconvex problems as well. A convergence rate estimate is obtained for the adaptive mirror descent with switchings on the class of weakly α-quasiconvex objective functionals and constraint functionals. A convergence rate estimate of a proposed subgradient method with switchings with relative accuracy with respect to the objective functional is proved for problems of minimization of a convex homogeneous objective functional with a weakly α-quasiconvex constraint functional. We also consider a method for problems of minimization of a convex homogeneous Lipschitz functional with unimodal Lipschitz constraint functional and derive an estimate for its convergence rate. All convergence rate estimates proved in the paper show the optimality of the proposed algorithmic procedures from the viewpoint of the theory of lower oracle bounds.

Keywords: relative accuracy, convex homogeneous functional, weakly α-quasiconvex functional, mirror descent, Lipschitz-continuous functional, unimodal functional.

MSC: 90C25, 90C06, 49J52

DOI: 10.21538/0134-4889-2020-26-3-198-210

1. Введение

Предлагаемая работа посвящена некоторым субградиентным методам для минимизации, вообще говоря, негладкого выпуклого липшицева однородного функционала с относительной точностью при наличии одного или нескольких ограничений-неравенств, заданных выпуклыми функционалами. Указанная постановка задачи в значительной степени восходит к работам [1; 2] (см. также [3, гл. 6]). Как показано в данных источниках, подход к оцениванию

[1]Исследования Ф. С. Стонякина по разработке алгоритма 1 и доказательству теорем 1, 2 и 3, а также исследования И. В. Баран по разработке алгоритма 2 выполнены при поддержке гранта Президента Российской Федерации для государственной поддержки молодых российских ученых — кандидатов наук, код МК-15.2020.1.

качества решения задачи с точки зрения именно относительной точности вполне оправдан для разных прикладных задач (линейное программирование, проектирование механических конструкций и др.), если нет необходимости искать слишком точное решение. В настоящей статье рассматривается задача минимизации на выпуклом замкнутом подмножестве $X \subset \mathbb{R}^n$ выпуклого однородного функционала

$$f(x) \to \min_{x \in X} \tag{1.1}$$

с функционалами ограничений $g_p(x) \leqslant 0$ ($p = \overline{1,m}$ для некоторого натурального m). Стандартно будем обозначать $g(x) := \max\limits_{1 \leqslant p \leqslant m} \{g_p(x)\}$ и полагать далее, что ограничение одно.

Уточним, что найденное в допустимой точке x значение $f(x)$ приближает минимальное значение $f^* > 0$ с *относительной точностью* $\delta > 0$, если $f(x) < (1+\delta) f^*$. Всюду далее для задач минимизации с относительной точностью будем допускать, что задача разрешима и $f^* > 0$. Известно (см. [3, гл. 6]), что достаточно широкий класс задач оптимизации с относительной точностью можно сводить к минимизации выпуклой однородной функции. Например, можно рассмотреть такую постановку:

$$\text{найти} \quad f^* = \min_x \{f(x) \colon x \in \mathcal{L}\}, \quad \mathcal{L} = \{x \in \mathbb{R}^n \colon Cx = b\}, \tag{1.2}$$

где выпуклая функция $f(x)$ является однородной степени 1, C — $p \times n$-матрица (без ограничения общности можно считать, что матрица C имеет полный строчный ранг) и ненулевой вектор b. Основное предположение (см. [3]) о задаче (1.2) выглядит следующим образом:

$$\mathrm{dom} f \equiv \mathbb{R}^n, \quad 0 \in \mathrm{int} \partial f(0). \tag{1.3}$$

Другими словами, имеется в виду, что f — опорный функционал некоторого выпуклого компактного множества, у которого 0 — внутренняя точка. Поэтому при условии (1.3) искомое минимальное значение f^* положительно и задача отыскания решения задачи (1.2) с некоторой относительной точностью поставлена корректно. Заметим при этом, что любая задача безусловной минимизации

$$\min_{y \in \mathbb{R}^{n-1}} \phi(y)$$

с выпуклой целевой функцией $\phi(\cdot)$ может быть переписана в форме (1.2) следующим образом:

$$x = (y, \tau) \in \mathbb{R}^{n-1} \times \mathbb{R}^1_+, \quad f(x) = \tau \phi(y/\tau), \quad Cx \equiv \tau.$$

Стоит однако отметить, что в общем случае нет возможности гарантировать, что f удовлетворяет (1.3).

Для выделенного класса задач (1.1) в настоящей статье рассматриваются два метода, для которых удалось вывести оценки скорости сходимости, указывающие на оптимальность соответствующих алгоритмических процедур с точки зрения теории нижних оракульных оценок [4]. Рассмотренные методы (алгоритмы 1 и 2) основаны на использовании аналогов субградиентных схем с переключениями (см. [5;6]) (в случае неевклидовых расстояний — зеркальных спусков [7]). Однако при этом применяется похожая на [8–10] методика выбора шагов и критериев остановки методов, но с учетом специфики рассматриваемых в данной статье задач. Более того, оказывается возможным существенное расширение класса функционалов ограничений с сохранением оптимальных оценок скорости сходимости (сложности). Для первого из предложенных методов можно утверждать сохранение оптимальных с точностью до умножения на постоянный множитель оценок скорости сходимости (сложности) на классе слабо α-квазивыпуклых ($\alpha \in (0;1]$) функционалов ограничения (которые, вообще говоря, не выпуклы), удовлетворяющих условию Липшица. Отметим, что слабо α-квазивыпуклые функционалы естественно возникают во многих задачах (см., например, [11, замечание 2.1], а также работы [12; 13] и имеющиеся в них ссылки на источники). Получена также оценка скорости сходимости другого метода зеркального спуска (алгоритм 2) на классе выпуклых однородных задач с унимодальным M_g-липшицевым функционалом ограничения.

Выделим основные результаты (вклад) настоящей работы.

— Предложен метод зеркального спуска с переключениями по продуктивным и непродуктивным шагам (алгоритм 1) для задачи минимизации слабо α-квазивыпуклого целевого функционала со слабо α-квазивыпуклым функционалом ограничения и получена оценка его скорости сходимости с абсолютной точностью (теорема 1).

