

УДК: 519.86

Об одном методе минимизации выпуклой липшицевой функции двух переменных на квадрате

Д. А. Пасечнюк^{1,a}, Ф. С. Стонякин^{2,3,b}

¹Президентский физико-математический лицей № 239,
Россия, 191028, г. Санкт-Петербург, ул. Кирочная, д. 8

²Крымский федеральный университет имени В. И. Вернадского,
Россия, 295007, г. Симферополь, пр. Академика Вернадского, д. 4

³Московский физико-технический институт,
Россия, 141070, г. Долгопрудный, Институтский пер., д. 9

E-mail: ^a pasechnyuk2004@gmail.com, ^b fedyor@mail.ru

Получено 08.01.2019, после доработки — 08.02.2019.

Принято к публикации 22.04.2019.

В статье получены оценки скорости сходимости по функции для недавно предложенного Ю. Е. Нестеровым метода минимизации выпуклой липшицевой функции двух переменных на квадрате с фиксированной стороной. Идея метода — деление квадрата на меньшие части и постепенное их удаление так, чтобы в оставшейся достаточно малой части все значения целевой функции были достаточно близки к оптимальному. При этом метод заключается в решении вспомогательных задач одномерной минимизации вдоль разделяющих отрезков и не предполагает вычисления точного значения градиента целевого функционала. Основным результатом работы о необходимом количестве итераций для достижений заданной точности доказан в классе гладких выпуклых функций, имеющих липшицев градиент. При этом отмечено, что свойство липшицевости градиента достаточно потребовать не на всем квадрате, а лишь на некоторых отрезках. Показано, что метод может работать при наличии погрешностей решения вспомогательных одномерных задач, а также при вычислении направлений градиентов. Также описана ситуация, когда возможно пренебречь временными затратами (или уменьшить их) на решение вспомогательных одномерных задач. Для некоторых примеров экспериментально продемонстрировано, что метод может эффективно работать и на некоторых классах негладких функций. При этом построен пример простой негладкой функции, для которой при неудачном выборе субградиента даже в случае точного решения вспомогательных одномерных задач может не наблюдаться сходимость метода. Проведено сравнение работы метода Ю. Е. Нестерова, метода эллипсоидов и градиентного спуска для некоторых гладких выпуклых функций. Эксперименты показали, что метод Ю. Е. Нестерова может достигать желаемой точности решения задачи за меньшее (в сравнении с другими рассмотренными методами) время. В частности, замечено, что при увеличении точности искомого решения время работы метода Ю. Е. Нестерова может расти медленнее, чем время работы метода эллипсоидов.

Ключевые слова: задача минимизации, выпуклый функционал, липшицев функционал, липшицев градиент, негладкий функционал, субградиент, градиентный спуск, метод эллипсоидов, скорость сходимости

Исследования Ф. С. Стонякина, посвященные плану доказательства теоремы 1, разработке замечаний 2, 4 и 5, а также примера 2, выполнены при поддержке гранта Российского научного фонда 18-71-00048.

UDC: 519.86

One method for minimization a convex Lipschitz-continuous function of two variables on a fixed square

D. A. Pasechnyuk^{1,a}, F. S. Stonyakin^{2,3,b}¹Presidential Physics and Mathematics Lyceum No. 239,
8 Kirochnaya st., Saint-Petersburg, 191028, Russia²V. I. Vernadsky Crimean Federal University,
4 V. Vernadsky ave., Simferopol, 295007, Russia³Moscow Institute of Physics and Technology,
9 Institutsky ave., Dolgoprudny, 141070, RussiaE-mail: ^a pasechnyuk2004@gmail.com, ^b fedyor@mail.ru*Received 08.01.2019, after completion — 08.02.2019.**Accepted for publication 22.04.2019.*

In the article we have obtained some estimates of the rate of convergence for the recently proposed by Yu. E. Nesterov method of minimization of a convex Lipschitz-continuous function of two variables on a square with a fixed side. The idea of the method is to divide the square into smaller parts and gradually remove them so that in the remaining sufficiently small part. The method consists in solving auxiliary problems of one-dimensional minimization along the separating segments and does not imply the calculation of the exact value of the gradient of the objective functional. The main result of the paper is proved in the class of smooth convex functions having a Lipschitz-continuous gradient. Moreover, it is noted that the property of Lipschitz-continuity for gradient is sufficient to require not on the whole square, but only on some segments. It is shown that the method can work in the presence of errors in solving auxiliary one-dimensional problems, as well as in calculating the direction of gradients. Also we describe the situation when it is possible to neglect or reduce the time spent on solving auxiliary one-dimensional problems. For some examples, experiments have demonstrated that the method can work effectively on some classes of non-smooth functions. In this case, an example of a simple non-smooth function is constructed, for which, if the subgradient is chosen incorrectly, even if the auxiliary one-dimensional problem is exactly solved, the convergence property of the method may not hold. Experiments have shown that the method under consideration can achieve the desired accuracy of solving the problem in less time than the other methods (gradient descent and ellipsoid method) considered. Partially, it is noted that with an increase in the accuracy of the desired solution, the operating time for the Yu. E. Nesterov's method can grow slower than the time of the ellipsoid method.

Keywords: minimization problem, convex functional, Lipschitz-continuous functional, Lipschitz-continuous gradient, non-smooth functional, subgradient, gradient method, ellipsoid method, rate of convergence

Citation: *Computer Research and Modeling*, 2019, vol. 11, no. 3, pp. 379–395 (Russian).

Research by F. S. Stonyakin, devoted to the plane of the proof of Theorem 1, Remarks 2, 4, 5, and Example 2 was supported by the Russian Science Foundation grant 18-71-00048.

Введение

В последние десятилетия активно исследуются самые разные подходы к численным методам для условных и безусловных задач многомерной оптимизации [Васильев, 2011; Гасников, 2018; Bubeц, 2015; Nesterov, 2018]. Весьма часто для многомерных задач оптимизации практикуют переход к более простой двойственной задаче [Chen et al., 2014; Nesterov, 2009; Nesterov, Shpirko, 2014]. В частности, условные задачи оптимизации с небольшим количеством функциональных ограничений с помощью метода множителей Лагранжа могут сводиться к маломерным двойственным задачам. А для задач маломерной оптимизации обычно используют разные геометрические методы (см., например, [Bubeц, 2015; Lee et al., 2015; Nesterov, 2018]), среди которых весьма популярен метод центров тяжести Левина–Ньюмена, а также метод эллипсоидов (см., например, [Гасников, 2018; Bubeц, 2015; Lee et al., 2015; Nesterov, 2018]). Отметим [Немировский, Юдин, 1979; Bubeц, 2015], что метод центров тяжести имеет оптимальную оценку числа итераций n на классе выпуклых функционалов для достижения необходимой точности ε решения задачи (d — размерность пространства):

$$n \geq Cd \log \left(\frac{1}{\varepsilon} \right). \quad (1)$$

Однако стоимость самих итераций метода центра тяжести может быть довольно дорогой за счет трудоемкой операции нахождения центра тяжести выпуклого множества (см., например, [Гасников, 2018; Bubeц, 2015]). Поэтому на практике часто применяют метод эллипсоидов, где нахождение центра тяжести заменяется на нахождение центра описанного эллипсоида, что делает итерацию метода менее затратной. Указанные методы работают с линейной скоростью сходимости (сопоставимой с (1)) как для гладких, так и негладких функций и требуют для достижения достаточно приемлемого качества решения вычисления (суб)градиента на каждой итерации.

