

УДК 519.85

Ф. С. Стонякин

Крымский федеральный университет им. В. И. Вернадского
Московский физико-технический институт (национальный исследовательский университет)

Адаптивные градиентные методы для некоторых классов задач негладкой оптимизации

Предложено несколько адаптивных алгоритмических методов, применимых к задачам негладкой выпуклой оптимизации. Первый из них основан на введении специальной искусственной неточности, и для его реализации предложен соответствующий аналог концепции абстрактной неточной модели целевой функции. Для этой концепции предложены аналоги градиентного метода, а также быстрого градиентного метода с адаптивной настройкой некоторых параметров неточной модели, и получена оценка качества найденного решения. Показано, что для негладких задач возможно модифицировать предложенные методы так, чтобы гарантированно выполнялась сходимость по функции со скоростью, близкой к оптимальной. Введена аналогичная концепция неточной модели для оператора поля вариационного неравенства, а также для седловой задачи и приведена оценка скорости сходимости соответствующего адаптивного варианта проксимального зеркального метода. Предложены аналоги субградиентных схем с переключениями для задач выпуклой оптимизации с ограничениями. При этом рассмотрены предположения, связанные с недавно предложенным условием относительной липшицевости. Это позволило выписать оценку качества решения с относительной точностью для задачи минимизации однородного выпуклого функционала при достаточно общих предположениях.

Ключевые слова: градиентный метод, быстрый градиентный метод, адаптивный метод, липшицев градиент, негладкая оптимизация, зеркальный спуск, относительная липшицевость, относительная точность.

F. S. Stonyakin

V. I. Vernadsky Crimean Federal University
Moscow Institute of Physics and Technology

Adaptive gradient methods for some classes of nonsmooth optimization problems

We propose several adaptive algorithmic methods for problems of non-smooth convex optimization. The first method is based on a special artificial inexactness. For its implementation the corresponding analogue of the concept of an abstract inexact model of the objective functional is proposed. For this concept, analogues of gradient and fast gradient methods with adaptive tuning of some parameters of this model are proposed and an assessment of the quality of the solution is obtained. It is shown that it is possible for nonsmooth problems to modify the proposed method to guarantee the convergence in the function with a close to optimal rate. A similar concept of an inexact model is introduced for variational inequalities and saddle point problems. Estimate of the convergence rate of the corresponding adaptive version of the proximal mirror method is presented. Analogues of switching subgradient schemes are proposed for convex programming problems. In this case, assumptions close to the recently proposed condition of the relative Lipschitz continuity are considered, which allows us to obtain an estimate of the quality of the solution with relative accuracy for the problem of minimizing a homogeneous convex functional using fairly general assumptions.

Key words: gradient method, fast gradient method, adaptive method, Lipschitz continuous gradient, nonsmooth optimization, mirror descent, relative Lipschitz continuity, relative accuracy.

Введение

Во многих прикладных задачах возникает необходимость подходящих алгоритмических методов для задач негладкой выпуклой оптимизации. Однако оценки эффективности таких процедур в случае большой размерности переменных весьма пессимистичны. Так, к примеру, ε -точное решение по функции задачи выпуклой негладкой оптимизации возможно достичь за $O(\varepsilon^{-2})$ обращений к подпрограмме нахождения (суб)градиента, и в общем случае такая оценка не улучшаема [1]. Для гладких задач оценки эффективности выше, что приводит к естественной идее для негладких задач обосновать возможность использования какого-нибудь приближения оптимизационной модели к гладкому случаю. Эту идею, в частности, реализуют так называемые *универсальные методы*, исследованию которых было положено начало в работе [2]. Универсальные градиентные методы основаны на построении для задач выпуклой оптимизации с гёльдеровым (суб)градиентом целевого функционала аналога стандартной квадратичной интерполяции с искусственно введённой погрешностью. Универсальность метода при этом понимается как возможность адаптивной настройки при работе метода на оптимальный в некотором смысле уровень гладкости задачи и величину, соответствующую константе Гёльдера L_ν (суб)градиента целевого функционала. Оказывается, что возможность такой настройки может позволить экспериментально для некоторых задач улучшить скорость сходимости по сравнению с оптимальными теоретическими оценками [2].

Искусственную неточность для негладких задач можно вводить по-разному [3, 4]. При этом естественно возникает проблема описания влияния погрешностей задания целевого функционала и градиента на оценки скорости сходимости методов. Для градиентных методов известен подход к этой проблеме, основанный на недавно предложенной концепции неточного оракула [1, 6]. Известно, что для неускоренных градиентных методов в оценках скорости сходимости не происходит накопления величин, связанных с погрешностями. Однако при этом для оптимального при отсутствии погрешностей на классе гладких выпуклых задач быстрого градиентного метода в итоговой оценке скорости сходимости величины погрешностей могут накапливаться. Концепция неточного оракула была обобщена в [2, 8], где были введены понятия (δ, L) -модели и (δ, L, μ) -модели целевой функции для задач оптимизации. Суть данных обобщений в замене $\langle \nabla f(x), y - x \rangle$ некоторой абстрактной выпуклой по первой переменной функцией $\psi(y, x)$, что позволяет рассматривать более широкий класс задач [8].

Различным методам градиентного типа посвящены все новые современные работы [2, 3, 8–10]. В частности, недавно в [9] введены условия относительной гладкости оптимизируемого функционала, которые предполагают замену условия липшицевости градиента на ослабленный вариант:

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + LV(y, x), \quad (1)$$

где $V(y, x)$ — широко используемый в оптимизации аналог расстояния между точками x и y , который называют *дивергенцией Брэгмана*. Обычно *дивергенция Брэгмана* вводится на базе вспомогательной 1-сильно выпуклой функции d (порождает расстояния), дифференцируемой во всех точках выпуклого замкнутого множества Q :

$$V(y, x) = d(y) - d(x) - \langle \nabla d(x), y - x \rangle \quad \forall x, y \in Q, \quad (2)$$

где $\langle \cdot, \cdot \rangle$ — скалярное произведение в \mathbb{R}^n . В частности, для стандартной евклидовой нормы $\|\cdot\|_2$ и расстояния в \mathbb{R}^n можно считать, что $V(y, x) = d(y - x) = \frac{1}{2}\|y - x\|_2^2$ для произвольных $x, y \in Q$. Однако часто возникает необходимость использовать и неевклидовы

нормы. Более того, рассмотренное в [3, 9] условие относительной гладкости предполагает лишь выпуклость (но не сильную выпуклость) порождающей функции d . Как показано в [3], концепция относительной гладкости позволяет применить вариант градиентного метода для некоторых задач, которые ранее решались лишь с помощью методов внутренней точки. В частности, речь идет об известной задаче построения оптимального эллипсоида, покрывающего заданный набор точек. Эта задача, в частности, представляет интерес для статистики и анализа данных. Отметим в этой связи также предложенный недавно в [12] подход к задачам негладкой оптимизации, связанный с релаксацией условия Липшица, которая предполагает замену ограниченности нормы субградиента $\|\nabla f(x)\|_* \leq M_f$ так называемой *относительной липшицевостью*:

$$\|\nabla f(x)\|_* \leq \frac{M_f \sqrt{2V(y, x)}}{\|y - x\|} \quad \forall x, y \in Q, y \neq x.$$

При этом порождающая функция d не обязательно сильно выпукла. В работе [12] предложены детерминированный и стохастический алгоритмы зеркального спуска для задачи минимизации выпуклого относительно липшицева целевого функционала. Отметим также, что в [13] показана оптимальность зеркального спуска для задач математического программирования с квазивыпуклыми дифференцируемыми гёльдеровыми целевыми функционалами и выпуклыми липшицевыми функционалами ограничений.

Предлагаемая статья посвящена в основном развитию упомянутых выше идей. В первом разделе предложена модификация концепции (δ, L) -модели целевой функции с переменной неточностью, которая позволяет учесть возможность неточного задания не только значения целевой функции, но и самой модели. В частности, для стандартной модели $\psi(y, x) = \langle \nabla f(x), y - x \rangle$ описывается ситуация погрешности задания целевого функционала f , а также погрешности задания (суб)градиента ∇f .

В первых трёх разделах работы мы выделяем (в модельной общности по аналогии с [2]) класс задач, для которых ($\tilde{\nabla} f$ — возмущённое значение (суб)градиента ∇f) верно следующее неравенство с параметрами $\delta_1, \delta_2, \gamma, \Delta \geq 0$:

$$\begin{aligned} f(x) + \langle \tilde{\nabla} f(x), y - x \rangle - \delta_2 - \gamma \|y - x\| &\leq f(y) \leq \\ &\leq f(x) + \langle \tilde{\nabla} f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 + \Delta \|y - x\| + \delta_1 \quad \forall x, y \in Q. \end{aligned} \quad (3)$$

Смысл такого обобщения заключается в том, что возможны различные значения параметров γ и Δ в (3) и влияние величин δ_1 и Δ на итоговое качество решения может быть уменьшено. Случай $\gamma > 0$ более подробно рассмотрен в [4], а в настоящей работе мы полагаем $\gamma = 0$. Это предположение вполне естественно, если рассматривать Δ как искусственную неточность для негладкой задачи. В этом случае Δ в неравенстве

$$f(x) + \langle \tilde{\nabla} f(x), y - x \rangle \leq f(y) \leq f(x) + \langle \tilde{\nabla} f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2 + \Delta \|y - x\| \quad (4)$$

вполне можно рассматривать как оценку скачков субдифференциалов f вдоль всевозможных векторных отрезков $[x; y]$ [3]. Несложно понять, что если f в (4) удовлетворяет условию Липшица с константой $M > 0$, то $\Delta \leq 2M$. Отметим, что возможна ситуация, когда Δ существенно меньше M [3]. Похожие на (4) условия рассматривались в [6, 14] для случая, когда f представим в виде суммы гладкого и негладкого целевого функционала. В настоящей работе рассмотрен более широкий класс целевых функционалов, которые не обязательно представимы в виде суммы гладкого и негладкого слагаемых. В итоге предложена общая концепция неточной модели целевой функции, которая могла бы описать все указанные выше ситуации. Далее в разделе 4 введена аналогичная концепция неточной модели для оператора поля вариационного неравенства, а также для седловой задачи. Выписаны оценки скорости сходимости соответствующего адаптивного варианта проксимального зеркального метода. В последнем разделе статьи рассмотрены аналоги субградиентных схем с переключениями для задач выпуклой оптимизации с ограничениями. При

этом рассмотрены предположения, близкие к недавно предложенному условию относительной липшицевости [12], что позволило получить оценку качества решения с относительной точностью для задачи минимизации однородного выпуклого функционала при достаточно общих предположениях.

Работа состоит из введения, пяти основных разделов и заключения.

В разделе 1 обобщено ранее предложенное понятие (δ, L) -модели целевой функции [2] в запрошенной точке и введена концепция (δ, Δ, L) -модели функции (определение 1). Введена также более общая концепция (δ, Δ, L, μ) -модели целевой функции, которую можно, в частности, применить на классе гладких сильно выпуклых функционалов. Предложен аналог градиентного метода (алгоритм 1) градиентного типа с адаптивной настройкой параметров неточности этой модели.