— Доказана оценка скорости сходимости алгоритма 1 для задачи минимизации выпуклого однородного липшицева функционала с относительной точностью при наличии слабо α-квазивыпуклого липшицева функционала ограничения (теорема 2).

— Получена оценка скорости сходимости другого алгоритма зеркального спуска с переключениями для задачи минимизации выпуклого однородного липшицева функционала с относительной точностью, применимая в случае M_g-липшицева унимодального функционала ограничения (теорема 3).

Всюду далее под $\langle\cdot,\cdot\rangle$ понимается скалярное произведение векторов.

2. Адаптивный субградиентный метод для минимизации негладких слабо α-квазивыпуклых функционалов с ограничениями

В этом разделе мы рассмотрим вспомогательный результат о зеркальных спусках для задач вида

$$f(x) \longrightarrow \min_{x\in X}, \quad g(x) \leqslant 0 \tag{2.1}$$

на некотором классе не обязательно выпуклых функционалов f и g. Пусть имеется некоторое число $\alpha \in (0;1]$, x_* — точка минимума задачи (2.1). Напомним, что функционал f называется *слабо α-квазивыпуклым* относительно x_* на множестве X, если для произвольного $x \in X$ выполнено неравенство

$$f(x_*) \geqslant f(x) + \frac{1}{\alpha}\langle\widetilde{\nabla} f(x),\, x_* - x\rangle, \tag{2.2}$$

где $\widetilde{\nabla} f(x)$ — произвольный субградиент f в точке x (в гладком случае — обычный градиент). Под субградиентом мы здесь и всюду далее пониманием элемент субдифференциала Кларка f в точке x и предполагаем его существование. Это вполне естественно для липшицевых функционалов (для существования субдифференциала Кларка в точке достаточно локальной липшицевости f в окрестности этой точки). Если функционал f выпуклый, то субдифференциал Кларка совпадает с обычным субдифференциалом в смысле выпуклого анализа, и в таком случае условие слабой α-квазивыпуклости (2.2) верно при $\alpha = 1$.

Ясно, что если неравенство (2.2) верно для некоторого $\alpha = \alpha_0 \in (0;1]$, то оно верно и при $\alpha \in (0;\alpha_0]$. Примеры функционалов, для которых возможно проверить свойство слабой α-квазивыпуклости и оценить параметр α, приведены, в частности, в [13]. Отметим также, что субградиент f может быть нулевым только в точке минимума: равенство $\widetilde{\nabla} f(x) = 0$ влечет $f(x) \leqslant f(x_*)$, что автоматически означает $f(x) = f(x_*)$.

Для дальнейших рассуждений нам потребуются некоторые вспомогательные понятия. Введем так называемую *прокс-функцию* $d\colon X \to \mathbb{R}$, обладающую свойством непрерывной дифференцируемости и 1-сильной выпуклости относительно нормы $\|\cdot\|$, т. е.

$$\langle\nabla d(x) - \nabla d(y),\, x - y\rangle \geqslant \|x - y\|^2 \quad \forall x, y \in X,$$

и предположим, что $\min\limits_{x\in X} d(x) = d(0)$. Будем полагать, что существует такая константа $\Theta_0 > 0$, что $d(x_*) \leqslant \Theta_0^2$, где x_* — точное решение задачи (2.1), удовлетворяющее (2.2).

Для произвольных $x, y \in X$ рассмотрим соответствующее расхождение Брегмана

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

В зависимости от постановки конкретной задачи возможны различные подходы к определению прокс-структуры задачи и соответствующей функции расхождения Брегмана (евклидова, энтропийная и др.) Стандартно определим оператор проектирования

$$\mathrm{Mirr}_x(p) = \arg\min_{u \in X} \big\{ \langle p, u \rangle + V(x, u) \big\} \quad \text{для всяких} \quad x \in X \quad \text{и} \quad p \in \mathbb{R}^n.$$

Сделаем предположения о том, что оператор $\mathrm{Mirr}_x(p)$ легко вычислим, а также рассматриваемая задача (2.1) разрешима.

Следующее утверждение известно на классе выпуклых функционалов (см., например, [8, лемма 1]), а мы сформулируем его аналог для слабо α-квазивыпуклых функционалов. Стандартно будем обозначать через $\|\cdot\|_*$ двойственную (сопряженную) норму к $\|\cdot\|$.

Лемма 1. *Пусть $f\colon X \to \mathbb{R}$ — слабо α-квазивыпуклый функционал относительно решения x_* задачи* (2.1)*, субдифференцируемый по Кларку в любой точке $x \in X$. Тогда при $z = \mathrm{Mirr}_y(h\widetilde{\nabla} f(y))$ верно неравенство*

$$\alpha h(f(y) - f(x_*)) \leqslant h\langle \widetilde{\nabla} f(y),\, y - x_* \rangle \leqslant \frac{h^2}{2}\, ||\widetilde{\nabla} f(y)||_*^2 + V(y, x_*) - V(z, x_*) \tag{2.3}$$

для произвольного $h > 0$ и всякого субградиента $\widetilde{\nabla} f(y)$.

Отметим лишь, что, например, в [8, лемма 1] показано, что для всякого $x \in X$ верно неравенство

$$h\langle \widetilde{\nabla} f(y),\, y - x \rangle \leqslant \frac{h^2}{2}\, ||\widetilde{\nabla} f(y)||_*^2 + V(y, x) - V(z, x).$$

Если теперь положить $x = x_*$ и воспользоваться условием слабой α-квазивыпуклости f (2.2), то получим неравенство (2.3).

Перейдем к методу для случая слабо α-квазивыпуклых целевого функционала и функционала ограничения относительно искомой точки минимума x_* задачи (2.1). Расширение класса функционалов, в частности, приводит к видоизменению условия проверки продуктивности шага (см. п. 5 алгоритма 1).