Известны самые разные подходы к проблеме ускорения работы (суб)градиентных методов оптимизации [Гасников, 2018; Bayandina et al., 2018; Bubeц, 2015; Nesterov, 2018]. Например, для удешевления стоимости итераций метода часто применяют рандомизированные процедуры, предполагающие вместо вычисления точных значений важных параметров (в частности, градиента) на итерациях вычисление их вероятностных аналогов [Баяндина и др., 2018; Воронцова и др., 2019; Bogolubsky et al., 2016; Dang, Lan, 2015; Dvurechensky, Gasnikov, 2016; Tappenden et al., 2016]: вместо (суб)градиентов на итерациях метода вычисление их стохастических аналогов позволяет уменьшить стоимость итерации. Однако для такого типа неполноградиентных методов при этом теряются детерминированные оценки гарантированного качества найденного решения и заменяются на вероятностные (стохастические).

Также в последние годы активно исследуются оптимальные методы, при работе которых не накапливаются погрешности задания оптимизируемого функционала или его (суб)градиента. Для негладких задач в этом плане давно известны методы с использованием δ -субградиентов (см., например, [Поляк, 1983]). Для задач более высокого уровня гладкости (градиент целевого функционала удовлетворяет условию Липшица или Гельдера) в последние годы разрабатываются детерминированные и стохастические методы, основанные на концепции неточного оракула (см., например, [Devolder et al., 2014; Dvurechensky, Gasnikov, 2016]). Концепция неточного оракула позволяет, в частности, вместо градиента целевого функционала рассматривать его некоторый сдвиг, связанный с неточностью вычисления. При этом неточность вычисления градиента существенно не ухудшает оценки скорости сходимости для неускоренного градиентного метода [Devolder et al., 2014].

Мы же рассмотрим в настоящей статье предложенный Ю.Е. Нестеровым (см. упражнение 1.5 из пособия [Гасников, 2018]) неполноградиентный метод минимизации липшицевой

выпуклой функции двух переменных на квадрате, позволяющий получить детерминированные оценки. Особенность метода — сведение задачи минимизации к постепенному сужению области определения за счет учета направления (суб)градиентов целевой функции вблизи точек-решений вспомогательных одномерных задач. Иными словами, предлагается комбинировать быстро работающие методы одномерной минимизации и с учетом геометрической структуры области определения выяснять направление субградиентов в точках на каждой итерации. Важно при этом, что на каждой итерации метода два раза находится направление (суб)градиента (а не самого (суб)градиента!), что существенно удешевляет стоимость итерации по сравнению с полноградиентным методом. При этом сохраняется линейная скорость сходимости метода. Метод при этом достаточно прост, что приводит к его устойчивости по отношению к погрешностям решения вспомогательных задач, а также вычисления градиентов на итерациях метода (теорема 1 и замечание 2). Здесь есть некоторое сходство с устойчивостью к погрешностям вычисления градиента целевого функционала неускоренного градиентного метода [Devolder et al., 2014].

Основной результат работы — оценка скорости сходимости метода Ю. Е. Нестерова для гладких выпуклых липшицевых функций, имеющих липшицев градиент (теорема 1). Показано, что метод может быть использован и в случае вычисления (направления) градиента с некоторой погрешностью (замечание 2). При этом оказывается, что может теряться свойство сходимости метода по аргументу (пример 1). Отмечено, что свойство липшицевости градиента достаточно потребовать не на всем квадрате, а лишь на некоторых отрезках внутри этого квадрата (замечание 4). Показано, что метод может эффективно работать и для некоторых негладких функций (см. рис. 8). Однако построен пример простой негладкой функции, для которой при неверном выборе субградиента может не наблюдаться сходимость даже по функции (пример 2). Проведено сравнение работы метода Ю. Е. Нестерова, метода эллипсоидов и градиентного спуска для гладких выпуклых функций различного вида (рис. 9 и рис. 10). Проведенные эксперименты показали, что рассматриваемый метод может позволить достичь требуемой точности решения задачи за меньшее (в сравнении с другими рассмотренными методами) время как в случае известного точного решения, так и при использовании оценок на скорость сходимости методов. Замечено, что при увеличении точности искомого решения время работы метода Ю. Е. Нестерова может расти медленнее, чем время работы метода эллипсоидов (рис. 11).

Отметим, что норму и скалярное произведение на плоскости всюду далее мы считаем стандартными.

Описание метода

Пусть имеется квадрат $\Pi \subset \mathbb{R}^2$ со стороной R и на нем задана функция $f : \Pi \rightarrow \mathbb{R}$. Опишем итерацию предлагаемого метода (см. [Гасников, 2018, упражнение 1.5]) как для гладких, так и негладких выпуклых целевых функционалов. Стоит сказать, что метод может расходиться на негладкой функции (см. пример 2 далее). При этом отметим, что для негладких выпуклых функционалов вместо обычного градиента используется *субградиент* $v \in \mathbb{R}^2$, определяемый неравенством

$$f(x) - f(x_0) \geq \langle v, x - x_0 \rangle,$$

справедливым для всех x из области определения. Отметим, что субградиентные методы широко используются в оптимизации (см., например, [Гасников, 2018; Поляк, 1983; Шор, 1979; Nesterov, 2018]). Опишем итерацию предлагаемого метода (см. [Гасников, 2018, упражнение 1.5]).

1. Через центр имеющегося квадрата проводится горизонтальная прямая. На отрезке, высекаемом из квадрата этой прямой, с точностью δ по аргументу решается задача одномерной оптимизации. Далее при реализации для решения одномерной задачи мы используем метод золотого сечения.

2. В найденной точке x_δ вычисляется вектор (суб)градиента $\nabla f(x_\delta)$ и определяется, в сторону какого из двух прямоугольников направлен вектор $\nabla f(x_\delta)$, и этот прямоугольник исключается из рассмотрения. При этом важно, что достаточно знать именно направление, а не точное значение (суб)градиента.

3. Через центр оставшегося прямоугольника проводится вертикальная прямая, на отрезке, высекаемом этой прямой в прямоугольнике, также с точностью δ по аргументу решается задача одномерной оптимизации. В найденной точке вычисляется вектор (суб)градиента функции и определяется, в сторону какого из двух квадратов он направлен. Этот квадрат исключается из рассмотрения.

Если вектор (суб)градиента в точке-приближении решения вспомогательной одномерной задачи нулевой, то процесс можно остановить и выдать указанную точку (она и даст точное решение).

Также возможно, что (суб)градиент в данной точке будет направлен вдоль отрезка, на котором решается вспомогательная одномерная задача, и неясно, какую часть оставшейся фигуры отсекать. В таком случае условимся отсекать любую из частей на выбор.

Обоснование оценки скорости сходимости

Легко понять, что на каждой итерации метода Ю. Е. Нестерова линейные размеры квадрата уменьшаются вдвое, и в какой-то момент оставшийся квадрат будет настолько мал, что в силу условия Липшица значения целевого функционала в его точках будут достаточно близкими друг к другу и можно выбирать любую точку. Мы докажем, что при подходящем выборе точности решения вспомогательных одномерных задач можно гарантировать, что значения целевой функции в точках оставшегося квадрата будут достаточно близки к оптимальному.

Теорема 1. Пусть на некотором квадрате $\Pi \subset \mathbb{R}^2$ со стороной R задана выпуклая дифференцируемая функция $f: \Pi \rightarrow \mathbb{R}$, для которой при любых $x, y \in \Pi$

- 1) $|f(x) - f(y)| \leq L\|x - y\|$,
- 2) $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$.