В разделе 2 предложен вариант быстрого градиентного метода (алгоритм 2) для задач выпуклой минимизации с адаптивным выбором шага и адаптивной настройкой на величины параметров (δ, Δ, L) -модели и получена оценка качества найденного решения.

Раздел 3 посвящен специальному классу задач выпуклой негладкой оптимизации, к которым применима концепция определения 1 ($\delta = 0, \Delta > 0$). Показано, что для таких задач возможно модифицировать алгоритм 2 так, чтобы гарантированно достигалось ε -точное решение задачи минимизации f за

$$O\left(\sqrt{\frac{L}{\varepsilon}}\right) + O\left(\frac{\Delta^2}{\varepsilon^2}\right)$$

обращений к (суб)градиенту целевого функционала. По сути, получен некоторый аналог результата ([6], п. 4.7.2), однако оптимизируемый функционал уже не обязательно имеет вид суммы гладкого и негладкого слагаемого и рассмотрена модельная общность. Отметим, что при этом для адаптивного ускоренного метода оценка может увеличиться на некоторый логарифмический множитель вида $O(\log_2(\varepsilon^{-3}))$, для адаптивного неускоренного — на множитель вида $O(\log_2(\varepsilon^{-1}))$. Отметим также, что параметр Δ можно понимать и как величину, которая соответствует неточно заданному (суб)градиенту.

Далее в разделе 4 показано, как можно вводить аналог концепции (δ, Δ, L) -модели оптимизируемого функционала, но уже для вариационных неравенств и седловых задач.

Наконец, в последнем разделе 5 обсуждаются альтернативные подходы к алгоритмическим схемам для задач негладкой оптимизации с липшицевыми выпуклыми функциональными ограничениями. Предложен адаптивный метод с заменой обычного условия Липшица на некоторые его аналоги. Рассмотрено условие, близкое к относительной липшицевости [12], которое не предполагает 1-сильной выпуклости дивергенции Брэгмана. Это позволило предложить метод, который гарантирует достижение качества приближённого решения с фиксированной относительной точностью для задачи минимизации выпуклого однородного целевого функционала f в случае, когда субдифференциал f в нулевой точке может не содержать 0 как внутреннюю точку.

Всюду далее через $\|\cdot\|$ обозначается норма в пространстве \mathbb{R}^n , а через $\|\cdot\|_*$ — норма в двойственном пространстве.

1. Концепция (δ, Δ, L) -модели функции в запрошенной точке и оценка скорости сходимости для градиентного метода

Введем анонсированный выше аналог понятия (δ, Δ, L) -модели целевой функции, который учитывает погрешность Δ задания градиента и применим также для задач с относительно гладкими целевыми функционалами [3].

Определение 1. Будем говорить, что f допускает (δ, Δ, L) -модель ψ в точке $x \in Q$, если для некоторой выпуклой по первой переменной функции $\psi(y, x)$, такой, что при всяком x $\psi(x, x) = 0$, будет верно неравенство

$$f(x) + \psi(y, x) - \delta_1 \leq f_\delta(x) + \psi(y, x) \leq f(y) \leq f_\delta(x) + \psi(y, x) + \delta_2 + \Delta \|y - x\| + LV(y, x) \quad (5)$$

для произвольных $x, y \in Q$ при фиксированных $\delta, \delta_1, \delta_2 \geq 0$ и $f_\delta(x) \in [f(x) - \delta, f(x)] \forall x \in Q$.

Всюду далее будем полагать, что $\delta = \delta_1 = \delta_2$, а также $f_\delta(x) \in [f(x) - \delta, f(x)]$. Отметим, что предложенная концепция есть модификация введенного в [4] понятия $(\delta, \Delta, \gamma, L)$ -модели целевого функционала в запрашиваемой точке.

Замечание 1. Для неускоренного метода возможно заменить левое неравенство в (5) на ослабленный вариант

$$f(x_*) \geq f(x) + \psi(x_*, x) - \delta, \quad (6)$$

где x_* — ближайшее к x решение задачи минимизации f с точки зрения дивергенции Брэгмана $V(x_*, x)$. Неравенство (6) будет, в частности, верно для задачи минимизации квазивыпуклой целевой функции при достаточно малой величине погрешности градиента.

Можно также ввести и аналог концепции (δ, L, μ) -оракула в оптимизации с переменной неточностью. При $\mu > 0$ эта концепция позволяет обосновать скорость сходимости предлагаемого нами метода, близкую к линейной.

Определение 2. Будем полагать, что f допускает (δ, Δ, L, μ) -модель $\psi(y, x)$, если для произвольных $x, y \in Q$ верно

$$f(y) \leq f_\delta(x) + \psi(y, x) + \Delta \|y - x\| + \delta + LV(y, x), \quad (7)$$

а также

$$f(x) - \delta + \psi(x_*, x) + \mu V(x_*, x) \leq f_\delta(x) + \psi(x_*, x) + \mu V(x_*, x) \leq f(x_*), \quad (8)$$

где x_* — ближайшее к x решение задачи минимизации f с точки зрения дивергенции Брэгмана $V(x_*, x)$. Неравенство (8) будет, в частности, верно для задачи минимизации сильно квазивыпуклой целевой функции [10] при достаточно малой величине погрешности градиента.

Пример 1. В качестве примера по аналогии с примером 1 из [4] отметим задачу сильно выпуклой композитной оптимизации $f(x) = g(x) + h(x) \rightarrow \min$, где g — гладкая выпуклая функция, а h — выпуклая не обязательно гладкая функция простой структуры. Если при этом для градиента ∇g задано его приближение $\tilde{\nabla} g$: $\|\tilde{\nabla} g(x) - \nabla g(x)\| \leq \Delta$, то можно положить

$$\psi(y, x) = \langle \tilde{\nabla} g(x), y - x \rangle + h(y) - h(x)$$

и в случае μ -сильной выпуклости g или h будет верно (7).

Для задачи минимизации функционала, допускающего (δ, Δ, L, μ) -модель в произвольной запрошенной точке, предложим такой метод.

Алгоритм 1 Адаптивный градиентный метод для функций, допускающих (δ, Δ, L, μ) -модель в запрошенной точке.

Require: $x^0 \in Q$ — начальная точка, $V(x_*, x^0) \leq R^2$, параметры $L_0, \Delta_0, \delta_0 > 0$: $2\mu < L_0 \leq 2L$, $\Delta_0 \leq 2\Delta$, $\delta_0 \leq 2\delta$.

1: $L_{k+1} := L_k/2$, $\Delta_{k+1} := \Delta_k/2$, $\delta_{k+1} := \delta_k/2$.

2: $x^{k+1} := \arg \min_{x \in Q} \{\psi(x, x^k) + L_{k+1}V(x, x^k)\}$.

3: **repeat**

4: **if** $f_\delta(x^{k+1}) \leq f_\delta(x^k) + \psi(x^{k+1}, x^k) + L_{k+1}V(x^{k+1}, x^k) + \delta_{k+1} + \Delta_{k+1} \|x^{k+1} - x^k\|$ **then**

5: $k := k + 1$ и выполнение п. 1.

6: **else**

7: $L_{k+1} := 2L_{k+1}$, $\delta_{k+1} := 2 \cdot \delta_{k+1}$; $\Delta_{k+1} := 2 \cdot \Delta_{k+1}$ и выполнение п. 2.

8: **end if**

9: **until** $k \geq N$

Ensure: $y^{k+1} := \arg \min_{i=0, \dots, k} f(x^{i+1})$.

Всюду далее условимся полагать, что $\prod_{i=k+1}^k a_i = 1$ для некоторой числовой последовательности a_i . Справедлива следующая

Теорема 1. Пусть f имеет (δ, Δ, L, μ) -модель в каждой точке $x \in Q$. Тогда после k итераций справедливо неравенство

$$f(y^{k+1}) - f(x_*) \leq \frac{1}{\sum_{i=0}^k \frac{1}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right)} \times \\ \times \left(\prod_{i=0}^k \left(1 - \frac{\mu}{L_{i+1}}\right) V(x_*, x^0) + \sum_{i=0}^k \frac{\delta + \delta_{i+1} + \Delta_{i+1} \|x^{i+1} - x^i\|}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right) \right).$$

Отметим, что вспомогательная задача п. 2 листинга алгоритма 1 решается не более

$$2k + \max \left\{ \log_2 \frac{2L}{L_0}, \log_2 \frac{2\delta}{\delta_0}, \log_2 \frac{2\Delta}{\Delta_0} \right\} \quad (9)$$

раз.

Доказательство. Введем обозначения: $\widehat{\delta}_{k+1} := \delta_{k+1} + \Delta_{k+1} \|x^{k+1} - x^k\|$. После k итераций алгоритма 1 получаем

$$0 \leq \psi(x, x^k) - \psi(x^{k+1}, x^k) + L_{k+1}V(x, x^k) - L_{k+1}V(x, x^{k+1}) - L_{k+1}V(x^{k+1}, x^k), \text{ откуда}$$

$$L_{k+1}V(x, x^{k+1}) \leq \psi(x, x^k) - \psi(x^{k+1}, x^k) + L_{k+1}V(x, x^k) - L_{k+1}V(x^{k+1}, x^k). \quad (10)$$

Согласно неравенству (7):

$$-L_{k+1}V(x^{k+1}, x^k) \leq \widehat{\delta}_{k+1} - f_\delta(x^{k+1}) + f_\delta(x^k) + \psi(x^{k+1}, x^k).$$

Применяя теперь (10), получаем

$$L_{k+1}V(x, x^{k+1}) \leq \widehat{\delta}_{k+1} - f(x^{k+1}) + f_\delta(x^k) + \psi(x, x^k) + L_{k+1}V(x, x^k) + \delta. \quad (11)$$

Пусть $x = x_*$. Тогда, учитывая (8), имеем $f_\delta(x^k) + \psi(x_*, x^k) \leq f(x_*) - \mu V(x_*, x^k)$. Применим это неравенство к (11): $L_{k+1}V(x_*, x^{k+1}) \leq \delta + \widehat{\delta}_{k+1} + f(x_*) - f(x^{k+1}) + (L_{k+1} - \mu)V(x_*, x^k)$. Далее,

$$V(x_*, x^{k+1}) \leq \frac{f(x_*) - f(x^{k+1})}{L_{k+1}} + \frac{\delta + \widehat{\delta}_{k+1}}{L_{k+1}} + \left(1 - \frac{\mu}{L_{k+1}}\right) V(x_*, x^k) \leq$$

$$\begin{aligned} &\leq \frac{f(x_*) - f(x^{k+1})}{L_{k+1}} + \frac{1}{L_k} \left(1 - \frac{\mu}{L_{k+1}}\right) \left(f(x_*) - f(x^k)\right) + \frac{\delta + \widehat{\delta}_{k+1}}{L_{k+1}} + \\ &\quad + \frac{\delta + \widehat{\delta}_k}{L_k} \left(1 - \frac{\mu}{L_{k+1}}\right) + \left(1 - \frac{\mu}{L_{k+1}}\right) \left(1 - \frac{\mu}{L_k}\right) V(x_*, x^{k-1}) \leq \\ &\leq \sum_{i=0}^k \frac{f(x_*) - f(x^{i+1})}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right) + \sum_{i=0}^k \frac{\delta + \widehat{\delta}_{i+1}}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right) + \prod_{i=0}^k \left(1 - \frac{\mu}{L_{i+1}}\right) V(x_*, x_0). \end{aligned}$$

С учетом $y^{k+1} = \arg \min_{i=0, \dots, k} f(x^{i+1})$ и $V(x_*, x^{k+1}) \geq 0$ имеем $f(y^{k+1}) - f(x_*) \leq$

$$\leq \frac{1}{\sum_{i=0}^k \frac{1}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right)} \left(\prod_{i=0}^k \left(1 - \frac{\mu}{L_{i+1}}\right) V(x_*, x^0) + \sum_{i=0}^k \frac{\delta + \widehat{\delta}_{i+1}}{L_{i+1}} \prod_{j=i+1}^k \left(1 - \frac{\mu}{L_j}\right) \right).$$

Оценка (9) обосновывается аналогично п. 2 доказательства теоремы 2.1 из [4].