З а м е ч а н и е 1. На непродуктивном шаге субградиент функционала ограничения не может быть нулевым. Действительно, если $\widetilde{\nabla} g(x^N) = 0$, то

$$g(x_*) \geqslant g(x^N) + \frac{1}{\alpha}\langle \widetilde{\nabla} g(x^N),\, x_* - x^N \rangle = g(x^N).$$

Поэтому $g(x^N) \leqslant g(x_*) \leqslant 0$, что противоречит неравенству

$$\alpha g(x^N) \geqslant \varepsilon\, \|\widetilde{\nabla} g(x^N)\|$$

вне зависимости от значений $\varepsilon > 0$ и $\alpha \in (0; 1]$.

Ясно, что либо на каком-то продуктивном шаге целевая функция будет иметь нулевой субградиент (и тогда метод можно остановить, поскольку задача решена точно), либо после некоторого количества итераций будет выполнен критерий остановки (п. 14 листинга). Допустим, что субградиент целевой функции отличен от 0 на всех продуктивных итерациях предложенного метода. Выпишем теоретический результат о качестве решения после остановки алгоритма 1.

А л г о р и т м 1: Адаптивный зеркальный спуск, слабо α-квазивыпуклые функционалы

Require: $\varepsilon > 0,\ \Theta_0 \colon d(x_*) \leqslant \Theta_0^2$
$x^0 = \arg\min\limits_{x \in X} d(x)$
$I =: \varnothing$
$N \leftarrow 0$
repeat
  if $g(x^N) \leqslant \dfrac{\varepsilon}{\alpha} \|\widetilde{\nabla} g(x^N)\|_*$ **then**
    $M_N = \|\widetilde{\nabla} f(x^N)\|_*,\ h_N = \dfrac{\varepsilon}{M_N^2}$
    $x^{N+1} = Mirr_{x^N}(h_N \widetilde{\nabla} f(x^N))$ // *“продуктивный шаг”*
    $N \to I$
  else
    $M_N = \|\widetilde{\nabla} g(x^N)\|_*,\ h_N = \dfrac{\varepsilon}{M_N}$
    $x^{N+1} = Mirr_{x^N}(h_N \widetilde{\nabla} g(x^N))$ // *“непродуктивный шаг”*
  end if
  $N \leftarrow N + 1$
until $2\dfrac{\Theta_0^2}{\varepsilon^2} \leqslant \sum\limits_{j \in I} \dfrac{1}{M_j^2} + N - |I|$
Ensure: $\widehat{x} := \arg\min\limits_{x^k,\, k \in I} f(x^k)$

Теорема 1. *Пусть функционалы f и g слабо α-квазивыпуклы для некоторого $\alpha \in (0;1]$ относительно искомой точки минимума x_* задачи* (2.1). *Тогда после выполнения критерия остановки алгоритма* 1 *для $\widehat{x} = \arg\min\limits_{k \in I} f(x^k)$ верны соотношения*

$$f(\widehat{x}) - f(x_*) \leqslant \frac{\varepsilon}{\alpha}, \quad g(\widehat{x}) \leqslant \frac{\varepsilon}{\alpha} \|g(\hat{x})\|_*. \tag{2.4}$$

Д о к а з а т е л ь с т в о. Предположим, что критерий остановки выполнен после N итераций алгоритма 1 и $k = 0, 1, \ldots, N-1$.

Если $k \in I$, то по лемме 1

$$\alpha h_k \left(f(x^k) - f(x_*)\right) \leqslant h_k \left\langle \widetilde{\nabla} f(x^k), x^k - x_* \right\rangle \leqslant \frac{h_k^2}{2} ||\widetilde{\nabla} f(x^k)||_*^2 + V(x^k, x_*) - V(x^{k+1}, x_*)$$
$$= \frac{\varepsilon^2}{2} \frac{1}{||\widetilde{\nabla} f(x^k)||_*^2} + V(x^k, x_*) - V(x^{k+1}, x_*). \tag{2.5}$$

Если $k \notin I$, то $\dfrac{\alpha g(x^k)}{||\widetilde{\nabla} g(x^k)||_*} > \varepsilon$ и $\dfrac{\alpha(g(x^k) - g(x_*))}{||\widetilde{\nabla} g(x^k)||_*} \geqslant \dfrac{\alpha g(x^k)}{||\widetilde{\nabla} g(x^k)||_*} > \varepsilon$. Поэтому по лемме 1 верны неравенства

$$\varepsilon^2 < \alpha h_k \left(g(x^k) - g(x_*)\right) \leqslant \frac{h_k^2}{2} ||\widetilde{\nabla} g(x^k)||_*^2 + V(x^k, x_*) - V(x^{k+1}, x_*)$$
$$= \frac{\varepsilon^2}{2} + V(x^k, x_*) - V(x^{k+1}, x_*) \tag{2.6}$$

или

$$\frac{\varepsilon^2}{2} < V(x^k, x_*) - V(x^{k+1}, x_*).$$

После суммирования неравенств (2.5) и (2.6) имеем:

$$\alpha \sum_{k \in I} h_k \left(f(x^k) - f(x_*)\right) \leqslant \sum_{k \in I} \frac{\varepsilon^2}{2\,||\widetilde{\nabla} f(x^k)||_*^2} - \frac{\varepsilon^2\,|J|}{2} + V(x^0, x_*) - V(x^{k+1}, x_*)$$

$$= \frac{\varepsilon}{2}\sum_{k\in I} h_k - \frac{\varepsilon^2|J|}{2} + \Theta_0^2 = \varepsilon \sum_{k\in I} h_k - \frac{\varepsilon^2}{2}\left(\sum_{k\in I}\frac{1}{||\widetilde{\nabla} f(x^k)||_*^2}\right) + |J| + \Theta_0^2.$$

После выполнения критерия остановки алгоритма (п. 14 листинга алгоритма 1):

$$\alpha \sum_{k\in I} h_k \left(f(x^k) - f(x_*)\right) \leqslant \varepsilon \sum_{k\in I} h_k,$$

поэтому для выхода $\widehat{x} = \arg\min\limits_{k\in I} f(x^k)$ будет верно $f(\hat{x}) - f(x_*) \leqslant \dfrac{\varepsilon}{\alpha}$, причем для произвольного $k \in I$ $g(x^k) \leqslant \dfrac{\varepsilon}{\alpha}\, ||\widetilde{\nabla} g(x^k)||_*$, откуда $g(\widehat{x}) \leqslant \dfrac{\varepsilon}{\alpha}\, \|\widetilde{\nabla} g(\hat{x})\|_*$.