Тогда для всякого $\varepsilon > 0$ после

$$n = \log_2 \frac{2LR\sqrt{2}}{\varepsilon}$$

итераций метода Ю. Е. Нестерова при условии, что вспомогательные одномерные задачи решаются с точностью по аргументу

$$\delta = \frac{\varepsilon}{2MR(\sqrt{2} + \sqrt{5})\left(1 - \frac{\varepsilon}{LR\sqrt{2}}\right)},$$

получим квадрат $K_n \subset \Pi = K_0$ такой, что

$$f(x) - \min_{x \in \Pi} f(x) \leq \varepsilon \quad \forall x \in K_n.$$

Доказательство.

1. На каждом шаге метода линейные размеры квадрата уменьшаются в два раза. Поэтому после n шагов метода длина диагонали оставшегося квадрата K_n будет равна $\frac{R\sqrt{2}}{2^n}$. Из условия Липшица для f имеем

$$f(x) - \min_{x \in K_n} f(x) \leq L\|x - \arg \min_{x \in K_n} f(x)\| \quad \forall x \in K_n.$$

Наибольшее возможное расстояние между точками в полученном квадрате равно длине его диагонали. Это позволяет оценить правую часть неравенства сверху:

$$f(x) - \min_{x \in K_n} f(x) \leq \frac{LR\sqrt{2}}{2^n} \quad \forall x \in K_n. \quad (2)$$

Тогда получаем критерий останова метода:

$$\frac{LR\sqrt{2}}{2^n} \leq \frac{\varepsilon}{2} \quad \text{или} \quad n \geq \log_2 \frac{2LR\sqrt{2}}{\varepsilon}. \quad (3)$$

2. Пусть после некоторого количества итераций метода мы рассматриваем квадрат P со стороной R_0 . На текущей итерации он разделяется на два меньших прямоугольника, Q и Z , причем на разделяющем отрезке l решается задача одномерной оптимизации f (рис. 1).

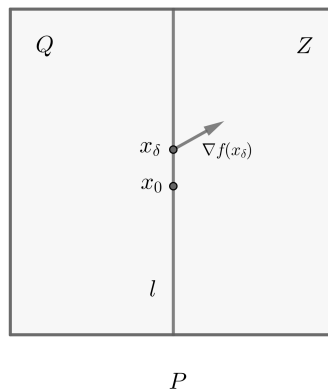


Рис. 1. Решение вспомогательной задачи оптимизации на квадрате P . Q, Z — прямоугольники, на которые был разделен квадрат на первом шаге рассматриваемой итерации, x_0 — точное решение задачи одномерной оптимизации на разделяющем отрезке, x_δ — δ -приближение решения

Как известно, для ее решения x_0 можно найти приближение x_δ с любой сколь угодно малой точностью по аргументу δ (используя, например, метод золотого сечения):

$$\|x_\delta - x_0\| \leq \delta.$$

3. Покажем, как в ходе одной итерации метода (отсечение двух частей исходного квадрата) может накапливаться погрешность.

1) Обозначим через x_0 решение вспомогательной задачи

$$f(x) \rightarrow \min_{x \in l}.$$

Ввиду свойства Липшица для градиента f : $\|\nabla f(x) - \nabla f(y)\| \leq M\|x - y\|$ имеем

$$\|\nabla f(x_\delta) - \nabla f(x_0)\| \leq M\delta. \quad (4)$$

Допустим, что $\nabla f(x_\delta)$ направлен в сторону Z и x_δ не совпадает ни с одним из концов отрезка l . Докажем, что

$$\min_{x \in Q} f(x) - \min_{x \in P} f(x) \leq \frac{MR\delta\sqrt{5}}{2}. \quad (5)$$

Согласно лемме Ферма (не уменьшая общности рассуждений, мы считаем, что решение вспомогательной задачи не совпадает с концом отрезка)

$$x_0 = \arg \min_{x \in l} f(x) \Rightarrow \frac{\partial f}{\partial x_i} = 0.$$

Тогда $\nabla_i f(x_0) = 0$, откуда

$$\begin{cases} \nabla f(x_0) = \bar{0}, \\ \nabla f(x_0) \perp l. \end{cases}$$

- 2) Пусть $\nabla f(x_0) = \bar{0}$. Тогда $x_0 \in l \subset Q$ — точка локального (а вследствие выпуклости f и глобального) минимума функции f на P . В таком случае

$$\min_{x \in Q} f(x) = \min_{x \in P} f(x). \quad (6)$$

- 3) Если ненулевой вектор $\nabla f(x_0)$ направлен в сторону Z , то также верно (6), так как производная по всякому направлению, лежащему в Z , в таком случае будет неотрицательной.
- 4) Если ненулевой вектор $\nabla f(x_0)$ направлен в сторону Q , то векторы $\nabla f(x_0)$ и $\nabla f(x_\delta)$ направлены в разные части делимого прямоугольника. Также вектор $\nabla f(x_0)$ перпендикулярен l , и поэтому векторы $\nabla f(x_0)$ и $\nabla f(x_\delta)$ образуют тупой угол (рис. 2).

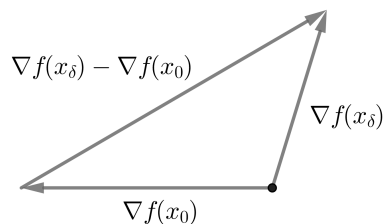


Рис. 2. Тупоугольный треугольник, образованный векторами градиента функции в точках x_0 (точное решения вспомогательной задачи) и x_δ (приближенное решение)

В тупоугольном треугольнике, который образован этими векторами, против большего угла лежит большая сторона. Поэтому

$$\|\nabla f(x_\delta)\| \leq \|\nabla f(x_\delta) - \nabla f(x_0)\| \stackrel{(4)}{\leq} M\delta.$$

Тогда

$$\|\nabla f(x_\delta)\| \leq M\delta. \quad (7)$$

Пусть $x_* = \arg \min_{x \in Z} f(x)$. Тогда

$$\begin{aligned} f(x_*) - f(x_\delta) &\geq \langle \nabla f(x_\delta), x_* - x_\delta \rangle \Rightarrow \\ \Rightarrow f(x_\delta) - f(x_*) &\leq \langle -\nabla f(x_\delta), x_* - x_\delta \rangle \leq |\langle -\nabla f(x_\delta), x_* - x_\delta \rangle|. \end{aligned}$$