Замечание 2. Оценка (9) показывает, что в среднем трудоемкость итерации предложенного адаптивного алгоритма превышает трудоемкость неадаптивного метода не более чем в постоянное число раз. Отметим также, что при $k = 0, 1, 2, \dots$ $L_{k+1} \leq 2CL$, $C = \max \left\{ 1, \frac{2\delta}{\delta_0}, \frac{2\Delta}{\Delta_0} \right\}$.

Следствие 1. При $\mu = 0$ полученная оценка качества решения принимает вид

$$\begin{aligned} f(y^{k+1}) - f(x_*) &\leq \frac{V(x_*, x^0)}{\sum_{i=0}^k \frac{1}{L_{i+1}}} + \left(\sum_{i=0}^k \frac{1}{L_{i+1}} \right)^{-1} \sum_{i=0}^k \frac{\delta_{i+1} + \Delta_{i+1} \|x^{i+1} - x^i\|}{L_{i+1}} + \delta \leq \quad (12) \\ &\leq 2CLV(x_*, x^0) + \left(\sum_{i=0}^k \frac{1}{L_{i+1}} \right)^{-1} \sum_{i=0}^k \frac{\delta_{i+1} + \Delta_{i+1} \|x^{i+1} - x^i\|}{L_{i+1}} + \delta. \end{aligned}$$

2. Оценка скорости сходимости для варианта быстрого градиентного метода, использующего концепцию (δ, Δ, L) -модели целевой функции в запрошенной точке

Оказывается, что концепция (δ, Δ, L) -модели оптимизируемого функционала позволяет получить аналог теоремы 1 для адаптивного варианта быстрого градиентного метода Ю. Е. Нестерова (БГМ) [15]. Смысл использования этого метода в том, что при отсутствии погрешностей он даёт оптимальные оценки скорости сходимости на классе задач выпуклой оптимизации с липшицевым градиентом. Мы отправляемся от [2], где предложен адаптивный быстрый градиентный метод с оракулом, использующий (δ, L) -модель целевой функции в запрошенной точке (в нашей терминологии это случай $\Delta = 0$). Здесь уже существенна 1-сильная выпуклость прокс-функции $d(x)$ относительно нормы.

Замечание 3. Аналогично замечанию 2 для всякого $k \geq 0$ при некотором $C \geq 1$ выполнено $L_k \leq 2CL$. При $k = 0$ это верно из того, что $L_0 \leq 2L$. Для $k \geq 1$ это следует из того, что мы выйдем из внутреннего цикла, где подбирается L_k , ранее, чем L_k станет больше $2CL$. Выход из цикла гарантируется тем, что согласно предположению существует (δ, Δ, L) -модель для $f(x)$ в любой точке $x \in Q$.

По схеме рассуждений [2] можно проверить следующий результат.

Алгоритм 2 Быстрый градиентный метод с оракулом, использующий (δ, Δ, L) -модель в запрошенной точке.

Require: $x^0 \in Q$ — начальная точка, $V(x_*, x^0) \leq R^2$, параметры $L_0 > 0$, $\Delta_0 > 0$, $\delta_0 > 0$ ($L_0 \leq 2L$, $\Delta_0 \leq 2\Delta$, $\delta_0 \leq 2\delta$).

1: **0-шаг:** $y^0 := x^0$, $u^0 := x^0$, $L_1 := \frac{L_0}{2}$, $\Delta_1 := \frac{\Delta_0}{2}$, $\delta_1 := \frac{\delta_0}{2}$, $\alpha_0 := 0$, $A_0 := \alpha_0$

2: **for** $k = 1, \dots$ **do**

3: Находим наибольший корень $\alpha_{k+1} : A_k + \alpha_{k+1} = L_{k+1}\alpha_{k+1}^2$;

$$A_{k+1} := A_k + \alpha_{k+1}, y_{k+1} := \frac{\alpha_{k+1}u^k + A_kx^k}{A_{k+1}}, \varphi_{k+1}(x) = V(x, u^k) + \alpha_{k+1}\psi(x, y^{k+1});$$

$$u^{k+1} := \arg \min_{x \in Q} \varphi_{k+1}(x), x^{k+1} := \frac{\alpha_{k+1}u^{k+1} + A_kx^k}{A_{k+1}};$$

4: **if** $f_\delta(x^{k+1}) \leq f_\delta(y^{k+1}) + \psi(x^{k+1}, y^{k+1}) + \frac{L_{k+1}}{2}\|x^{k+1} - y^{k+1}\|^2 + \Delta_{k+1}\|x^{k+1} - y^{k+1}\| + \delta_{k+1}$
then

5: $L_{k+2} := \frac{L_{k+1}}{2}$, $\Delta_{k+2} := \frac{\Delta_{k+1}}{2}$ и $\delta_{k+2} := \frac{\delta_{k+1}}{2}$ и перейти к следующему шагу

6: **else**

7: $L_{k+1} := 2L_{k+1}$, $\Delta_{k+1} := 2\Delta_{k+1}$ и $\delta_{k+1} := 2\delta_{k+1}$ и повторить текущий шаг

8: **end if**

9: **end for**

Теорема 2. Пусть $V(x_*, x^0) \leq R^2$, где x^0 — начальная точка, а x_* — ближайшая точка минимума к точке x^0 в смысле дивергенции Брэгмана. Для предложенного алгоритма 2 после k итераций выполнено следующее неравенство:

$$f(x^k) - f(x_*) \leq \frac{R^2}{A_k} + \frac{1}{A_k} \sum_{i=0}^{k-1} (\Delta_{i+1} \|x^{i+1} - y^{i+1}\| + \delta_{i+1} + \delta) A_{i+1}.$$

Отметим, что вспомогательная задача пункта 3 листинга алгоритма 2 решается не более

$$2k + \max \left\{ \log_2 \frac{2L}{L_0}, \log_2 \frac{2\delta}{\delta_0}, \log_2 \frac{2\Delta}{\Delta_0} \right\} \quad (13)$$

раз.

Полученную оценку можно несколько конкретизировать с использованием следующего вспомогательного утверждения, которое также проверяется по схеме рассуждений [2] (по сути, оно уже доказано в [2] при $C = 1$).

Лемма 1. Пусть для последовательности α_k выполнено

$$\alpha_0 = 0, A_k = \sum_{i=0}^k \alpha_i, A_k = L_k \alpha_k^2,$$

где для фиксированного $C \geq 1$ верно $L_k \leq 2CL$ при всяком $k \geq 0$ согласно замечанию 3 выше. Тогда для любого $k \geq 1$ верно следующее неравенство:

$$A_k \geq \frac{(k+1)^2}{8CL}. \quad (14)$$

Поэтому из теоремы 2 вытекает

Следствие 2. Пусть $V(x_*, x^0) \leq R^2$, где x^0 — начальная точка, а x_* — ближайшая точка минимума к точке x^0 в смысле дивергенции Брэгмана. Для предложенного алгоритма 2 выполнено следующее неравенство:

$$f(x^k) - f(x_*) \leq \frac{8CLR^2}{(k+1)^2} + \frac{1}{A_k} \sum_{i=0}^{k-1} (\Delta_{i+1} \|x^{i+1} - y^{i+1}\| + \delta_{i+1} + \delta) A_{i+1}.$$

3. О применимости предложенной концепции неточной модели функции к негладким задачам за счет введения искусственных неточностей

Отметим, что на величину Δ в (12) можно смотреть и как на характеристику негладкости функционала f . Точнее говоря, Δ можно понимать, например, как верхнюю оценку суммы диаметров субдифференциалов f в точках негладкости вдоль всевозможных векторных отрезков $[x; y]$ из области определения f . Оказывается, в случае известной величины $\Delta < +\infty$ возможно несколько модифицировать алгоритм 1, обеспечив уменьшение $\Delta_{i+1} \|x^{i+1} - x^i\|$ в (12) до любой заданной величины. Это позволит показать оптимальность данного метода в теории нижних оракульных оценок [1] с точностью до логарифмического множителя. Заметим, что в данном пункте мы всюду требуем 1-сильную выпуклость прокс-функции в определении 1. Также обозначим всюду далее через $f^* = f(x_*)$ — искомое минимальное значение f .

Покажем, как это возможно сделать в предположении, что целевой функционал f допускает (δ, Δ, L) -модель ψ .

Пусть на $(k+1)$ -й итерации алгоритма 1 ($k = 0, 1, \dots, N - 1$) верно неравенство $L \leq L_{k+1} \leq 2L$ (как показано в п. 2 доказательства теоремы 2.1 из [4], этого можно всегда добиться выполнением не более чем постоянного числа операций п. 2 листинга алгоритма 1). Для каждой итерации алгоритма 1 ($k = 0, 1, \dots, N - 1$) предложим такую процедуру:

$$\boxed{\text{Повторяем операции п. 2 } p \text{ раз, увеличивая } L_{k+1} \text{ в два раза при неизменной } \Delta_{k+1} \leq 2\Delta.} \quad (15)$$

Процедуру (15) остановим в случае выполнения одного из неравенств:

$$\Delta_{k+1} \|x^{k+1} - x^k\| \leq \frac{\varepsilon}{2} \quad (16)$$

или

$$f(x^{k+1}) \leq f(x^k) + \psi(x^{k+1}, x^k) + 2^{p-1}L \|x^{k+1} - x^k\|^2. \quad (17)$$

Отметим, что здесь мы полагаем функционал f точно заданным, то есть $f_\delta = f$ ($\delta = 0$) и $\psi(y, x) = \langle \nabla f(x), y - x \rangle$, где ∇f — некоторый субградиент f . В работе [4] доказано, что p можно выбрать так, чтобы $2^p \geq 1 + \frac{16\Delta^2}{\varepsilon L}$, и получен следующий результат.

Теорема 3. Пусть $\mu = 0$. Тогда для выхода \hat{y}^{k+1} модифицированного алгоритма 1 с учетом дополнительной процедуры (15) неравенство $f(\hat{y}^{k+1}) - f^* \leq \varepsilon$ будет гарантированно выполнено не более чем после

$$\left\lceil \frac{4LR^2}{\varepsilon} + \frac{64\Delta^2 R^2}{\varepsilon^2} \right\rceil \cdot \left\lceil \log_2 \left(1 + \frac{16\Delta^2}{\varepsilon L} \right) \right\rceil \quad (18)$$

вычислений субградиента f .