Остается лишь показать, что множество продуктивных шагов I непусто. Если $I = \varnothing$, то $|J| = N$ и из выполнения критерия остановки (п. 14 листинга алгоритма 1) следует, что $N \geqslant \dfrac{2\,\Theta_0^2}{\varepsilon^2}$. С другой стороны, из (2.6) имеем

$$\frac{\varepsilon^2 N}{2} < V(x^0, x_*) \leqslant \Theta_0^2,$$

т. е. получаем противоречие, и $I \neq \varnothing$.

Теорема доказана.

Ясно, что в случае M_g-липшицевости на X функционала ограничения g второе неравенство в (2.4) означает $g(\widehat{x}) \leqslant \dfrac{\varepsilon}{\alpha} M_g$. При этом если использовать второе неравенство в (2.4), то оценка по ограничению может быть лучше. Также стоит заметить, что можно использовать соотношение (2.4) и для g, не удовлетворяющих условию Липшица. Но при этом уже нет возможности гарантировать достаточно хорошее качество по ограничению, поскольку субградиенты в общем случае не ограничены.

З а м е ч а н и е 2. Оценим количество итераций, необходимых для выполнения критерия остановки алгоритма 1 для M_f-липшицева целевого функционала на X. Ясно, что для всякого $k \in I$ верно неравенство $\|\widetilde{\nabla} f(x^k)\|_* \leqslant M_f$, и поэтому

$$|J| + \sum_{k\in I}\frac{1}{\|\widetilde{\nabla} f(x^k)\|_*^2} \geqslant |J| + \frac{|I|}{\max\limits_{k\in I}\|\widetilde{\nabla} f(x^k)\|_*^2} \geqslant |J| + \frac{|I|}{M_f^2} \geqslant (|I| + |J|)\frac{1}{\max\{1, M_f^2\}}.$$

Это означает, что при

$$N \geqslant \frac{2\,\Theta_0^2 \max\{1, M_f^2\}}{\varepsilon^2} \tag{2.7}$$

критерий остановки заведомо выполнен, т. е. искомая точность достигается за $O\big(\varepsilon^{-2}\big)$ итераций. Это указывает на оптимальность алгоритма 1 с точки зрения нижних оракульных оценок (см. [4]). В силу адаптивности алгоритм 1 применим и к задачам для целевых функционалов, которые не удовлетворяют условию Липшица. Однако при этом уже нельзя гарантировать оптимальную скорость сходимости рассматриваемого метода.

З а м е ч а н и е 3. Очевидно, что полученные оценки можно уточнять в случае, когда параметры слабой α-квазивыпуклости для f и g различны (в частности, если один из функционалов f или g выпуклый в обычном смысле). Например, если целевой функционал f выпуклый, а g — слабо α-квазивыпуклый, то оценка (2.4) принимает вид

$$f(\widehat{x}) - f(x_*) \leqslant \varepsilon, \quad g(\widehat{x}) \leqslant \frac{\varepsilon}{\alpha}\,\|\widetilde{\nabla} g(\widehat{x})\|_*.$$

3. Оценка скорости сходимости субградиентного метода с переключениями с относительной точностью по целевому функционалу

Теперь рассмотрим задачу (1.1) минимизации выпуклой однородной функции на выпуклом замкнутом подмножестве $X \subset \mathbb{R}^n$. В данном разделе работы будем аналогично [3, гл. 6] предполагать (1.3), откуда следует $f(x) \geqslant \gamma_0\|x\| \ \forall x \in X$ для некоторого фиксированного $\gamma_0 > 0$. Далее, если не оговорено иное, в качестве начальной точки методов будем выбирать проекцию x^0 начала координат на множество X относительно рассматриваемой нормы

$$\|x^0\| := \min_{x\in X}\{\|x\|\}. \tag{3.1}$$

Напомним, что в [3, теорема 6.1.1] при сделанных предположениях обосновано следующее неравенство:

$$\|x^0 - x_*\| \leqslant \frac{2}{\gamma_0} f^*. \tag{3.2}$$

Если допустимое множество задачи есть аффинное подпространство, x^0 — проекция 0 на это подпространство и норма евклидова $\|\cdot\| = \|\cdot\|_2$, то согласно [3, теорема 6.1.1] неравенство (3.2) можно усилить

$$\|x^0 - x_*\|_2 \leqslant \frac{1}{\gamma_0} f^*. \tag{3.3}$$

Это позволит далее улучшать полученные в общем случае вычислительные гарантии в несколько раз.

Очевидно, что аналогичное (3.3) неравенство верно уже для произвольной нормы и любого допустимого множества задачи в случае $0 \in X$, если ограничения возможно записать в виде системы неравенств с выпуклыми функционалами, а точку x^0 выбрать нулевой. В частности, для задачи (1.2) вполне возможно ограничение вида $x \in \mathcal{L}$ ($\mathcal{L}$ — аффинное подпространство) записать в виде системы двух неравенств для линейных функционалов ограничения, а в качестве X выбрать все пространство. Такой подход даст возможность по сравнению с [3], в частности, избежать при выборе начальной точки x^0 дополнительной операции проектирования 0 на аффинное подпространство $\mathcal{L}$.

Теперь покажем, как можно видоизменить алгоритм 1, чтобы было возможно гарантировать достижение качества решения задачи (1.1) с заданной относительной точностью $\delta > 0$. Слабую α-квазивыпуклость будем предполагать лишь для функционала ограничения g относительно искомого решения x_* задачи (1.1).