По неравенству Коши – Буняковского имеем

$$f(x_\delta) - f(x_*) \leq \|\nabla f(x_\delta)\| \cdot \|x_* - x_\delta\|.$$

Известно, что x_* находится в одном из двух прямоугольников, на которые был поделен исходный прямоугольник. Диагональ этого прямоугольника (длина которой есть также и наибольшее возможное расстояние между точками в нем) равна $\frac{R_0 \sqrt{5}}{2}$. То есть

$$\|x_\delta - x_*\| \leq \frac{R_0 \sqrt{5}}{2}.$$

Так как в x_δ достигается минимум $f(x)$ на Q , а в x_* — минимум $f(x)$ на P , то верно следующее неравенство:

$$\min_{x \in Q} f(x) - \min_{x \in P} f(x) \leq \frac{MR\delta \sqrt{5}}{2}.$$

- 5) Если точка x_δ находится на одном из концов l и градиент в этой точке направлен в сторону Z , то все производные по направлениям, отклоняющимся от направления градиента не более чем на 90 градусов, будут неотрицательны. Таким образом получаем, что на полуплоскости, отделяемой прямой, проходящей через конец отрезка и содержащей Z , значения функции не меньше, чем в самом конце отрезка. Тогда

$$\min_{x \in Q} f(x) = \min_{x \in P} f(x).$$

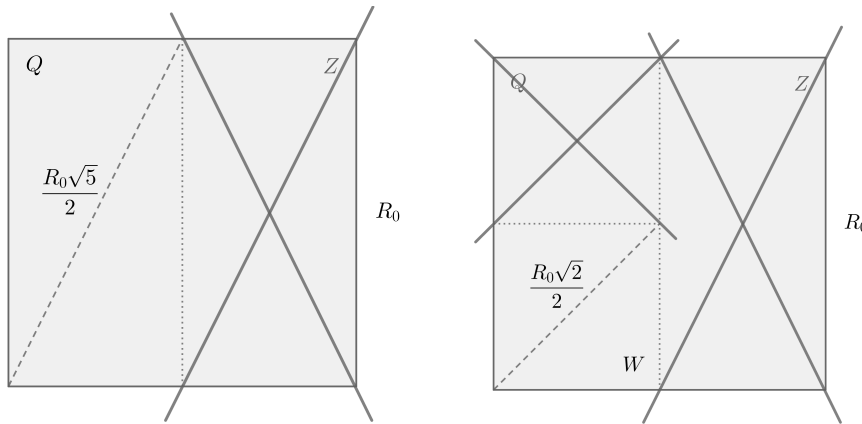


Рис. 3. Пошаговое исключение частей рассматриваемого квадрата со стороной R_0 на некоторой итерации: Z, Q — прямоугольники (длина диагонали равна $\frac{R_0\sqrt{5}}{2}$), на которые делится квадрат на первом шаге итерации, W — квадрат (длина диагонали равна $\frac{R_0\sqrt{2}}{2}$), оставшийся после исключения части прямоугольника Q на втором шаге

4. Пусть на некоторой итерации имеем квадрат P со стороной R_0 (рис. 3). На первом шаге из рассмотрения исключается один из прямоугольников, на которые был разделен данный квадрат. Длина диагонали оставшегося прямоугольника равна $\frac{R_0\sqrt{5}}{2}$. При этом ввиду (5) имеем

$$\min_{x \in Q} f(x) - \min_{x \in P} f(x) \leq M\delta \cdot \frac{R_0\sqrt{5}}{2}.$$

На втором шаге пополам делится уже оставшийся прямоугольник, в результате чего получаем квадрат, длина диагонали которого равна $\frac{R_0\sqrt{2}}{2}$. Соответственно,

$$\min_{x \in W} f(x) - \min_{x \in Q} f(x) \leq M\delta \cdot \frac{R_0\sqrt{2}}{2}.$$

5. На каждой итерации метода линейные размеры квадрата уменьшаются вдвое, и поэтому на i -й итерации длина стороны квадрата будет равна $R \cdot \frac{1}{2^{i-1}}$. Каждая же итерация предполагает

отсечение сначала прямоугольника, а затем — квадрата. Просуммировав оценки накапливающихся погрешностей (см. пункты 3 и 4 выше), получим

$$\min_{x \in K_n} f(x) - \min_{x \in \Pi} f(x) \leq M\delta \cdot \left(1 + \frac{1}{2} + \dots + \frac{1}{2^{n-1}}\right) R \cdot \left(\frac{\sqrt{5}}{2} + \frac{\sqrt{2}}{2}\right).$$

С учетом формулы суммы первых n членов геометрической прогрессии

$$\sum_{i=0}^{n-1} \frac{1}{2^i} = \frac{1 - \frac{1}{2^n}}{1 - \frac{1}{2}} = 2 \left(1 - \frac{1}{2^n}\right)$$

получаем

$$\min_{x \in K_n} f(x) - \min_{x \in \Pi} f(x) \leq MR\delta (\sqrt{2} + \sqrt{5}) \left(1 - \frac{1}{2^n}\right). \quad (8)$$

6. Если потребовать, чтобы выполнялось неравенство

$$\min_{x \in K_n} f(x) - \min_{x \in \Pi} f(x) \leq \varepsilon/2, \quad (9)$$

то ввиду (3) и (8) получим

$$MR\delta (\sqrt{2} + \sqrt{5}) \left(1 - \frac{\varepsilon}{LR\sqrt{2}}\right) \leq \frac{\varepsilon}{2},$$

т. е. необходимая точность решения вспомогательных одномерных задач оптимизации:

$$\delta \leq \frac{\varepsilon}{2MR (\sqrt{2} + \sqrt{5}) \left(1 - \frac{\varepsilon}{LR\sqrt{2}}\right)}. \quad (10)$$

Тогда (3) и (9) означают, что

$$f(x) - \min_{x \in \Pi} f(x) \leq \varepsilon \quad \forall x \in K_n. \quad (11)$$

□

ЗАМЕЧАНИЕ 1. Из представленного выше результата по стандартной схеме рассуждений (см., например, [Поляк, 1983]) можно вывести оценку скорости сходимости метода по аргументу для сильно выпуклых целевых функционалов.

ЗАМЕЧАНИЕ 2. Важно отметить, что схема доказательства основного результата (теорема 1) позволяет учитывать не только погрешность решения вспомогательных одномерных задач, но и погрешность вычисления градиента в точках x_δ (см. пп. 3 и 4 доказательства теоремы 1). Точнее говоря, в точках x_δ вектор градиента функции может вычисляться с точностью Δ : если $\nabla f(x_\delta)$ — точная величина градиента, то нам доступен некоторый вектор $v(x_\delta)$, для которого

$$\|v(x_\delta) - \nabla f(x_\delta)\| \leq \Delta.$$

В том случае, если вектор $v(x_\delta)$ направлен в ту же часть делимого квадрата, что и вектор $\nabla f(x_\delta)$, все рассуждения п. 3 доказательства теоремы 1 остаются неизменными, так как исключается та же часть квадрата, что и при $\Delta = 0$. Если же это не так, то либо вектор $\nabla f(x_\delta)$ направлен в ту часть, где лежит x^* — точное решение задачи, а вектор $v(x_\delta)$ — в иную (очевидно, что в результате будет исключена часть, которой не принадлежит x^* , так что рассуждения в п. 3 также остаются справедливыми), либо же вектор $v(x_\delta)$ в ту часть, где лежит x^* , а $\nabla f(x_\delta)$ — в иную (рис. 4).