Пусть теперь $\mu > 0$. Если положить $\delta_{i+1} = \delta = 0$ для всякого $i = \overline{0, k}$ и $\Delta_{i+1} \|x^{i+1} - x^i\| \leq \frac{\varepsilon}{2}$, то согласно теореме 1

$$f(y^{k+1}) - f(x_*) \leq 2CL \left(1 - \frac{\mu}{2CL} \right)^{k+1} V(x_*, x^0) + \frac{\varepsilon}{2}.$$

В частности, данное неравенство позволяет при понимании величины Δ в (12) как характеристики негладкости функционала f применить процедуру, аналогичную (15) (при этом уже необходимо требовать 1-сильную выпуклость прокс-функции в определении 2). Тогда ε -точность решения задачи минимизации f будет достигаться за

$$O\left(\frac{1}{\varepsilon} \log_2^2 \frac{1}{\varepsilon}\right)$$

шагов градиентного метода (пункта 2 листинга алгоритма 1). Действительно, если положить $\hat{\delta} = \delta = 0$ (то есть $f_{\delta} = f$) и $V(x_*, x^0) \leq R^2$, то для достижения качества решения

$$f(y^{k+1}) - f(x_*) \leq \varepsilon$$

необходимо выполнить не более

$$\left\lceil \frac{2CL}{\mu} \ln \frac{4CLR^2}{\varepsilon} \right\rceil - 1$$

итераций алгоритма 1. Ввиду оценки (22) с учетом процедуры (15) на каждом шаге итоговое число обращений к (суб)градиенту целевого функционала для алгоритма 1 можно оценить как

$$\left\lceil \left(\frac{2CL}{\mu} + \frac{32C\Delta^2}{\mu\varepsilon} \right) \ln \left(\frac{4CLR^2}{\varepsilon} + \frac{64CR^2}{\varepsilon^2} \right) \right\rceil \left\lceil \log_2 \left(1 + \frac{16\Delta^2}{\varepsilon L} \right) \right\rceil - 1.$$

Покажем, что даст применение похожей схемы для быстрого градиентного метода на классе выпуклых функционалов f , для которых при некоторых $L > 0$ и $\Delta > 0$ верно

$$f(y) \leq f(x) + \psi(y, x) + \frac{L}{2} \|y - x\|^2 + \Delta \|y - x\| \quad \forall x, y \in Q. \quad (19)$$

В частности, для задач выпуклой негладкой оптимизации $\psi(y, x) = \langle \nabla f(x), y - x \rangle$ для некоторого субградиента ∇f .

Рассмотрим вариант быстрого градиентного метода (алгоритм 2), который использует концепцию (δ, Δ, L) -модели целевого функционала в запрошенной точке. Применим неадаптивный вариант этого метода с постоянным шагом $L_{k+1} = 2^p L$ для некоторого $p \in \mathbb{N}$ и $\delta = \frac{\varepsilon}{2\gamma}$ для некоторых $\varepsilon > 0$ и фиксированной постоянной $\gamma > 0$. Будем при этом считать $f_{\delta} = f$ (то есть функционал f задан точно). В таком случае на $(k+1)$ -й итерации ($k = 0, 1, 2, \dots$) верно неравенство

$$f(x^{k+1}) \leq f(x^k) + \psi(x^{k+1}, x^k) + \frac{2^p L}{2} \|x^{k+1} - x^k\|^2 + \frac{\varepsilon}{2\gamma}$$

и после N итераций (это можно проверить аналогично [2])

$$f(x^N) - f^* \leq \frac{8 \cdot 2^p \cdot LR^2}{(N+1)^2} + \frac{\varepsilon N}{2\gamma}. \quad (20)$$

Выберем натуральное p так, чтобы гарантированно выполнялась альтернатива

$$\left[\begin{array}{l} \Delta \|x^{k+1} - y^{k+1}\| \leq \frac{\varepsilon}{2\gamma}, \\ \frac{(2^p - 1)L}{2} \|x^{k+1} - y^{k+1}\|^2 \geq \Delta \|x^{k+1} - y^{k+1}\|. \end{array} \right. \quad (21)$$

Если

$$\Delta \|x^{k+1} - y^{k+1}\| > \frac{\varepsilon}{2\gamma},$$

то

$$\frac{(2^p - 1)L}{2} \|x^{k+1} - y^{k+1}\|^2 > \frac{(2^p - 1)L\varepsilon}{4\gamma\Delta}.$$

Тогда второе неравенство альтернативы (21) заведомо выполнится при

$$2^p > 1 + \frac{4\gamma\Delta^2}{L\varepsilon}. \quad (22)$$

Поэтому положим

$$p = \left\lceil \log_2 \left(1 + \frac{4\gamma\Delta^2}{L\varepsilon} \right) \right\rceil. \quad (23)$$

Теперь покажем, каким можно выбрать количество итераций N , чтобы гарантированно (46) обеспечивало $f(x^N) - f^* \leq \varepsilon$. Для этого потребуем выполнения неравенств

$$\frac{2^{p+3}LR^2}{(N+1)^2} \leq \frac{\varepsilon}{2} \text{ и } \frac{\varepsilon N}{2\gamma} \leq \frac{\varepsilon}{2},$$

откуда $\gamma \geq N$ и $(N+1)^2 \geq \frac{2^{p+4}LR^2}{\varepsilon}$. Для упрощения выкладок усилим последнее требование: $N^2 \geq \frac{2^{p+4}LR^2}{\varepsilon}$, откуда

$$N^2 > \frac{16LR^2}{\varepsilon} \left(1 + \frac{4\gamma\Delta^2}{L\varepsilon}\right) \geq \frac{16LR^2}{\varepsilon} + \frac{64N\Delta^2R^2}{\varepsilon^2}.$$

Это означает, что N можно выбирать как $\lceil N_2 \rceil$, где N_2 — больший корень уравнения

$$N^2 - \frac{64\Delta^2R^2}{\varepsilon}N - \frac{16LR^2}{\varepsilon} = 0,$$

$$N_2 = \frac{32\Delta^2R^2}{\varepsilon^2} + \sqrt{\left(\frac{32\Delta^2R^2}{\varepsilon^2}\right)^2 + \frac{16LR^2}{\varepsilon}}.$$

Далее, в силу неравенства $\sqrt{a+b} > \sqrt{\frac{a}{2}} + \sqrt{\frac{b}{2}}$ ($a, b > 0$) имеем

$$N \geq N_2 \geq \frac{(32 + 16\sqrt{2})\Delta^2R^2}{\varepsilon^2} + \frac{2R\sqrt{2L}}{\sqrt{\varepsilon}}.$$

Таким образом, можно гарантировать достижение ε -точного решения задачи минимизации f за

$$O\left(\sqrt{\frac{L}{\varepsilon}}\right) + O\left(\frac{\Delta^2}{\varepsilon^2}\right)$$

итераций быстрого градиентного метода, то есть обоснован некоторый аналог результата о слайдинге [14], но уже целевая функция не обязательно есть сумма гладкого и негладкого слагаемых.

При этом возможно использовать метод с адаптивной настройкой констант L , а также $\Delta_{k+1} \leq \Delta$. Тогда оценка числа итераций может измениться в постоянное число раз. Также при этом за счёт p дополнительных процедур адаптивного подбора L_{k+1} (при условии $L \leq L_{k+1} \leq 2L$ для всякого $k = 0, 1, \dots, N-1$) на каждой итерации метода добавится множитель (23):

$$p \geq \left\lceil \log_2 \left(1 + \frac{4N\Delta^2}{L\varepsilon}\right) \right\rceil \geq \log_2 \left(1 + \frac{(128 + 64\sqrt{2})\Delta^4R^2}{L\varepsilon^3} + \frac{8R\Delta^2\sqrt{2}}{\sqrt{L\varepsilon^3}}\right).$$

Теорема 4. Для выхода x^N модифицированного алгоритма 2 с учетом дополнительной процедуры (15) при $\Delta_{k+1} = \Delta$ на $(k+1)$ -й итерации ($k = 0, 1, \dots, N-1$) неравенство $f(x^N) - f^* \leq \varepsilon$ будет гарантированно выполнено не более чем после

$$\left\lceil \frac{(32 + 16\sqrt{2})\Delta^2R^2}{\varepsilon^2} + \frac{2R\sqrt{2L}}{\sqrt{\varepsilon}} \right\rceil \cdot \left\lceil \log_2 \left(1 + \frac{(128 + 64\sqrt{2})\Delta^4R^2}{L\varepsilon^3} + \frac{8R\Delta^2\sqrt{2}}{\sqrt{L\varepsilon^3}}\right) \right\rceil. \quad (24)$$

вычислений субградиента f .

Замечание 4. Если не предполагать, что на $(k+1)$ -й итерации ($k = 0, 1, \dots, N-1$) модифицированных алгоритмов 1 и 2 выполнено неравенство $L \leq L_{k+1} \leq 2L$ и предусмотреть

полностью адаптивную настройку параметров L и Δ в методе, то оценки (18) и (24) могут увеличиться не более чем в

$$\left[\max \left\{ \frac{2L}{L_0}, \frac{2\Delta}{\Delta_0} \right\} \right]$$

раз. Отметим также, что логарифмические множители в (18) и (24) можно опустить, если рассматривать неадаптивные варианты методов 1 и 2 с фиксированным параметром $L_{k+1} = 2^p L$ при подходящем натуральном p .

4. Адаптивная настройка на величину погрешностей для вариационных неравенств и седловых задач

Метод с адаптивной настройкой погрешностей можно предложить и для вариационных неравенств [16], а также седловых задач. В данном пункте мы покажем, как это можно сделать в модельной общности. Мы отправляемся от [17], где введена концепция неточной модели для вариационных неравенств и седловых задач, но с постоянной неточностью и без адаптивной настройки к величинам погрешностей.

Напомним постановку задачи решения вариационного неравенства, а также необходимые понятия и результаты. Для оператора $G : Q \rightarrow \mathbb{R}^n$, заданного на выпуклом компакте $Q \subset \mathbb{R}^n$ под *сильным вариационным неравенством* понимаем неравенство вида

$$\langle G(x_*), x_* - x \rangle \leq 0. \quad (25)$$

Отметим, что в (25) требуется найти $x_* \in Q$ (это x_* и называется решением ВН) для которого

$$\max_{x \in Q} \langle G(x_*), x_* - x \rangle \leq 0. \quad (26)$$

Для монотонного оператора поля G можно рассматривать *слабые вариационные неравенства*

$$\langle G(x), x_* - x \rangle \leq 0. \quad (27)$$

Обычно в (27) требуется найти $x_* \in Q$, для которого (27) верно при всех $x \in Q$.

Предложен аналог концепции неточной модели целевой функции в оптимизации для вариационных неравенств и седловых задач. Для удобства будем рассматривать задачу нахождения решения $x_* \in Q$ абстрактной задачи равновесного программирования

$$\psi(x, x_*) \geq 0 \quad \forall x \in Q \quad (28)$$

для некоторого выпуклого компакта $Q \subset \mathbb{R}^n$, а также функционала $\psi : Q \times Q \rightarrow \mathbb{R}$. Если предположить абстрактную монотонность функционала ψ :

$$\psi(x, y) + \psi(y, x) \leq 0 \quad \forall x, y \in Q, \quad (29)$$

то всякое решение (28) будет также и решением двойственной задачи равновесия

$$\psi(x_*, x) \leq 0 \quad \forall x \in Q. \quad (30)$$

В общем случае сделаем предположение о существовании решения x_* задачи (28). Приведем пару примеров задания ψ , для которых данное условие заведомо выполнено.