Теорема 2. *Пусть выпуклый однородный функционал f M_f-липшицев на X при некотором $M_f > 0$ и удовлетворяет условию (1.3), g — слабо α-квазивыпуклый относительно искомого решения x_* задачи (1.1) на X, а также для некоторого $C > 0$ верно*

$$2\,V(x^0, x_*) \leqslant C^2\,\|x^0 - x_*\|^2 \tag{3.4}$$

и начальная точка x^0 выбрана согласно (3.1). Тогда для всякого $\delta > 0$ можно подобрать входные параметры $\varepsilon > 0$ и $\Theta_0 > 0$ алгоритма 1 так, что после

$$N \geqslant \frac{4\,C^2\,\max\{1, M_f^2\}}{\gamma_0^2\delta^2} \tag{3.5}$$

итераций этого метода гарантированно будут выполнены следующие неравенства:

$$f(\widehat{x}) \leqslant f^*(1+\delta), \quad g(\widehat{x}) \leqslant \frac{\delta\, f^*\, \|\widetilde{\nabla} g(\widehat{x})\|_*}{\alpha}.$$

Доказательство. В силу предположения (3.4) можно подобрать входной параметр Θ_0 алгоритма 1 так, чтобы для некоторого $R > 0$ было верно

$$2\,V(x^0, x_*) \leqslant 2\,\Theta_0^2 = R^2 \leqslant C^2\,\|x^0 - x_*\|^2.$$

Выберем теперь $\varepsilon = \dfrac{R\,\gamma_0\,\delta}{2\,C}$. Тогда после остановки алгоритма 1 согласно теореме 1 и замечанию 3 будут выполняться следующие неравенства:

$$f(\widehat{x}) - f^* \leqslant \frac{R\,\gamma_0\,\delta}{2\,C}, \quad g(\widehat{x}) \leqslant \frac{R\,\gamma_0\,\delta\,\|\widetilde{\nabla} g(\hat{x})\|_*}{2\,C\,\alpha}.$$

Теперь исходя из сделанного предположения $R \leqslant C\,\|x^0 - x^*\|$ и в силу неравенства (3.2) имеем

$$f(\widehat{x}) - f^* \leqslant \frac{R\,\gamma_0\,\delta}{2\,C} \leqslant \frac{\gamma_0\,\delta\,\|x^0 - x^*\|}{2} \leqslant f^*\delta,$$

а также $g(\hat{x}) \leqslant \dfrac{\delta\,f^*\,\|\widetilde{\nabla} g(\hat{x})\|_*}{\alpha}$.

Вспоминая оценку количества итераций (2.7) (после которых гарантированно будет выполнен критерий остановки алгоритма 1) с учетом $\varepsilon = \dfrac{R\,\gamma_0\,\delta}{2\,C}$, получим оценку (3.5) числа итераций алгоритма 1, после которых для точки-выхода заведомо будет достигнута δ-относительная точность решения поставленной задачи по целевому функционалу.

Теорема доказана.

Замечание 4. Отметим (см. [4; 14]), что предположение (3.4) вполне естественно для многих известных примеров выбора прокс-функций вне зависимости от x^0 и x_*. Хорошо известно [4; 14], что если для произвольных $x, y \in X$ существует зависящая от величины размерности пространства константа $C_n > 0$ (причем $C_n = O(\log n)$) такая, что $d(x - y) \leqslant C_n\,\|x - y\|^2$, то $V(x, y) \leqslant C_n\,\|x - y\|^2$ при любых $x, y \in X$. Тогда (3.4) будет верно при всяком $C^2 \geqslant \dfrac{C_n}{2}$ вне зависимости от x^0 и x_*.

Замечание 5. Если воспользоваться адаптивным критерием остановки алгоритма 1 (п. 14 листинга) и не заменять $\displaystyle\sum_{k \in I} \frac{1}{\|\widetilde{\nabla} f(x^k)\|^2}$ на $\dfrac{|I|}{M_f^2}$, то при практической реализации метода достижение приемлемого качества решения возможно за меньшее число итераций по сравнению с доказанной оценкой (3.5). В частности, неравенства (2.7) и (3.5) можно заменить на следующие:

$$N \geqslant \frac{2\,\Theta_0^2}{\varepsilon^2} \max\bigl\{1, \max_{k \in I} \|\widetilde{\nabla} f(x^k)\|_*^2\bigr\} = \frac{4\,C^2}{\gamma_0^2\,\delta^2} \max\bigl\{1, \max_{k \in I} \|\widetilde{\nabla} f(x^k)\|_*^2\bigr\}.$$

Эти оценки могут оказаться лучшими по сравнению с (2.7) и (3.5) на классе M_f-липшицевых целевых функционалов.

Замечание 6. Рассмотрим теперь случай евклидовой нормы и прокс-структуры, т. е.

$$V(x, y) = d(x - y) = \frac{1}{2}\,\|x - y\|_2^2$$

для произвольных $x, y \in X$, и предположим, что допустимое множество задачи есть аффинное подпространство, а начальная точка x^0 — евклидова проекция начала координат на допустимое множество задачи. В этом случае рассуждения существенно опираются на неравенство (3.3) для выпуклого и однородного функционала f, что позволяет уменьшить константу в полученной оценке (3.5). Кроме того, в этом случае нет необходимости вводить предположение (3.4). Действительно, для евклидовой нормы возможно подобрать входной параметр Θ_0 алгоритма 1 так, чтобы для некоторых $R > 0$ и $C \geqslant 1$ было верно

$$\|x^0 - x_*\|_2^2 \leqslant 2\,\Theta_0^2 = R^2 \leqslant C^2\,\|x^0 - x_*\|^2,$$

и положить $\varepsilon = \dfrac{R\,\gamma_0\,\delta}{C}$. Это означает, что после остановки алгоритма 1 будут выполняться неравенства

$$f(\widehat{x}) - f^* \leqslant \frac{R\,\gamma_0\,\delta}{C}, \quad g(\widehat{x}) \leqslant \frac{R\,\gamma_0\,\delta\,\|g(\hat{x})\|_*}{C\,\alpha}.$$

Поэтому для описанных выше предположений ввиду (3.3) имеем

$$f(\widehat{x}) \leqslant f^*(1+\delta), \quad g(\widehat{x}) \leqslant \frac{\delta\,f^*\,\|g(\hat{x})\|_*}{\alpha},$$

при чем в силу выбора $\varepsilon = \dfrac{R\,\gamma_0\,\delta}{C}$ оценка числа итераций (3.5), необходимых для выполнения критерия остановки алгоритма, уменьшается в 4 раза. В отличие от предыдущего результата параметр $C \geqslant 1$ можно выбрать произвольно. Чем больше выбирается параметр C, тем выше будет точность достигнутого решения, но вычислительные затраты (необходимое количество итераций) будут большими. Ясно, что улучшение оценки (3.5) в 4 раза в условиях теоремы 2 возможно и для произвольной нормы, если $0 \in X$.