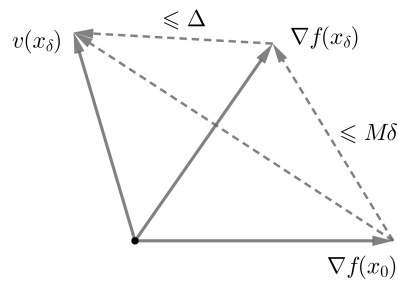


Рис. 4. Треугольники, образованные векторами градиента функции в точках x_0 точного решения вспомогательной задачи и x_δ (δ -приближенное решение), а также вектором $v(x_\delta)$, который есть Δ -приближение градиента функции в точке x_δ

В таком случае имеем треугольник, образованный векторами $\nabla f(x_\delta) - \nabla f(x_0)$ и $v(x_\delta) - \nabla f(x_\delta)$. Из неравенства треугольника следует, что если $\|\nabla f(x_\delta) - \nabla f(x_0)\| \stackrel{(4)}{\leq} M\delta$ и $\|v(x_\delta) - \nabla f(x_\delta)\| \leq \Delta$, то

$$\|v(x_\delta) - \nabla f(x_0)\| \leq M\delta + \Delta.$$

Векторы $v(x_\delta)$ и $\nabla f(x_0)$ образуют тупоугольный треугольник, причем длина его наибольшей стороны $\|v(x_\delta) - \nabla f(x_0)\| \leq M\delta + \Delta$. Следовательно, $\|v(x_\delta)\| \leq M\delta + \Delta$. Из треугольника, образуемого векторами $v(x_\delta)$ и $\nabla f(x_\delta)$ ввиду неравенства треугольника, имеем

$$\|\nabla f(x_\delta)\| \leq \|v(x_\delta)\| + \|v(x_\delta) - \nabla f(x_\delta)\| \leq M\delta + 2\Delta.$$

Далее, следуя схеме рассуждений доказательства теоремы 1, получим

$$(M\delta + 2\Delta) \cdot R(\sqrt{2} + \sqrt{5}) \left(1 - \frac{\varepsilon}{LR\sqrt{2}}\right) \leq \frac{\varepsilon}{2},$$

откуда

$$2\Delta + M\delta \leq \frac{\varepsilon}{2R(\sqrt{2} + \sqrt{5}) \left(1 - \frac{\varepsilon}{LR\sqrt{2}}\right)}.$$

Таким образом, метод можно применить при наличии двух типов погрешностей: при решении вспомогательных одномерных задач оптимизации и при вычислении (направлений) градиентов. Эти два типа погрешностей аккумулируют итоговую погрешность. При увеличении одной из погрешностей можно уменьшать значение другой так, чтобы сохранялось последнее неравенство, гарантирующее достижение приемлемого качества решения.

Отметим, что при этом можно построить такой пример функции, что для нее даже при небольшой погрешности вычисления градиента может не наблюдаться свойство сходимости метода по аргументу.

ПРИМЕР 1. Рассмотрим на квадрате $[0; 1]^2$ функцию

$$f(x_1, x_2) = x_1 - 0.001x_2,$$

минимум которой достигается в точке $(0; 1)$. Градиент данной функции постоянен и равен $\nabla f(x) = (1; -0.001)$, причем вторая координата мала по модулю. Если при некотором Δ эта компонента будет найдена с погрешностью большей 0.001, то может оказаться, что она станет положительной, и полученный вектор будет направлен в прямоугольник, содержащий точку точного решения задачи, в результате чего он будет исключен из рассмотрения. Из-за этого после удаления прямоугольника, содержащего точное решение задачи на первой итерации, метод не будет сходиться по аргументу.

Тем не менее по функции будет достигнуто необходимое качество решения, поскольку ввиду малого коэффициента при x_2 слагаемое $0.001x_2$ не будет значительно влиять на значение функции. На рис. 5 рассмотрен похожий пример нелинейной целевой функции.

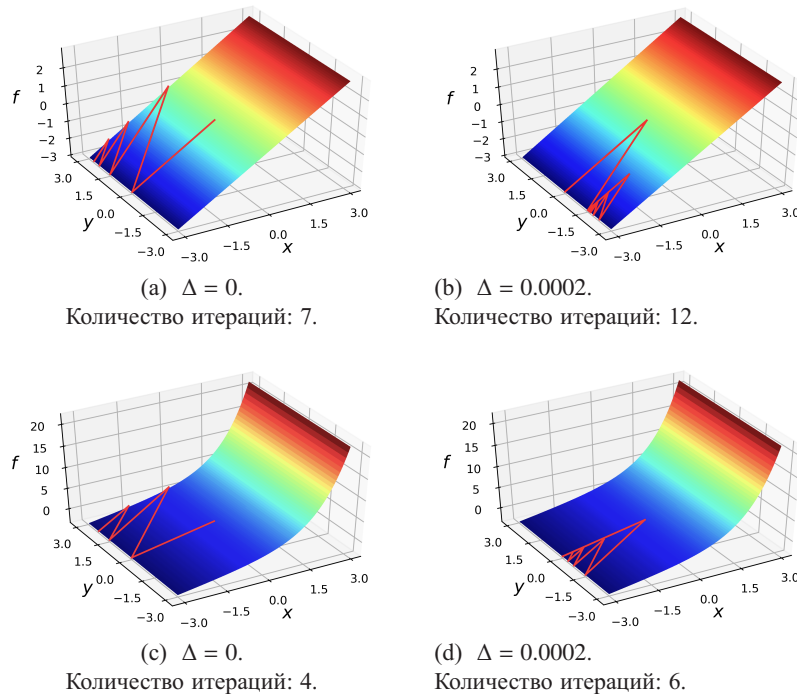


Рис. 5. Результат работы метода Ю. Е. Нестерова для функций $f(x_1, x_2) = x_1 - 0.0001x_2$ (рис. 5, а и 5, b) и $g(x_1, x_2) = e^{x_1} + x_1 - 0.0001x_2$ (рис. 5, c и 5, d) с различными значениями наибольшей погрешности Δ вычисления градиента в точках приблизительного решения вспомогательных одномерных задач. При определенном значении Δ на первой итерации метода может быть исключен прямоугольник, содержащий точное решение задачи. Критерий останова — достижение заданной точности по функции

Функции $f(x_1, x_2)$ и $g(x_1, x_2)$ определены на квадрате с вершинами в точках $(-3; -3)$, $(-3; 3)$, $(3; -3)$ и $(3; 3)$, их наименьшие значения $f_{\min} = -3.0003$ и $g_{\min} \approx -2.95$ достигаются в точке $(-3; 3)$. Как видно из графиков, ввиду ошибки при вычислении градиента метод на первой итерации может исключить из рассмотрения прямоугольник, содержащий точное решение задачи. Это в результате приводит к тому, что остановка произойдет в точке, довольно далекой от точного решения. Однако при этом необходимая точность по функции достигается.

ЗАМЕЧАНИЕ 3. Функция, рассмотренная в примере 1, обладает также тем свойством, что константа Липшица ее градиента равна 0, что позволяет выбирать в качестве значения M любую сколь угодно малую величину. Раз $\delta \sim \frac{1}{M}$ (см. (10)), то в этом случае погрешность решения одномерной подзадачи по аргументу может быть сколь угодно большой. То есть для функций (в том числе и нелинейных), для которых значение M на данном квадрате достаточно мало, может оказаться допустимым вообще не решать вспомогательные задачи на разделяющих отрезках.

О работе метода Ю. Е. Нестерова для негладких функций

Итак, мы обосновали сходимость за линейное время метода Ю. Е. Нестерова в классе гладких функций двух переменных с липшицевым градиентом.

ЗАМЕЧАНИЕ 4. На самом деле условие Липшица для градиента в доказательстве теоремы 1 было использовано лишь для точек, лежащих на высекаемых отрезках, вдоль которых рассматривались задачи одномерной минимизации (теорема 1, п. 3 доказательства). Если известна точность ε для желаемого решения, а также константа Липшица L функционала f , то можно определить число итераций n метода Ю. Е. Нестерова. По этому числу n можно построить $2 \cdot (2^n - 1)$ прямых — $2^n - 1$ вертикальных и то же количество горизонтальных, вдоль которых, возможно, будут решаться вспомогательные задачи (рис. 6).