Пример 2. Если для некоторого оператора $G : R \rightarrow \mathbb{R}^n$ положить

$$\psi(x, y) = \langle G(y), x - y \rangle \quad \forall x, y \in Q, \quad (31)$$

то (28) и (30) будут равносильны соответственно стандартным сильному и слабому вариационному неравенству с оператором G .

Пример 3. Для некоторого оператора $G : Q \rightarrow \mathbb{R}^n$ и выпуклого функционала $h : Q \rightarrow \mathbb{R}^n$ простой структуры выбор функционала

$$\psi(x, y) = \langle G(y), x - y \rangle + h(x) - h(y) \quad (32)$$

приводит к смешанному вариационному неравенству

$$\langle G(y), y - x \rangle + h(y) - h(x) \leq 0, \quad (33)$$

которое в случае монотонности оператора G влечет

$$\langle G(x), y - x \rangle + h(y) - h(x) \leq 0. \quad (34)$$

Концепцию (δ, Δ, L) -модели для выделенного выше класса задач возможно ввести следующим образом.

Определение 3. Будем говорить, что функционал ψ допускает (δ, Δ, L) -модель $\psi_\delta(x, y)$ при некоторых фиксированных значениях параметров $L, \delta, \Delta > 0$ в произвольной точке y относительно дивергенции Брэгмана $V(y, x)$, если для произвольных $x, y, z \in Q$ верны:

(i) $\psi(x, y) \leq \psi_\delta(x, y) + \delta;$

(ii) $\psi_\delta(x, y)$ выпуклый функционал по первой переменной и $\psi_\delta(x, x) = 0;$

(iii) (абстрактная δ -монотонность)

$$\psi_\delta(x, y) + \psi_\delta(y, x) \leq \delta; \quad (35)$$

(iv) (обобщённая относительная гладкость)

$$\psi_\delta(x, y) \leq \psi_\delta(x, z) + \psi_\delta(z, y) + LV(x, z) + LV(z, y) + \delta + \Delta \|y - z\|. \quad (36)$$

Естественно возникает идея обобщить этот метод на абстрактные задачи (28) и (30) в предположениях их разрешимости, а также (i) – (iv). При этом будем учитывать погрешность δ в (36), а также погрешность $\tilde{\delta}$ решения вспомогательных задач на итерациях согласно одному из достаточно известных в алгоритмической оптимизации подходов:

$$x := \arg \min_{y \in Q} \tilde{\delta} \varphi(y), \text{ если } \langle \nabla \varphi(x), x - y \rangle \leq \tilde{\delta}. \quad (37)$$

По сути, предполагается, что вспомогательные задачи минимизации решаются с некоторой точностью $\tilde{\delta}$ по функции.

К задачам вида (28) и (30) при условии наличия (δ, Δ, L) -модели возможно применить аналог проксимального зеркального метода А.С. Немировского с адаптивным выбором шага. Опишем $(N + 1)$ -ю итерацию этого метода ($N = 0, 1, 2, \dots$), выбрав начальное приближение $x^0 = \arg \min_{x \in Q} d(x)$, зафиксировав точность $\varepsilon > 0$, а также некоторые константы $L_0 \leq 2L$, $\delta_0 \leq 2\delta$ и $\Delta_0 \leq 2\Delta$.

Алгоритм 3 Адаптивный метод для концепции (δ, Δ, L) -модели для ВН.

$$1. N := N + 1, L_{N+1} := \frac{L_N}{2}, \delta_{N+1} := \frac{\delta_N}{2}, \Delta_{N+1} := \frac{\Delta_N}{2}.$$

2. Вычисляем:

$$y^{N+1} := \arg \min_{x \in Q}^{\tilde{\delta}} \{ \psi_\delta(x, x^N) + L_{N+1} V(x, x^N) \},$$

$$x^{N+1} := \arg \min_{x \in Q}^{\tilde{\delta}} \{ \psi_\delta(x, y^{N+1}) + L_{N+1} V(x, x^N) \}$$

до тех пор, пока не будет выполнено:

$$\begin{aligned} \psi_\delta(x^{N+1}, x^N) \leq & \psi_\delta(y^{N+1}, x^N) + \psi_\delta(x^{N+1}, y^{N+1}) + \\ & + L_{N+1} V(y^{N+1}, x^N) + L_{N+1} V(y^{N+1}, x^{N+1}) + \Delta_{N+1} \|y^{N+1} - x^{N+1}\| + \delta_{N+1}. \end{aligned} \quad (38)$$

3. Если (38) не выполнено, то $L_{N+1} := 2L_{N+1}$, $\delta_{N+1} := 2\delta_{N+1}$, $\Delta_{N+1} := 2\Delta_{N+1}$ и повторяем п. 2.

4. Иначе переход к п. 1.

5. Критерий остановки метода:

$$\sum_{k=0}^{N-1} \frac{1}{L_{k+1}} \geq \frac{\max_{x \in Q} V(x, x^0)}{\varepsilon}. \quad (39)$$

Для краткости будем всюду далее обозначать

$$S_N := \sum_{k=0}^{N-1} \frac{1}{L_{k+1}}. \quad (40)$$

Справедлива следующая

Теорема 5. После остановки рассматриваемого метода для всякого $x \in Q$ будет заведомо выполнено неравенство

$$-\frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\psi_\delta(x, y^{k+1})}{L_{k+1}} \leq \varepsilon + 2\tilde{\delta} + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1} + \Delta_{k+1} \|y^{k+1} - x^{k+1}\|}{L_{k+1}}, \quad (41)$$

а также

$$\psi(\tilde{y}, x) \leq \varepsilon + 2\tilde{\delta} + 2\delta + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1} + \Delta_{k+1} \|y^{k+1} - x^{k+1}\|}{L_{k+1}} \quad (42)$$

при

$$\tilde{y} := \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{y^{k+1}}{L_{k+1}}. \quad (43)$$

Замечание 5. Для обычных слабых вариационных неравенств (27) неравенство (42) можно заменить на

$$\max_{x \in Q} \langle G(x), \tilde{y} - x \rangle \leq \varepsilon + 2\tilde{\delta} + 2\delta + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1} + \Delta_{k+1} \|y^{k+1} - x^{k+1}\|}{L_{k+1}}. \quad (44)$$

Аналогично теореме 3 можно сформулировать результат о том, что при $\gamma = \delta = \tilde{\delta} = 0$ процедурой типа (15) можно добиться выполнения $\max_{x \in Q} \langle G(x), \tilde{y} - x \rangle \leq \varepsilon$ за

$$\left\lceil \frac{4LR^2}{\varepsilon} + \frac{64\Delta^2}{\varepsilon^2} \right\rceil \left\lceil \log \left(1 + \frac{16\Delta^2}{\varepsilon L} \right) \right\rceil$$

обращения к оракулу для G .

Можно ввести также аналог концепции (δ, Δ, L) -модели для седловых задач. Напомним, что постановка седловой задачи предполагает, что для выпуклого по u и вогнутого по v функционала $f(u, v) : \mathbb{R}^{n_1+n_2} \rightarrow \mathbb{R}$ ($u \in Q_1 \subset \mathbb{R}^{n_1}$ и $v \in Q_2 \subset \mathbb{R}^{n_2}$) требуется найти (u_*, v_*) такую, что

$$f(u_*, v) \leq f(u_*, v_*) \leq f(u, v_*) \quad (45)$$

для произвольных $u \in Q_1$ и $v \in Q_2$. Мы считаем Q_1 и Q_2 выпуклыми компактами в пространствах \mathbb{R}^{n_1} и \mathbb{R}^{n_2} и поэтому $Q = Q_1 \times Q_2 \subset \mathbb{R}^{n_1+n_2}$ также есть выпуклый компакт. Для всякого $x = (u, v) \in Q$ будем полагать, что $\|x\| = \sqrt{\|u\|_1^2 + \|v\|_2^2}$, где $\|\cdot\|_1$ и $\|\cdot\|_2$ — нормы в пространствах \mathbb{R}^{n_1} и \mathbb{R}^{n_2} . Условимся обозначать $x = (u_x, v_x)$, $y = (u_y, v_y) \in Q$.

Хорошо известно, что для достаточно гладкой функции f по u и v задача (45) сводится к вариационному неравенству с оператором

$$G(x) = \begin{pmatrix} f'_u(u_x, v_x) \\ -f'_v(u_x, v_x) \end{pmatrix}. \quad (46)$$

Предложим некоторую вариацию концепции (δ, Δ, L) -модели для вариационных неравенств, но уже на более узком классе седловых задач.

Определение 4. Будем говорить, что для некоторой постоянной $\delta > 0$ функция $\psi_\delta(x, y)$ ($\psi : \mathbb{R}^{n_1+n_2} \times \mathbb{R}^{n_1 \times n_2} \rightarrow \mathbb{R}$) есть (δ, Δ, L) -модель для седловой задачи (45), если для некоторого функционала ψ_δ при произвольных $x, y, z \in Q$ выполнены предположения (i) – (iv) определения 3, а также справедливо неравенство:

$$f(u_y, v_x) - f(u_x, v_y) \leq -\psi_\delta(x, y) + \delta \quad \forall x, y \in Q. \quad (47)$$

Из теоремы 5 вытекает

Теорема 6. Если для седловой задачи (45) существует (δ, Δ, L) -модель $\psi_\delta(x, y)$, то после остановки алгоритма 3 получим точку

$$\tilde{y} = (u_{\tilde{y}}, v_{\tilde{y}}) := (\tilde{u}, \tilde{v}) := \frac{1}{S_N} \sum_{k=0}^{N_1} \frac{y^{k+1}}{L_{k+1}}, \quad (48)$$

для которой верна оценка величины-качества решения седловой задачи:

$$\max_{v \in Q_2} f(\tilde{u}, v) - \min_{u \in Q_1} f(u, \tilde{v}) \leq \varepsilon + 2\tilde{\delta} + \delta + \frac{1}{S_N} \sum_{k=0}^{N-1} \frac{\delta_{k+1} + \Delta_{k+1} \|y^{k+1} - x^{k+1}\|}{L_{k+1}}. \quad (49)$$

5. Адаптивные алгоритмы зеркального спуска для негладких задач выпуклой оптимизации с функциональными ограничениями

В данном разделе мы рассмотрим некоторые методы для задачи минимизации выпуклой негладкой функции $f(x) \rightarrow \min$ с липшицевым выпуклым функциональным ограничением $g(x) \leq 0$. Далее будем полагать, что x_* — одно из решений такой задачи. Мы будем рассматривать для выделенного класса задач методы, аналогичные схемам с переключениями, которые восходят к пионерским работам [18, 19]. Недавно в [20] были предложены методы зеркального спуска такого типа для условных задач с адаптивным выбором шага и адаптивными критериями остановки. Более того, оказалось [13, 21], что для указанных процедур

оптимальную на классе выпуклых липшицевых целевых функционалов оценку скорости сходимости $O(\varepsilon^{-2})$ можно распространить и на класс квазивыпуклых гёльдеровых целевых функционалов. Рассмотрены и приложения к задаче оптимизации компьютерной сети для логарифмического целевого функционала [21]. По итогам проведённых экспериментов выяснилось, что, как правило, быстрее всех из рассмотренных схем [13, 20, 21] работает следующий метод.