Теперь рассмотрим еще один метод зеркального спуска с переключениями (см. [9, алгоритм 3]), для которого при соответствующих предположениях можно выписать оценки скорости сходимости для задачи минимизации выпуклого однородного M_f-липшицева функционала f на множестве X с унимодальным M_g-липшицевым на X функционалом ограничения g ($M_f, M_g > 0$). Напомним, что функционал $g\colon X \to \mathbb{R}$ называется *унимодальным* (*квазивыпуклым*), если для произвольных $x, y \in X$

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

Обратим внимание, что при анализе этого метода для унимодального функционала ограничения g вместо (суб)градиента $\widetilde{\nabla} g(y)$ возможно рассматривать некоторый элемент набора векторов нормалей ко множеству уровня функции g в точке x

$$\hat{D}\,g(x) = \{p \mid \langle p, x - y\rangle \geqslant 0 \;\; \forall y \in X\colon g(y) < g(x)\}.$$

Как правило, $\hat{D}g(x)$ — непустой, замкнутый и выпуклый конус. Следуя [15], мы предполагаем, что $\hat{D}\,g(x) \neq \{0\}$ при $x \neq x_*$. Здесь уже x_* — произвольное решение задачи (2.1). Будем обозначать через $D\,g(x)$ произвольный ненулевой вектор из $\hat{D}\,g(x)$. Однако вместо $D\,g(x)$ можно использовать и субградиент в смысле Кларка, если он существует, конечен и отличен от нуля. Для выпуклого целевого функционала будем использовать субградиент $\widetilde{\nabla} f$ в смысле выпуклого анализа. В этом случае очевидно, что $\widetilde{\nabla} f(y) \neq 0$ при $y \neq x_*$.

Справедлива следующая теорема.

Теорема 3. *Пусть выпуклый однородный функционал f M_f-липшицев на X при некотором $M_f > 0$ и удовлетворяет условию* (1.3)*, g унимодален и M_g-липшицев на X при $M_g > 0$, а также верно* (3.4) *и начальная точка x^0 выбрана согласно* (3.1)*. Тогда для всякого $\delta > 0$ можно подобрать входные параметры $\varepsilon > 0$ и $\Theta_0 > 0$ алгоритма* 2 *так, чтобы после*

$$N \geqslant \frac{4\,C^2\,M_f^2}{\gamma_0^2\,\delta^2}$$

итераций этого метода гарантированно будут выполнены следующие неравенства:

$$f(\widehat{x}) \leqslant f^*(1+\delta), \quad g(\widehat{x}) \leqslant \frac{M_g\,f^*\,\delta}{M_f}.$$

Доказательство. В [9, теорема 4] показано, что для унимодального (и в частности выпуклого) M_f-липшицева функционала f $(M_f > 0)$ после выполнения критерия остановки алгоритма 2 (выполнения достаточного количества итераций) верны следующие неравенства:

$$\min_{k\in I} f(x^k) - f(x_*) \leqslant M_f\varepsilon, \quad \max_{k\in I} g(x^k) \leqslant \varepsilon\,M_g. \tag{3.6}$$

Алгоритм 2: Адаптивный зеркальный спуск, случай унимодального функционала ограничения

Require: $\varepsilon > 0,\ \Theta_0 \colon d(x_*) \leqslant \Theta_0^2$

$x^0 = \arg\min\limits_{x\in X} d(x)$

$I =: \varnothing$

$N \leftarrow 0$

repeat

  if $g(x^N) \leqslant \varepsilon M_g$ **then**

    $x^{N+1} = Mirr_{x^N}\Big(\dfrac{\varepsilon \widetilde{\nabla} f(x^N))}{\|\widetilde{\nabla} f(x^N))\|_*}\Big)$ // *"продуктивный шаг"*

  else

    $x^{N+1} = Mirr_{x^N}\Big(\dfrac{\varepsilon D\, g(x^N))}{\|D\, g(x^N))\|_*}\Big)$ // *"непродуктивный шаг"*

  end if

  $N \leftarrow N + 1$

until $2\,\dfrac{\Theta_0^2}{\varepsilon^2} \leqslant N$

Ensure: $\widehat{x} := \arg\min\limits_{x^k,\ k\in I} f(x^k)$

В силу предположения (3.4) возможно подобрать входной параметр Θ_0 алгоритма 1 так, чтобы для некоторого $R > 0$ было верно $2\,V(x_*, x^0) \leqslant 2\,\Theta_0^2 = R^2 \leqslant C^2\,\|x^0 - x_*\|^2$. Выберем теперь $\varepsilon = \dfrac{R}{\sqrt{N}}$, что вполне корректно в силу критерия остановки алгоритма 2. Тогда (3.2) и (3.6) означают, что

$$f(\hat{x}) - f^* = \min_{k\in I} f(x^k) - f^* \leqslant \frac{M_f\,R}{\sqrt{N}} \leqslant \frac{2\,C\,f^*\,M_f}{\gamma_0\,\sqrt{N}},$$

а также

$$g(\hat{x}) \leqslant \max_{k\in I} g(x^k) \leqslant \frac{2\,C\,f^*\,M_g}{\gamma_0\,\sqrt{N}} \leqslant \frac{M_g\,f^*\,\delta}{M_f}.$$

Теорема доказана.