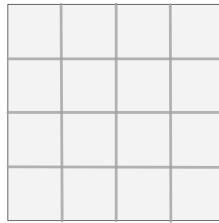
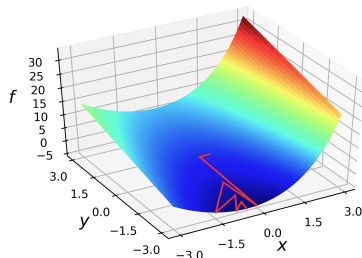


Рис. 6. Отрезки внутри исходного квадрата, вдоль которых, возможно, будут решаться вспомогательные задачи. Для сохранения оценок скорости сходимости метода достаточно потребовать выполнение условия липшицевости градиента целевого функционала только вдоль данных отрезков

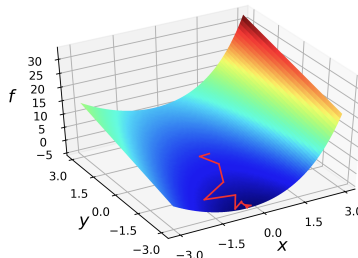
Условие Липшица для градиента целевого функционала достаточно потребовать только в точках построенных отрезков. Это, в частности, означает, что вне точек этих отрезков функция даже может иметь точки негладкости.

ЗАМЕЧАНИЕ 5. Условие липшицевости градиента используется лишь для отрезков, вдоль которых решаются вспомогательные задачи оптимизации, и в силу того, что эти отрезки параллельны осям координат, допустимо использовать условие Липшица не для полного градиента, а лишь для одной из частных производных функции по соответствующей переменной. Если окажется, что по одной из переменных функция линейна (рис. 7), то при решении вспомогательных задач константа Липшица соответствующей частной производной будет равна нулю, что означает, что для некоторых вспомогательных задач δ может быть сколь угодно велико, и эти задачи не требуют решения.

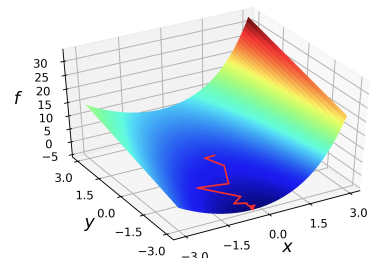
Проиллюстрируем это наблюдение на следующем примере.



(а) Вспомогательные задачи решаются с точностью δ , зависящей от константы Липшица для полного градиента.
Время работы: 8.04 мс.
Количество итераций: 7.



(б) Вспомогательные задачи не решаются (случайный выбор точки на отрезке). Критерий останова — достижение необходимой точности по функции.
Время работы: 5.91 мс.
Количество итераций: 14.



(с) Вспомогательные задачи не решаются. Критерий останова определяется оценкой на время работы.
Время работы: 6.09 мс.
Количество итераций: 14.

Рис. 7. Результат работы метода Ю. Е. Нестерова для функции $f(x_1, x_2) = e^{x_1} + x_1^2 + 2x_2$. Время работы и количество итераций метода зависят от того, производится ли решение вспомогательных задач на отрезках, вдоль которых константа Липшица градиента $M = 0$ и точность решения вспомогательных задач δ могут быть сколь угодно велики

Естественный интерес представляет вопрос о возможности обобщения теоремы 1 на случай негладкого выпуклого липшицева функционала f . Оказывается, что можно привести пример, для которого такая схема будет эффективно работать (см. рис. 8). Напомним, что для гладких функций f_1 и f_2 функция

$$f(x) = \max\{f_1(x), f_2(x)\}$$

будет уже негладкой и по теореме Демьянова–Данскина субградиентом f может считаться градиент $\nabla f_1(x)$ при $f_1(x) \geq f_2(x)$ и, наоборот, $\nabla f_2(x)$ при $f_2(x) \geq f_1(x)$.

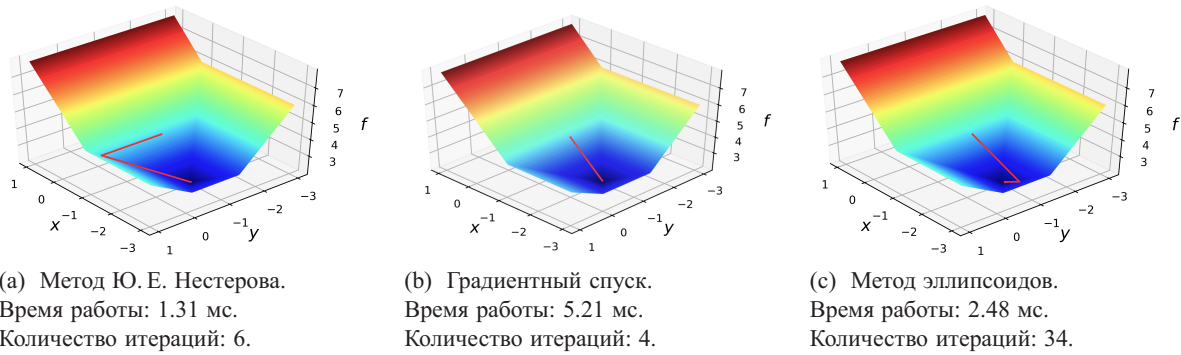


Рис. 8. Сравнение результатов работы метода Ю. Е. Нестерова и некоторых распространенных методов оптимизации на негладкой функции $g(x_1, x_2) = \max\{2x_1 + 6; -x_1; 1.2x_2 + 3.2; -2x_2\}$. Критерий останова для каждого из методов определяется в соответствии с теоретическими оценками

Тем не менее возможен простой пример негладкой выпуклой функции, для которой не будет сходимости метода Ю. Е. Нестерова даже по функции.

ПРИМЕР 2. Рассмотрим на квадрате с вершинами в точках $(0; 0)$, $(0; 1)$, $(1; 0)$ и $(1; 1)$ функцию

$$f(x_1, x_2) = |x_1 - x_2| + 0.9x_1.$$

В качестве первой разделяющей прямой выберем $x_2 = 1/2$. Тогда при $x_1 < 1/2$ имеем $f(x_1, 1/2) = x_2 - 0.1x_1$, а при $x_1 \geq 1/2$ получаем $f(x_1, 1/2) = 1.1x_1 - x_2$. Поэтому

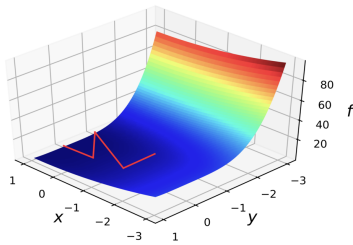
$$\nabla f(x_1, 1/2) = \begin{cases} (-0.1; 1), & \text{если } x_1 < 1/2, \\ (1.1; -1), & \text{если } x_1 > 1/2. \end{cases}$$

Иными словами, в зависимости от выбора точки приближения решения (и даже при точном решении) вспомогательной задачи (суб)градиент может быть направлен в любую из полуплоскостей относительно $x_2 = 1/2$ (отметим, что $\min_{0 \leq x_1 \leq 1} f(x_1, 1/2) = f(1/2, 1/2)$). Вместе с тем при отбрасывании части квадрата, для которой $x_2 < 1/2$, наименьшее значение целевой функции f в оставшейся части будет $f(1/2, 1/2) = 9/20 > f(0, 0) = 0$.