Алгоритм 4 Адаптивный зеркальный спуск

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

- 1: $x^0 = \operatorname{argmin}_{x \in Q} d(x)$
- 2: $I =: \emptyset$
- 3: $k \leftarrow 0$
- 4: **repeat**
- 5: **if** $g(x^k) \leq \varepsilon \|\nabla g(x^k)\|_*$ **then**
- 6: $h_k \leftarrow \frac{\varepsilon}{\|\nabla f(x^k)\|_*^2}$
- 7: $x^{k+1} \leftarrow \operatorname{Mirr}_{x^k}(h_k \nabla f(x^k))$ // "продуктивные шаги"
- 8: $k \rightarrow I$
- 9: **else**
- 10: $h_k \leftarrow \frac{\varepsilon}{\|\nabla g(x^k)\|_*}$
- 11: $x^{k+1} \leftarrow \operatorname{Mirr}_{x^k}(h_k \nabla g(x^k))$ // "непродуктивные шаги"
- 12: **end if**
- 13: $k \leftarrow k + 1$
- 14: **until**

$$\frac{2\Theta_0^2}{\varepsilon^2} \leq \sum_{k \in I} \frac{1}{\|\nabla f(x^k)\|_*^2} + |J|, \quad (50)$$

где $|J|$ — количество непродуктивных шагов (мы обозначим через $|I|$ количество продуктивных шагов, то есть $|I| + |J| = N$).

Ensure: $\hat{x} = \frac{1}{\sum_{k \in I} h_k} \sum_{k \in I} h_k x^k$.

Обозначим через $\varepsilon > 0$ фиксированную точность, x_0 — начальное приближение такое, что для некоторого $\Theta_0 > 0$ верно неравенство $V(x_*, x^0) \leq \Theta_0^2$. Пусть

$$|g(x) - g(y)| \leq M_g \|x - y\| \quad \forall x, y \in Q. \quad (51)$$

Тогда справедлив следующий результат об оценке качества найденного решения предложенного метода (в [21] имеется полное доказательство в случае евклидовой прокс-структуры, для произвольной прокс-структуры это обоснование аналогично).

Теорема 7. После остановки предложенного алгоритма 4 справедливы неравенства

$$f(\hat{x}) - f(x_*) \leq \varepsilon \text{ и } g(\hat{x}) \leq \varepsilon M_g.$$

На базе теоремы 7 в предположении липшицевости целевого функционала

$$|f(x) - f(y)| \leq M_f \|x - y\| \quad (52)$$

можно оценить количество итераций, необходимых для выполнения критерия остановки (1). Ясно, что $\forall k \in I \|\nabla f(x^k)\|_* \leq M_f$, и поэтому

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

Это означает, что при

$$N \geq \frac{2\Theta_0^2 \max\{1, M_f^2\}}{\varepsilon^2}$$

критерий останковки (1) заведомо выполнен, то есть искомая точность достигается за $O\left(\frac{1}{\varepsilon^2}\right)$ итераций.

По аналогии с ([20], теорема 3.2) можно проверить следующий результат, означающий прямо-двойственность алгоритма 4 в случае ограниченного множества Q .

Теорема 8. Пусть Q ограничено и $\max_{x \in Q} d(x) \leq \Theta_0^2$. Тогда если применить алгоритм 4 к задаче

$$\varphi(\lambda) = \min_{x \in Q} \left\{ f(x) + \sum_{i=1}^m \lambda_i g_i(x) \right\} \rightarrow \max_{\lambda_i \geq 0, i=1, \dots, m},$$

то после останковки для найденной пары $(\hat{x}^k, \hat{\lambda}^k)$ будет верно:

$$f(\hat{x}^k) - \varphi(\hat{\lambda}^k) \leq \varepsilon, \quad g(\hat{x}^k) \leq M_g \varepsilon.$$

Похожий на алгоритм 4 метод можно предложить и для задач со следующими релаксациями известного неравенства Коши–Буняковского для субградиентов $\nabla f(x)$ и $\nabla g(x)$:

$$\langle \nabla f(x), x - y \rangle \leq \omega \|\nabla f(x)\|_* \sqrt{2V(y, x)} \quad \forall y \in Q \quad (53)$$

и всякого $x \in Q$ такого, что $V(x_*, x) \geq \frac{\varepsilon^2}{2}$, где x_* – ближайшее решение к начальной точке x_0 с точки зрения дивергенции V , $\omega > 0$ – некоторая фиксированная постоянная; а также

$$\langle \nabla g(x), x - y \rangle \leq M_g \sqrt{2V(y, x)} \quad \forall x, y \in Q \quad (54)$$

для константы $M_g > 0$.

Ясно, что при $V(y, x) \geq \frac{1}{2} \|y - x\|^2$ неравенство (53) верно для $\omega = 1$. Неравенство (54), в частности, верно в случае относительной липшицевости g [12]. Рассмотрим следующий алгоритм 5.

Тогда из (53) и (54) имеем

$$h_k \langle \nabla f(x^k), x^k - x_* \rangle \leq \frac{\omega^2 \varepsilon^2}{2 \|\nabla f(x^k)\|_*^2} + V(x_*, x^k) - V(x_*, x^{k+1}) \quad \forall k \in I,$$

$$\varepsilon^2 < h_k g(x^k) \leq h_k \langle \nabla g(x^k), x^k - x_* \rangle \leq \frac{\varepsilon^2}{2} + V(x_*, x^k) - V(x_*, x^{k+1}) \quad \forall k \in J.$$

После суммирования указанных неравенств получаем

$$\sum_{k \in I} h_k (f(x^k) - f(x_*)) \leq \frac{\omega^2 \varepsilon}{2} \sum_{k \in I} h_k - \frac{\varepsilon^2}{2} |J| + V(x_*, x^0).$$

Теорема 9. После выполнения критерия останковки (55) справедлива оценка:

$$f(\hat{x}) - f(x_*) \leq \omega^2 \varepsilon \text{ и } g(\hat{x}) \leq \varepsilon M_g \quad (56)$$

или $V(x_*, x^k) < \frac{\varepsilon^2}{2}$ для некоторого k .

Замечание 6. Вместо условия (53) можно рассмотреть и неравенство $\langle \nabla f(x), x - y \rangle \leq M_f \sqrt{2V(y, x)}$, которое верно в случае относительной липшицевости f [12]. Для этого необходимо выбирать в алгоритме 5 продуктивные шаги $h_k^f = \frac{\varepsilon}{M_f^2}$, а также критерий останковки

$$2V(x_*, x^0) = 2\Theta_0^2 \leq \frac{\varepsilon^2 |I|}{M_f^2} + \varepsilon^2 |J|.$$

Алгоритм 5 Адаптивный зеркальный спуск**Require:** $\varepsilon > 0$, $\Theta_0 : d(x_*) \leq \Theta_0^2$

- 1: $x^0 = \operatorname{argmin}_{x \in Q} d(x)$
- 2: $I =: \emptyset$
- 3: $k \leftarrow 0$
- 4: **repeat**
- 5: **if** $g(x^k) \leq \varepsilon M_g$ **then**
- 6: $h_k^f \leftarrow \frac{\varepsilon}{\|\nabla f(x^k)\|_*^2}$
- 7: $x^{k+1} \leftarrow \operatorname{Mirr}_{x^k}(h_k \nabla f(x^k))$ // "продуктивные шаги"
- 8: $k \rightarrow I$
- 9: **else**
- 10: $h_k^g \leftarrow \frac{\varepsilon}{M_g}$
- 11: $x^{k+1} \leftarrow \operatorname{Mirr}_{x^k}(h_k \nabla g(x^k))$ // "непродуктивные шаги"
- 12: **end if**
- 13: $k \leftarrow k + 1$
- 14: **until**

$$\frac{2\Theta_0^2}{\varepsilon^2} \leq \sum_{k \in I} \frac{\omega^2}{\|\nabla f(x^k)\|_*^2} + |J|, \quad (55)$$

где $|J|$ — количество непродуктивных шагов (мы обозначим через $|I|$ количество продуктивных шагов, то есть $|I| + |J| = N$).

Ensure: $\hat{x} = \frac{1}{\sum_{k \in I} h_k^f} \sum_{k \in I} h_k^f x^k$.

После выполнения этого критерия будет верно неравенство

$$f(\hat{x}) - f(x_*) \leq \varepsilon \text{ и } g(\hat{x}) \leq \varepsilon M_g. \quad (57)$$

Полученный результат позволяет применить предложенную алгоритмическую схему для негладких задач выпуклой однородной оптимизации с относительной точностью. Такая постановка восходит к работам Ю. Е. Нестерова (см. главу 6 диссертации [25]). Как показано Ю. Е. Нестеровым, подход к оценке качества решения задачи с точки зрения именно относительной точности вполне оправдан для разных прикладных задач (линейное программирование, проектирование механических конструкций, задача оптимальной эллипсоидальной аппроксимации выпуклого множества и пр.), если желаемая относительная точность не очень мала. Известно, что достаточно широкий класс задач оптимизации с относительной точностью можно сводить к минимизации выпуклой однородной функции. Итак, рассматривается на выпуклом замкнутом множестве $Q \subset \mathbb{R}^n$ задача минимизации выпуклой однородной функции вида

$$f(x) \rightarrow \min_{x \in Q} \quad (58)$$

с выпуклыми функционалами ограничениями

$$g_p(x) \leq 0, \quad p = \overline{1, m}. \quad (59)$$

Стандартно будем обозначать $g(x) := \max_{1 \leq p \leq m} \{g_p(x)\}$.

С использованием рассмотренных вариантов концепции относительной липшицевости покажем, как можно выписать оценки сходимости для зеркальных спусков с переключениями в случае, когда 0 не обязательно есть внутренняя точка субдифференциала f в нуле. Рассмотрен следующий ослабленный вариант этого условия

$$B_{\gamma_0}^{K^*}(0) \subseteq \partial f(0) \subseteq B_{\gamma_1}^{K^*}(0), \quad (60)$$

где K^* — сопряженный конус к некоторому полунормированному конусу $K \subset \mathbb{R}^n$ с законом сокращения и конус-полунормой $\|\cdot\|_K$ (отличие от обычной полунормы в том, что $\|\alpha x\|_K = \alpha\|x\|_K$ лишь для $\alpha \geq 0$). Здесь под *сопряженным конусом* K^* понимается набор функционалов вида $\psi_\ell = \max\{0, \ell(x)\}$ для линейных функционалов $\ell : K \rightarrow \mathbb{R} : \ell(x) \leq C_\ell\|x\|_K$ при некотором $C_\ell > 0 \forall x \in K$. Ясно [22, 23], что K^* будет выпуклым конусом с операциями сложения

$$\psi_{\ell_1} \oplus \psi_{\ell_2} := \psi : \psi(x) = \max\{0, \ell_1(x) + \ell_2(x)\}$$

и умножения на скаляр $\lambda \geq 0$ $\psi_{\lambda\ell}(x) = \lambda\psi_\ell(x) = \lambda \max\{0, \ell(x)\} \forall x \in K$. На K^* можно ввести норму $\|\psi_\ell\|_{K^*} = \sup_{\|x\|_K \leq 1} \max\{0, \ell(x)\} = \sup_{\|x\|_K \leq 1} \ell(x)$ и шар радиуса r $B_r^{K^*}(0) = \{\psi_\ell \in K^* \mid \|\psi_\ell\|_{K^*} \leq r\}$.