Аналогично замечанию 6 можно заключить, что если для произвольной нормы возможно дополнительно предположить $x^0 = 0 \in X$ или же норма евклидова и допустимое множество задачи — это аффинное подпространство (начальная точка есть проекция 0 на это подпространство), то полученную в теореме 3 оценку количества итераций возможно уменьшить в 4 раза:

$$f(\hat{x}) \leqslant \Big(1 + \frac{C\,M_f}{\gamma_0\,\sqrt{N}}\Big) f^* \leqslant (1+\delta) f^*,$$

откуда $N \geqslant \dfrac{C^2\,M_f^2}{\gamma_0^2\,\delta^2}$, а также

$$g(\hat{x}) \leqslant \frac{C\,M_g\,f^*}{\gamma_0\,\sqrt{N}} \leqslant \frac{M_g\,f^*\,\delta}{M_f}.$$

4. Заключение

В работе предложены некоторые субградиентные методы, применимые к задачам минимизации выпуклого однородного функционала при наличии одного или нескольких выпуклых липшицевых функционалов ограничения. Полученные результаты основаны на специальных подходах к выбору шага и критерия остановки в зеркальных спусках с переключениями. Показано, что выполнение предложенных критериев остановки гарантирует для точки-выхода достижение заданной относительной точности $\delta > 0$ решения задачи по целевому функционалу, причем значение функционала ограничения сопоставимо с δ. При этом попутно предложен алгоритм 1 и в качестве базового результата доказана оценка скорости сходимости с абсолютной точностью для задачи минимизации слабо α-квазивыпуклого целевого функционала со слабо α-квазивыпуклым функционалом ограничения.

Если говорить о доказанных теоретических оценках для алгоритмов 1 и 2, то они оптимальны с точностью умножения на константу на классе задач выпуклого программирования с выпуклыми липшицевыми функционалами [4]. Тем более это касается более широкого класса слабо α-квазивыпуклых или унимодальных липшицевых функционалов. Теоретические результаты для алгоритма 1 применимы и на более широком классе задач со слабо α-квазивыпуклым функционалом ограничения, для алгоритма 2 — на классе задач с унимодальным M_g-липшицевым функционалом ограничения. Однако алгоритм 2 не адаптивен в том смысле, что явно использует значение константы Липшица $M_g > 0$ при проверке продуктивности шага (а не только в оценке скорости сходимости). Алгоритм 1 позволяет этого избежать, но для другого класса слабо α-квазивыпуклых функционалов ограничения. Однако при этом для реализации алгоритма 1 необходимо знать $\alpha > 0$, да и условие унимодальности представляется проще для проверки на практике. Отметим, что для задач с унимодальным целевым функционалом и выпуклым функционалом ограничениям возможно использовать аналог алгоритма 2 с условием проверки продуктивности шага, которое не использует константу M_g (см. [9, алгоритм 2]). Но в настоящей работе это опущено, поскольку выпуклость целевого функционала существенно использована для получения оценок с относительной точностью, а для выпуклых задач алгоритм 1 приводит к лучшим оценкам в сравнении с алгоритмом 2.

В качестве возможного развития настоящей работы представляется интересным исследование свойств прямо-двойственности предложенных зеркальных спусков по аналогии с [8], но уже для задач с относительной точностью. Также представляет интерес разработка методов с гарантией достижения приемлемой относительной точности решения для негладких задач со специальной структурой (например, седловых задач), которая может улучшать оценки скорости сходимости в плане необходимого числа итераций на базе использования различных ускоренных методов оптимизации.

Авторы выражают огромную признательность А. В. Гасникову, а также рецензенту за полезные обсуждения и рекомендации.

СПИСОК ЛИТЕРАТУРЫ

1. **Nesterov Yu.** Rounding of convex sets and efficient gradient methods for linear programming problems // Optimization Methods and Software. 2008. Vol. 23, no. 1. P. 109–128.
2. **Nesterov Yu.** Unconstrained convex minimization in relative scale // Math. Oper. Res. 2009. Vol. 34, no. 1. P. 180–193. 10.1287/moor.1080.0348 .
3. **Нестеров Ю. Е.** Алгоритмическая выпуклая оптимизация: дис. … д-ра физ.-мат. наук / Моск. физ.-техн. ин-т. 2013. 367 с.
4. **Немировский А. С., Юдин Д. Б.** Сложность задач и эффективность методов оптимизации. Москва: Наука, 1979. 384 с.
5. **Поляк Б. Т.** Один общий метод решения экстремальных задач // Докл. АН СССР. 1967. Т. 174, № 1. С. 33–36.
6. **Шор Н. З.** Применение обобщенного градиентного спуска в блочном программировании // Кибернетика. 1967. № 3. С. 53–55.
7. **Немировский А. С., Юдин Д. Б.** Эффективные методы решения задач выпуклого программирования большой размерности // Экономика и мат. методы. 1979. № 2. С. 135–152.
8. **Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A.** Mirror descentand convex optimization problems with non-smooth inequality constraints // Large-Scale and Distributed Optimization / eds. Pontus GiselssonAnders Rantzer. Springer: Cham. 2018. P. 181–213. (Lect. Notes Math.; vol. 2227) doi: 10.1007/978-3-319-97478-1_8 .