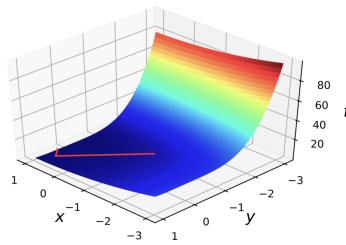
Эксперименты

Основное преимущество метода Ю. Е. Нестерова состоит в том, что данный метод является неполноградиентным, то есть не требует для своей работы вычисления полного значения градиента функции на каждой итерации. При этом некоторым недостатком может быть необходимость знать константы L и M (или их верхние оценки) для целевой функции. В то же время работа, например, метода эллипсоидов не зависит от уровня гладкости целевой функции. Также может оказаться не вполне удобным, что необходимая точность решения вспомогательных одномерных задач зависит от ε , что требует полного перезапуска метода при изменении точности решения задачи.

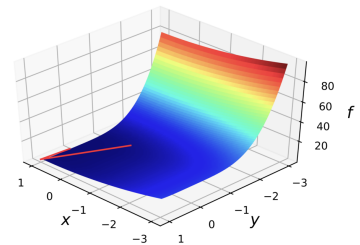
Рассмотрим гладкую выпуклую функцию $f(x_1, x_2) = (x_1 - 1)^2 + x_2^4$ на квадрате с вершинами в точках $(-3; -3)$, $(-3; 1)$, $(1; 1)$, $(1; -3)$. В данном эксперименте критерием останова методов будет достижение необходимой точности по функции $\varepsilon = 5 \times 10^{-3}$ в сравнении с минимумом функции на квадрате, достигающимся в точке $(1; 0)$.



(a) Метод Ю. Е. Нестерова.
Время работы: 2.78 мс.
Количество итераций: 4.



(b) Градиентный спуск.
Время работы: 7.53 мс.
Количество итераций: 5.



(c) Метод эллипсоидов.
Время работы: 4.05 мс.
Количество итераций: 2.

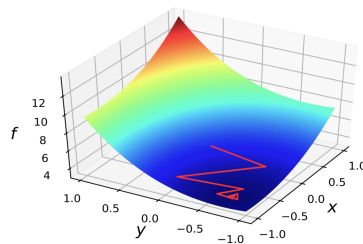
Рис. 9. Сравнение результатов работы метода Ю. Е. Нестерова и некоторых распространенных методов оптимизации на функции $f(x_1, x_2) = (x_1 - 1)^2 + x_2^4$. Критерий останова для каждого из методов — достижение необходимой точности по функции в сравнении с известным точным решением

Эксперименты показали (рис. 9), что в сравнении с другими рассматриваемыми методами (градиентным спуском и методом эллипсоидов) метод Ю. Е. Нестерова работает быстрее, несмотря на большее количество выполненных итераций. Это достигается за счет того, что другие рассмотренные полноградиентные методы тратят большее время на вычисление полного значения градиента функции. При этом рассматриваемый в работе метод Ю. Е. Нестерова предполагает для своей работы лишь вычисление направления (суб)градиента, что, вообще говоря, может удешевлять стоимость итерации.

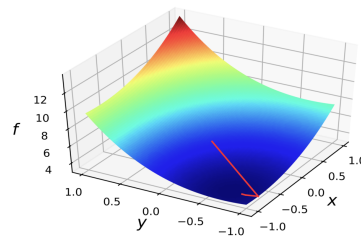
Сравним работу методов при отсутствии информации о точном решении задачи, используя имеющиеся оценки на количество итераций метода для достижения гарантированной точности ε . Рассмотрим функцию $f(x_1, x_2) = (x_1 + 1)^2 + x_2^2 - x_1 + e^{x_1} + e^{x_2 + 1}$ на квадрате $[-1, 1]^2$, достигающую наименьшего значения $f \approx 3.124$ в точке $(-0.74, -0.69)$. Значения необходимых констант для данной функции: $L = 10.993$, $M = 10.508$.

Результаты данного эксперимента показаны на рис. 10. При гарантированной точности $\varepsilon = 5 \times 10^{-2}$ с использованием метода Ю. Е. Нестерова достигнута точность 5×10^{-4} по функции и 2×10^{-2} по аргументу, тогда как метод эллипсоидов продемонстрировал точность 2.5×10^{-2} по функции и 1.23×10^{-1} по аргументу.

Если для данной задачи построить графики зависимости времени работы от логарифма точности для каждого из методов (рис. 11), то будет видно, что зависимость практически



(a) Метод Ю. Е. Нестерова.
Время работы: 6.42 мс.
Количество итераций: 11.



(b) Метод эллипсоидов.
Время работы: 15.7 мс.
Количество итераций: 42.

Рис. 10. Сравнение результатов работы метода Ю. Е. Нестерова и метода эллипсоидов на функции $f(x_1, x_2) = (x_1 + 1)^2 + x_2^2 - x_1 + e^{x_1} + e^{x_2 + 1}$. Критерий останова для метода эллипсоидов определяется в соответствии с теоретическими оценками (см., например, [Поляк, 1983])

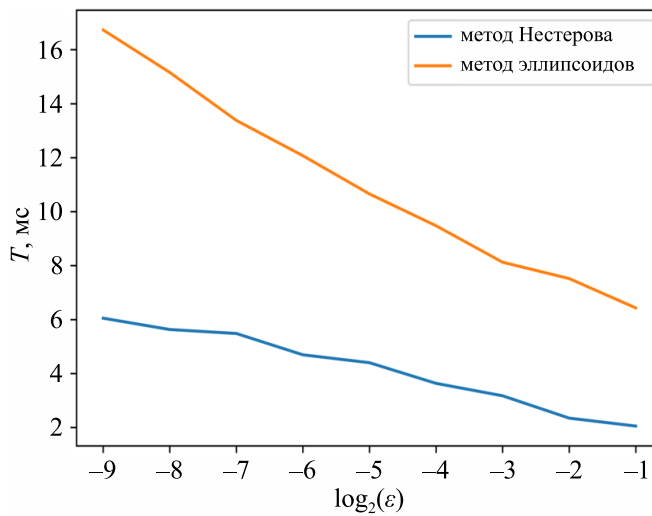


Рис. 11. Зависимость времени работы методов от требуемой точности решения задачи

линейна, и притом наклон линии, описывающей рост длительности работы метода эллипсоидов, значительно больше, чем наклон линии, соответствующей рассматриваемому нами методу Ю. Е. Нестерова.

Заключение

В работе получена оценка скорости сходимости одного неполноградиентного метода минимизации выпуклой липшицевой функции двух переменных на квадрате. Как показали результаты экспериментов, такой метод может работать существенно быстрее метода эллипсоидов.

Основной результат (теорема 1) предполагает, помимо условия Липшица для целевой функции, еще и условие липшицевости градиента. Показано, что метод может давать приемлемое качество решения за линейное время при условии достаточно малой погрешности для нахождения направления градиента (замечание 2). Однако, как показано далее, это требование можно существенно ослабить, потребовав его выполнение на отдельных отрезках (замечание 4). Очевидно, что если, например, количество точек негладкости целевого функционала конечно, то незначительным сдвигом разделяющих (на итерациях метода) квадрат отрезков можно исключить эти точки негладкости из рассмотрения. Однако нами предложен также и простой пример негладкой функции, для которой метод не сходится (пример 2).

Учитывая вышеизложенное, кажется полезной на будущее задача описания класса негладких выпуклых функций, в котором рассмотренный метод будет сходиться. Также весьма интересной была бы задача исследования эффективности предложенного подхода к условным задачам многомерной оптимизации с небольшим числом ограничений (которые за счет операции максимизации можно свести к двум). Заметим, что выполнение условия Слейтера позволяет компактифицировать значения двойственных переменных (см., например, § 4 из [Гасников, 2018]) и, по сути, рассматривать в качестве двойственной задачи двумерную задачу оптимизации (например, на квадрате).