Из аналога теоремы об опорном функционале в нормированных конусах [22] получаем, что

$$\|x\|_K = \max_{\psi_\ell \in B_1^{K^*}(0)} \ell(x). \quad (61)$$

Для полунормированного конуса при $\|x\|_K = 0$ достаточно выбрать $\ell \equiv 0$, и (61) будет верно. Приведем некоторый пример пары (K, K^*) .

Пример 4. Пусть $K = \{(x, y) \mid x, y \in \mathbb{R}\}$ и $\|(x, y)\|_K = \sqrt{x^2 + y^2} + y$. Можно проверить, что в таком случае $K^* = \{\psi_\ell \mid \ell((x, y)) = \lambda x + \mu y : \mu + \frac{\lambda^2}{\mu} < +\infty \text{ или } \lambda = \mu = 0\}$, а

$$\|\psi_\ell\|_{K^*} = \begin{cases} 0, & \text{если } \lambda = \mu = 0; \\ \frac{\mu}{2} + \frac{\lambda^2}{2\mu}, & \text{если } \mu + \frac{\lambda^2}{\mu} < +\infty \text{ при } \mu > 0. \end{cases}$$

Тогда $B_1^{K^*}(0)$ имеет вид круга на плоскости (λ, μ) радиуса 1 с центром в точке $\lambda = 0, \mu = 1$. Не уменьшая общности рассуждений, будем полагать $K = \bigcup_{r \geq 0} B_r^K(0)$, а также $x_* \in K$ для точного решения x_* рассматриваемой задачи минимизации f на Q .

Согласно схеме рассуждений ([25], глава 6) для вывода оценок скорости сходимости методов с относительной точностью необходимо знать оценку R величины расстояния от точки старта x^0 до ближайшего решения x_* . Однако в конусах, вообще говоря, не задана операция вычитания и поэтому в качестве аналога нормы разности можно использовать метрику $d^K(x^0, x_*)$, где

$$d^K(x, y) = \sup_{\|\psi_\ell\|_{K^*} \leq 1} |\psi_\ell(x) - \psi_\ell(y)| \quad \forall x, y \in Q.$$

Некоторые условия, при которых нормированный конус допускает существование метрики такого типа, исследованы автором работы в [22–24].

Получен аналог теоремы 6.1.1 [25] для указанного выше предположения $(x_0, x_* \in K)$.

Теорема 10. 1) $\forall x \in K \quad \gamma_0\|x\|_K \leq f(x) \leq \gamma_1\|x\|_K$. Более того,

$$\alpha f(x^0) \leq \gamma_0\|x^0\|_K \leq f(x_*) \leq f(x^0) \leq \gamma_1\|x^0\|_K.$$

2) Для всякого точного решения $x_* \in K$ справедливо неравенство

$$d^K(x^0, x_*) \leq \|x^0\|_K + \|x_*\|_K \leq \frac{2}{\gamma_0} f^* \leq \frac{2}{\gamma_0} f(x^0).$$

Для применимости к поставленной задаче приведенного выше метода зеркального спуска (алгоритм 5) с предложенными вариантами условий относительной липшицевости достаточно выбрать прокс-структуру так, чтобы

$$V(x_*, x^0) \leq \hat{\omega} d^K(x_*, x^0), \quad (62)$$

для некоторой постоянной $\hat{\omega} > 0$. Этого можно добиться, например, такими способами:

1. Если $x^0 = 0$, то при $d(x) = \|x\|_K$ можно положить $V(x_*, 0) = \|x_*\|_K = d^K(x_*, 0)$.
2. Если $x^0 = \arg \min_{x \in Q} d(x)$, то для некоторого субградиента $\nabla d(x^0)$ верно неравенство $\langle \nabla d(x^0), x_* - x^0 \rangle \geq 0$ и тогда $V(x_*, x^0) \leq d(x_*) - d(x^0)$. Если выбрать $d(x) = \|x\|_K$, то согласно теореме об опорном функционале в полунормированных конусах

$$\begin{aligned} V(x_*, x^0) &= \|x_*\|_K - \|x^0\|_K = \max\{0, \ell(x_*)\} - \|x^0\|_K \leq \\ &\leq \max\{0, \ell(x_*)\} - \max\{0, \ell(x^0)\} \leq d^K(x_*, x^0), \end{aligned}$$

то есть (62) выполнено при $\widehat{\omega} = 1$. Полагаем далее $\widehat{\omega} = 1$ (другое значение этой постоянной может привести лишь к изменению в итоговых оценках констант).

Поскольку функционал f однороден, то при условии ограниченности субдифференциала $\partial f(0)$ для некоторой константы $M_f > 0$ будет верно $\|\nabla f(x)\|_* \leq M_f$. Поэтому критерий останова модифицированного алгоритма 5 в соответствии с замечанием 6 заведомо будет выполнен после $2\Theta_0^2 \max\{1, M_f^2\} \varepsilon^{-2}$ итераций. Будем полагать, что $\Theta_0^2 = d^K(x^0, x_*) \geq V(x_*, x^0)$ и для некоторого N (число итераций) выберем $\varepsilon = \frac{\Theta_0^2}{\sqrt{N}}$. При данном ε после останова модифицированного алгоритма 5 будут верны неравенства $f(\widehat{x}) - f(x_*) \leq \frac{\Theta_0^2 \omega^2}{\sqrt{N}}$ и $g(\widehat{x}) \leq \frac{\Theta_0^2}{\sqrt{N}}$. Теперь согласно замечанию 6 имеем

$$\frac{\Theta_0^2}{\sqrt{N}} \leq \frac{2f(x_*)}{\gamma_0 \sqrt{N}}, \text{ то есть } f(\widehat{x}) \leq f(x_*) \left(1 + \frac{2}{\gamma_0 \sqrt{N}}\right) \text{ и } g(\widehat{x}) \leq \frac{M_g \Theta_0^2}{\sqrt{N}}.$$

Поэтому для достижения относительной точности $\delta > 0$ по функции заведомо достаточно

$$N \geq \frac{4}{\gamma_0^2 \delta^2}$$

шагов модифицированного алгоритма 5 с постоянными продуктивными шагами согласно замечанию 6. Отметим, что критерий останова при этом заведомо выполнен при $N \geq \frac{8}{\gamma_0^2 \delta^2}$, поскольку

$$\frac{2\Theta_0^2 \max\{1, M_f^2\}}{\frac{\Theta_0^4}{N}} = \frac{2N \max\{1, M_f^2\}}{\Theta_0^2} \geq \frac{8 \max\{1, M_f^2\}}{\gamma_0^2 \delta^2},$$

откуда вытекает следующая

Теорема 11. Пусть однородный выпуклый функционал f удовлетворяет предположению замечания 6 и при этом $M_f > 0$ такова, что $\|\nabla f(x)\|_* \leq M_f$. Тогда после $N \geq \frac{8}{\gamma_0^2 \delta^2}$ итераций модифицированного алгоритма 5 с постоянными продуктивными шагами согласно замечанию 6 гарантированно будет верно неравенство

$$f(\widehat{x}) \leq f(x_*)(1 + \delta) \text{ и } g(\widehat{x}) \leq \frac{M_g \Theta_0^2}{\sqrt{N}}.$$

Заключение

В настоящей статье предложены некоторые адаптивные методы градиентного типа, которые применимы к задачам негладкой оптимизации.

В первых трёх разделах рассмотрены подходы, основанные на введении искусственных погрешностей. Погрешности могут быть как естественными (задание целевой функции и градиента), так и искусственными (сведение негладкой задачи к гладкой с некоторой неточностью). При этом одна из основных отличительных особенностей предлагаемых подходов — наличие в оптимизационной модели целевого функционала переменной величины, соответствующей погрешности (суб)градиента или степени негладкости задачи.

Получены оценки скорости сходимости для адаптивного неускоренного градиентного метода с адаптивной настройкой на уровень гладкости задачи, а также величины погрешностей. Рассмотрен также вариант быстрого градиентного метода Ю.Е. Нестерова для соответствующей концепции неточной модели оптимизируемого функционала. Доказано, что для неускоренного метода не накапливаются все типы погрешностей модели. Для ускоренного метода обоснована возможность уменьшения влияния переменной погрешности на оценку качества решения до любой приемлемой величины при накоплении величин, соответствующих постоянным значениям величин погрешностей используемой концепции модели оптимизируемого функционала. При этом адаптивность метода может на практике улучшать качество найденного решения по сравнению с полученными теоретическими оценками. Однако в полученных оценках качества решения реализована адаптивная настройка не всех параметров неточной модели. Полная адаптивная настройка на величины погрешностей возможна для искусственных неточностей, связанных с рассмотрением негладких задач.

Обоснована применимость неускоренных процедур для относительно гладких целевых функционалов. В таком случае полученную оценку скорости сходимости $O(\varepsilon^{-1})$ можно считать оптимальной даже при отсутствии погрешностей [26]. Показано, как можно ввести аналогичную концепцию неточной модели для вариационных неравенств и седловых задач и предложить аналог экстраградиентного метода с адаптивной настройкой на величину детерминированного (как постоянного, так и переменного) шума.

В последнем шестом разделе работы рассмотрены адаптивные алгоритмические схемы с переключениями для негладких задач выпуклой оптимизации с липшицевыми ограничениями, которые достаточно эффективно работают для некоторых задач с целевыми функционалами более низкого уровня гладкости. В частности, речь может идти о задачах с дифференцируемыми гёльдеровыми целевыми функционалами [21] или с относительно липшицевыми целевыми функционалами [12] и функционалами ограничений. Введённые релаксации условия Липшица, в частности, позволяют получать оценки скорости сходимости с относительной точностью для однородных целевых функционалов при достаточно общих предположениях. Однако в отличие от результатов первых разделов работы по методам градиентного типа не удалось предложить методы с адаптивной настройкой на величины погрешностей. Тем не менее в таких схемах возможно использование δ -субградиентов вместо обычных субградиентов, а также возмущённых с точностью $\delta > 0$ значений функционалов. В итоговых оценках при этом не происходит накопление величин, соответствующих δ [27]. Это означает, в частности, что в итоговых оценках скорости сходимости не накапливаются величины, соответствующие погрешностям, возникающим при решении вспомогательных задач на итерациях алгоритма 5. Указанная методика применима к задачам с любым количеством функционалов ограничений. В этом плане интересной задачей может быть сравнение разработанной методики с подходом к негладким задачам выпуклого программирования с одним функционалом ограничения, который основан на переходе к одномерной двойственной задаче. При этом нахождение подходящего значения двойственного множителя может выполняться методом дихотомии [28] при допустимой погрешности значения производной, связанной с неточностью решения вспомогательных подзадач. Для задач с двумя ограничениями можно применять методику [29] с критерием останова, похожим на критерий останова алгоритма 1 из [28] и соответствующим подходящему значению возмущённого градиента двойственной задачи. Экспериментально показано, что такие подходы могут приводить к линейной скорости сходимости даже для негладких задач. Если же целевой функционал гладкий и сильно выпуклый, а функционалы ограничений гладкие и выпуклые, то линейную скорость сходимости можно обосновать по схеме рассуждений [28].