9. **Stonyakin F., Stepanov A., Gasnikov A., Titov A.** Mirror descent for constrained optimization problems with large subgradient values of functional constraints // Компьютерные исследования и моделирование. 2020. Vol. 12, no. 2. P. 301–317. doi: 10.20537/2076-7633-2020-12-2-301-317 .
10. **Стонякин Ф. С., Алкуса М. С., Степанов А. Н., Титов А. А.** Адаптивные алгоритмы зеркального спуска для задач выпуклой и сильно выпуклой оптимизации с функциональными ограничениями // Дискретный анализ и исслед. операций. 2019. Т. 26, № 3. С. 88–114. doi: 10.33048/daio.2019.26.636 .
11. **Гасников А. В.** Современные численные методы оптимизации. Метод универсального градиентного спуска. Москва: Изд-во МФТИ, 2018. 286 с.
12. **Гуминов С. В., Нестеров Ю. Е., Двуреченский П. Е., Гасников А. В.** Прямо-двойственный ускоренный градиентный метод с одномерным поиском для выпуклых, невыпуклых и негладких задач оптимизации // Докл. РАН. 2019. Т. 485, № 1. С. 15–18. doi: 10.31857/S0869-5652485115-18 .
13. **Hinder O., Sidford A., Sohoni N. S.** Near-optimal methods for minimizing star-convex functions and beyond [e-resource]. 2019. 37 p. URL: https://arxiv.org/pdf/1906.11985.pdf .
14. **Ben-Tal A., Nemirovski A.** Lectures on modern convex optimization analysis, algorithms, and engineering applications. Philadelphia: SIAM, 2019. 647 p.
15. **Нестеров Ю. Е.** Эффективные методы нелинейного программирования. Москва: Радио и связь, 1989. 301 с.

Поступила 9.06.2020
После доработки 14.08.2020
Принята к публикации 24.08.2020

Стонякин Федор Сергеевич
канд. физ.-мат. наук
доцент
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: fedyor@mail.ru

Баран Инна Викторовна
канд. физ.-мат. наук
старший преподаватель
Крымский федеральный университет им. В. И. Вернадского
г. Симферополь
e-mail: matemain@mail.ru

REFERENCES

1. Nesterov Yu. Rounding of convex sets and efficient gradient methods for linear programming problems. *Optimization Methods and Software*, 2008, vol. 23, no. 1, pp. 109–128.
2. Nesterov Yu. Unconstrained convex minimization in relative scale. *Math. Oper. Res.*, 2009, vol. 34, no. 1, pp. 180–193. doi: 10.1287/moor.1080.0348 .
3. Nesterov Yu. E. *Algoritmicheskaya vypuklaya optimizatsiya* [Algorithmic convex optimization]. Dr. Phys.-Math. Sci. Dissertation. Moscow: Publ. Mosk. Phys.-tech. Inst. (State University), 2013. 367‘p.
4. Nemirovskii A., Yudin D. *Problem complexity and method efficiency in optimization.* N Y: J. Wiley & Sons., 1983.388 p. Original Russian text published in Nemirovskii A., Yudin D. Slozhnost’ zadach i effektivnost’ metodov optimizatsii, Moscow: Nauka Publ., 384 p.
5. Polyak B. T. A general method of solving extremum problems. *Sov. Math. Dokl.*, 1967, vol. 8, no. 3, pp. 593—597.
6. Shor N. Z. Application of generalized gradient descent in block programming. *Kibernetika*, 1967, no. 3, pp. 53—55 (in Russian).
7. Nemirovskii A., Yudin D. Efficient methods for large-scale convex optimization problems. *Ekonomika Mat. Metody*, 1979, no. 2, pp. 135—152 (in Russian).
8. Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A. Mirror descentand convex optimization problems with non-smooth inequality constraints, In: Large-Scale and Distributed Optimization, eds. Pontus GiselssonAnders Rantzer, Springer: Cham., 2018, Ser. Lect. Notes Math., vol. 2227, Springer, Cham., 2018, pp. 181-–213. doi: 10.1007/978-3-319-97478-1_8 .

9. Stonyakin F., Stepanov A., Gasnikov A., Titov A. Mirror descent for constrained optimization problems with large subgradient values of functional constraints. *Computer Research and Modeling*, 2020, vol. 12, no. 2, pp. 301-–317. doi: 10.20537/2076-7633-2020-12-2-301-317 .
10. Stonyakin F., Alkousa M., Stepanov A., Titov A. Adaptive mirror descent algorithms for convex and strongly convex optimization problems with functional constraints. *J. Appl. Industr. Math.*, 2019, vol. 13, no. 3, pp. 557-–574. doi: 10.33048/daio.2019.26.636 .
11. Gasnikov A. V. *Sovremennye chislennye metody optimizatsii. Metod universal'nogo gradientnogo spuska* [Modern numerical optimization methods. The universal gradient descent method]. Moscow: MIPT Publ., 2018, 286 p.
12. Guminov S., Nesterov Yu., Dvurechensky P., Gasnikov A. Primal-dual accelerated gradient methods with small-dimensional relaxation oracle. *Reports of the Academy of Sciences*, 2019, vol. 485, no. 1, pp. 15–18 (in Russian). doi: 10.31857/S0869-5652485115-18 .
13. Hinder O., Sidford A., Sohoni N. S. *Near-optimal methods for minimizing star-convex functions and beyond* [e-resource]. 2019. 37 p. Available at https://arxiv.org/pdf/1906.11985.pdf .
14. Nemirovski A. Lectures on modern convex optimization analysis, algorithms, and engineering applications. *Philadelphia: SIAM*, 2013. 647 p.
15. Nesterov Yu. *Jeffektivnye metody nelinejnogo programmirovanija*, [Effective methods in nonlinear programming]. Moscow, Radio and Communication Publ., 1989. 301 p.

Received June 9, 2020
Revised August 14, 2020
Accepted August 24, 2020

Funding Agency: The research of F. Stonyakin in Algorithm 1 and Theorems 1, 2 and 3 and the research of I. Baran in Algorithm 2 was supported by the grant of the President of the Russian Federation for young Russian candidates of sciences, code MK-15.2020.1.

Fedor Sergeevich Stonyakin, Cand. Sci. (Phys.-Math.), V. I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: fedyor@mail.ru .

Inna Viktorovna Baran, Cand. Sci. (Phys.-Math.), V. I. Vernadsky Crimean Federal University, Simferopol, Republic of Crimea, 295007 Russia, e-mail: matemain@mail.ru .

Cite this article as: Cite this article as: F. S. Stonyakin, I. V. Baran, On some algorithms for constrained optimization problems with relative accuracy with respect to the objective functional, *Trudy Instituta Matematiki i Mekhaniki URO RAN*, 2020, vol. 26, no. 3, pp. 198–210.