Все численные эксперименты были выполнены с помощью программного обеспечения Anaconda 5.3.1 Python 3.6 version на компьютере с 2-ядерным процессором Intel Core i5-5250U с тактовой частотой 1.6 ГГц на каждое ядро. ОЗУ компьютера составляло 8 Гб.

Авторы благодарят Юрия Евгеньевича Нестерова за идею метода и полезные обсуждения.

Список литературы (References)

- Anaconda python дистрибутив, официальный веб-сайт [Электронный ресурс]: <https://www.anaconda.com> (дата обращения: 20.02.2019).
Anaconda python distribution official website [Electronic resource]: <https://www.anaconda.com> (accessed 20.02.2019).
- Баяндина А. С., Гасников А. В., Лагуновская А. А.* Безградиентные двухточечные методы решения задач стохастической негладкой выпуклой оптимизации при наличии малых шумов неслучайной природы // Автоматика и телемеханика. — 2018. — № 8. — С. 38–49.
Bayandina A. S., Gasnikov A. V., Lagunovskaya A. A. Gradient-free two-point methods for solving stochastic nonsmooth convex optimization problems with small non-random noises // Automation and Remote Control. — 2018. — Vol. 79, No. 8. — P. 1399–1408. (Original Russian paper: *Bayandina A. S., Gasnikov A. V., Lagunovskaya A. A. Bezgradientnye dvukhtochечные metody resheniya zadach stokhasticheskoi negladkoi vypukloi optimizatsii pri nalichii malykh шумов ne sluchainoi prirody // Automation and Remote Control. — 2018. — No. 8. — P. 38–49.*)
- Васильев Ф. П.* Методы оптимизации. — Т. 2. — М.: МЦНМО, 2011. — 433 с.
Vasiliev F. P. Metody optimizatsii [Methods of Optimization]. — Vol. 2. — Moscow: MCCME, 2011. — 433 p. (in Russian).
- Воронцова Е. А., Гасников А. В., Горбунов Э. А.* Ускоренные спуски по случайному направлению с неевклидовой прокс-структурой // Автоматика и телемеханика. — 2019. — № 4. — С. 126–143.
Vorontsova E. A., Gasnikov A. V., Gorbunov E. A. Uskorennые spuski po sluchainomu napravleniyu s neevklidovoi proks-strukturoi [Accelerated Directional Search with non-Euclidean prox-structure] // Automation and Remote Control. — 2019. — No. 4. — P. 126–143 (in Russian).
- Гасников А. В.* Современные численные методы оптимизации. Метод универсального градиентного спуска. — М.: МФТИ, 2018. — 240 с.
Gasnikov A. V. Sovremennye chislennye metody optimizatsii. Metod universal'nogo gradientnogo spuska [Modern numerical optimization methods. universal gradient descent method]. — Moscow: MIPT, 2018. — 240 p. (in Russian).
- Немировский А. С., Юдин Д. Б.* Сложность задач и эффективность методов оптимизации. — М.: Наука, 1979.
Nemirovski A. S., Yudin D. B. Problem Complexity and Method Efficiency in Optimization. — New York: Wiley-Interscience, 1983. (Russ. ed.: *Nemirovski A. S., Yudin D. B. Slozhnost' zadach i effektivnost' metodov optimizatsii. — Moscow: Nauka, 1979.*)
- Поляк Б. Т.* Введение в оптимизацию. — М.: Наука, 1983.
Polyak B. T. Introduction to Optimization. — New York: Optimization Software, 1987. (Russ. ed.: *Polyak B. T. Vvedenie v optimizatsiyu. — Moscow: Nauka, 1983.*)
- Шор Н. З.* Методы минимизации недифференцируемых функций и их приложения. — Киев: Наукова Думка, 1979.
Shor N. Z. Minimization Methods for Non-Differentiable Functions. — Springer Series in Computational Mathematics, 1985. (Russ. ed.: *Shor N. Z. Metody minimizatsii nedifferentsiruemykh funktsii i ikh prilozheniya. — Kiev: Naukova Dumka, 1979.*)
- Bayandina A., Dvurechensky P., Gasnikov A., Stonyakin F., Titov A.* Mirror descent and convex optimization problems with non-smooth inequality constraints // Large-Scale and Distributed Optimization. — P. Giselsson, A. Rantzer (eds.). — Chap. 8. — Springer International Publishing, 2018. — P. 181–215.
- Bogolubsky L., Dvurechensky P., Gasnikov A., Gusev G., Nesterov Yu., Raigorodskii A. M., Tikhonov A., Zhukovskii M.* Learning supervised pagerank with gradient-based and gradient-free optimization methods // Advances in Neural Information Processing Systems 29. — D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, R. Garnett (eds.). — Curran Associates, Inc., 2016. — P. 4914–4922.
- Bubeck S.* Convex optimization: algorithms and complexity // Foundations and Trends in Machine Learning. — 2015. — Vol. 8, No. 3–4. — P. 231–357.
- Chen Y., Lan G., Ouyang Y.* Optimal primal-dual methods for a class of saddle point problems // SIAM Journal on Optimization. — 2014. — Vol. 24, No. 4. — P. 1779–1814.

- Dang C. D., Lan G.* Stochastic Block Mirror Descent Methods for Nonsmooth and Stochastic Optimization // *SIAM J. on Optimization*. — 2015. — Vol. 25, No. 2. — P. 856–881.
- Devolder O., Glineur F., Nesterov Yu.* First-order methods of smooth convex optimization with inexact oracle // *Mathematical Programming*. — 2014. — Vol. 146, No. 1–2. — P. 37–75.
- Dvurechensky P., Gasnikov A.* Stochastic Intermediate Gradient Method for Convex Problems with Stochastic Inexact Oracle // *Journal of Optimization Theory and Applications*. — 2016. — Vol. 171, No. 1. — P. 121–145.
- Gasnikov A. V., Lagunovskaya A. A., Usmanova I. N., Fedorenko F. A.* Gradient-free proximal methods with inexact oracle for convex stochastic nonsmooth optimization problems on the simplex // *Automation and Remote Control*. — 2016. — Vol. 77, No. 11. — P. 2018–2034.
- Lee Y.-T., Sidford A., Wong S. C.-W.* A faster cutting plane method and its implications for combinatorial and convex optimization // E-print, 2015. — URL: <https://arxiv.org/pdf/1508.04874.pdf> (accessed 02.01.2019).
- Nemirovski A.* Lectures on modern convex optimization analysis, algorithms, and engineering applications. — Philadelphia: SIAM, 2015. — URL: http://www2.isye.gatech.edu/~nemirovs/Lect_ModConvOpt.pdf (accessed 02.01.2019).
- Nesterov Yu.* Primal-dual subgradient methods for convex problems // *Mathematical Programming*. — 2009. — Vol. 120, No. 1. — P. 221–259.
- Nesterov Yu., Shpirko S.* Primal-Dual Subgradient Method for Huge-Scale Linear Conic Problems // *SIAM Journal on Optimization*. — 2014. — Vol. 24, No. 3. — P. 1444–1457.
- Nesterov Yu.* Lectures on convex optimization. — Springer, 2018.
- Tappenden R., Richtárik P., Gondzio, J.* Inexact Coordinate Descent: Complexity and Preconditioning // *Journal of Optimization Theory and Applications*. — 2016. — Vol. 170, No. 1. — P. 144–176.