Автор благодарит П. Е. Двуреченского и А. В. Гасникова за указание некоторых литературных ссылок и полезные обсуждения.

Работа выполнена при поддержке гранта РФФИ мол-а-вед (разделы 1 и 2), гранта РФ 18-71-00048 (разделы 3 и 4, теоремы 7 и 8), а также гранта Президента РФ МК-15.2020.1 (теоремы 10 и 11).

Литература

1. *Немировский А.С., Юдин Д.Б.* Сложность задач и эффективность методов оптимизации. Москва : Наука, 1979. 384 с.
2. *Nesterov Yu.* Universal gradient methods for convex optimization problems // *Math. Program. Ser. A.* 2015. V. 152, N 1–2. P. 381–404.
3. *Стонякин Ф.С.* Аналог квадратичной интерполяции для специального класса негладких функционалов и одно его приложение к адаптивному методу зеркального спуска // *Динамические системы.* 2019. Т. 9(37), № 1. С. 3–16.
4. *Стонякин Ф.С.* Адаптация к погрешности для некоторых методов градиентного типа // *Труды ИММ УрО РАН.* 2019. Т. 25, № 4. С. 210–225.
5. *Devolder O., Glineur F., Nesterov Yu.* First-order methods of smooth convex optimization with inexact oracle // *Math. Program.* 2014. V. 146, N 1–2. n P. 37–75.
6. *Devolder O.* Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization // *PhD thesis,* 2013. 309 p.
7. *Tyurin A.I., Gasnikov A.V.* Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L) -model of a function in a requested point // *Computational Mathematics and Mathematical Physics.* 2019. V.59, N 7. P. 1137–1150.
8. *Stonyakin F.S., Dvinskikh D., Dvurechensky P., Kroshnin A., Kuznetsova O., Agafonov A., Gasnikov A., Tyurin A., Uribe C.A., Pasechnyuk D., Artamonov S.* Gradient Methods for Problems with Inexact Model of the Objective // In: *Khachay M., Kochetov Y., Pardalos P.* (eds) *Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science.* 2019. V. 11548. P. 97–114.
9. *Bauschke H.H., Bolte J., Teboulle M.* A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications // *Mathematics of Operations Research.* 2017. V. 42, N 2. P. 330–348.
10. *Necoara I., Nesterov Y., Glineur F.* Linear convergence of first order methods for non-strongly convex optimization // *Math. Program.* 2019. V. 175. P. 69–107.
11. *Lu H., Freund R.M., Nesterov Y.* Relatively smooth convex optimization by Firstorder methods, and applications // *SIAM Journal on Optimization.* 2018. V. 28, N 1. P. 333–354.
12. *Lu H.* «Relative-Continuity»for Non-Lipschitz Non-Smooth Convex Optimization using Stochastic (or Deterministic) Mirror Descent // *Arxiv preprint.* 2018. 22 p. Available at: <https://arxiv.org/abs/1710.04718v3>.
13. *Stonyakin F., Stepanov A., Titov A., Gasnikov A.* Mirror Descent for Constrained Optimization Problems with Large Subgradient Values // *Copmputer Research and Modelling.* 2020. V. 12, N 2. 23 p. Available at: <https://arxiv.org/abs/1908.00218v4>.
14. *Lan G.* Gradient sliding for composite optimization // *Math. Program.* 2016. V. 159, N 1–2. P. 201–235.
15. *Нестеров Ю.Е.* Метод минимизации выпуклых функций со скоростью сходимости $O(1/k^2)$ // *Доклады АН СССР.* 1983. Т. 269, № 3. С. 543–547.
16. *Stonyakin F., Vorontsova E., Alkousa M.* New Version of Mirror Prox for Variational Inequalities with Adaptation to Inexactness // *Communications on Computer and Information Sciences.* 2020. V. 1145. P. 427–442.

17. *Стонякин Ф.С.* Об адаптивном проксимальном методе для некоторого класса вариационных неравенств и смежных задач // Труды ИММ УрО РАН. 2019. Т. 25, № 2. С. 185–197.
18. *Поляк Б.Т.* Один общий метод решения экстремальных задач // Докл. АН СССР. 1967. Т. 174, № 1. С. 33–36.
19. *Шор Н.З.* Применение обобщённого градиентного спуска в блочном программировании // Кибернетика. 1967. № 3. С. 53–55.
20. *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. Lecture Notes in Math. 2018. V. 2227. P. 181–213.
21. *Ivanova A., Stonyakin F., Pasechnyuk D., Vorontsova E., Gasnikov A.* Adaptive Mirror Descent for the Network Utility Maximization Problem // Arxiv preprint. 2019. 7 p. Available at: <https://arxiv.org/abs/1911.07354v2>.
22. *Stonyakin F.S.* An analogue of the Hahn–Banach theorem for functionals on abstract convex cones // Eurasian Math. J. 2016. V.7, N 3. P. 89–99.
23. *Стонякин Ф.С.* Сублинейный аналог теоремы Банаха–Мазура в отделимых выпуклых конусах с нормой // Матем. заметки. 2018. Т. 104, № 1. С. 118–130.
24. *Stonyakin F.S.* Hahn–Banach type theorems on functional separation for convex ordered normed cones // Eurasian Math. J. 2019. V. 10, N 1. P. 59–79.
25. *Нестеров Ю.Е.* Алгоритмическая выпуклая оптимизация / Дисс. ... докт. физ.-мат. наук. Москва : МФТИ, 2013. 367 с.
26. *Dragomir R.-A., Taylor A., d’Áspremont A., Bolte J.* Optimal Complexity and Certification of Bregman First-Order Methods. Arxiv preprint. 2019. 32 p. Available at: <https://arxiv.org/abs/1911.08510>.
27. *Dvurechensky P., Gasnikov A., Nurminsky E. and Stonyakin F.* Advances in Low-Memory Subgradient Optimization. A.M. Bagirov et al. (eds.), Numerical Nonsmooth Optimization, Springer Nature Switzerland AG. 2020. 36 p. Available at: <https://arxiv.org/abs/1902.01572v1>.
28. *Stonyakin F.S., Alkousa M.S., Titov A.A., Piskunova V.V.* On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. M. Khachay et al. (eds.). Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science. 2019. V. 11548. P. 82–96.
29. *Пасечнюк Д.А., Стонякин Ф.С.* Об одном методе минимизации выпуклой липшицевой функции двух переменных на квадрате // Компьютерные исследования и моделирование. 2019. Т. 11, № 3. С. 379–395.

References

1. *Nemirovsky A.S., Yudin D.B.* Problem Complexity and Method Efficiency in Optimization. Moscow : Nauka, 1979. 384 p. (in Russian).
2. *Nesterov Yu.* Universal gradient methods for convex optimization problems. Math. Program. Ser. A. 2015. V. 152, N 1–2. P. 381–404.
3. *Stonyakin F.S.* An analog of quadratic interpolation for a special class of nonsmooth functionals and one of its applications to the adaptive method of mirror descent. Dynamical systems. 2019. V. 9 (37), N 1. P. 3–16. (in Russian).
4. *Stonyakin F.S.* Adaptation to inexactness for some gradient-type optimization methods. Proceedings of the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences. 2019. V. 25, N 4. P. 210–225. (in Russian).

5. *Devolder O., Glineur F., Nesterov Yu.* First-order methods of smooth convex optimization with inexact oracle. *Math. Program.* 2014. V. 146, N 1–2. P. 37–75.
6. *Devolder O.* Exactness, Inexactness and Stochasticity in First-Order Methods for Large-Scale Convex Optimization. PhD thesis, 2013. 309 p.
7. *Tyurin A.I., Gasnikov A.V.* Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L) -model of a function in a requested point. *Computational Mathematics and Mathematical Physics.* 2019. V. 59, N 7. P. 1137–1150.
8. *Stonyakin F.S., Dvinskikh D., Dvurechensky P., Kroshnin A., Kuznetsova O., Agafonov A., Gasnikov A., Tyurin A., Uribe C.A., Pasechnyuk D., Artamonov S.* Gradient Methods for Problems with Inexact Model of the Objective. *Khachay M., Kochetov Y., Pardalos P.* (eds) *Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science.* 2019. V. 11548. P. 97–114.
9. *Bauschke H.H., Bolte J., Teboulle M.* A Descent Lemma Beyond Lipschitz Gradient Continuity: First-Order Methods Revisited and Applications. *Mathematics of Operations Research.* 2017. V. 42, N 2. P. 330–348.
10. *Necoara I., Nesterov Y., Glineur F.* Linear convergence of first order methods for non-strongly convex optimization. *Math. Program.* 2019. V. 175. P. 69–107.
11. *Lu H., Freund R.M., Nesterov Y.* Relatively smooth convex optimization by Firstorder methods, and applications. *SIAM Journal on Optimization.* 2018. V. 28, N 1. P. 333–354.
12. *Lu H.* «Relative-Continuity» for Non-Lipschitz Non-Smooth Convex Optimization using Stochastic (or Deterministic) Mirror Descent. *Arxiv preprint.* 2018. 22 p. Available at: <https://arxiv.org/abs/1710.04718v3>.
13. *Stonyakin F., Stepanov A., Titov A., Gasnikov A.* Mirror Descent for Constrained Optimization Problems with Large Subgradient Values. *Copmuter Research and Modelling.* 2020. V. 12, N 2. 23 p. Available at: <https://arxiv.org/abs/1908.00218v4>.
14. *Lan G.* Gradient sliding for composite optimization. *Math. Program.* 2016. V. 159, N 1–2. P. 201–235.
15. *Nesterov Yu.E.* A minimization method for convex functions with a convergence rate $O(1/k^2)$. *Dokl. Akad. Nauk SSSR.* 1983. V. 269, N 3. P. 543–547. (in Russian).
16. *Stonyakin F., Vorontsova E., Alkousa M.* New Version of Mirror Prox for Variational Inequalities with Adaptation to Inexactness. *Communications on Computer and Information Sciences.* 2020. V. 1145. P. 427–442.
17. *Stonyakin F.S.* On the adaptive proximal method for a class of variational inequalities and related problems. *Proceedings of the Institute of Mathematics and Mechanics, Ural Branch of the Russian Academy of Sciences.* 2019. V. 25, N 2. P. 185–197. (in Russian).
18. *Polyak B.T.* One general method for solving extreme problems. *Dokl. Akad. Nauk SSSR.* 1967. V.174, N 1. P. 33–36. (in Russian).
19. *Shor N.Z.* Application of generalized gradient descent in block programming. *Cybernetics.* 1967. № 3. P. 53–55. (in Russian).
20. *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. Lecture Notes in Math.* 2018. V. 2227. P. 181–213.
21. *Ivanova A., Stonyakin F., Pasechnyuk D., Vorontsova E., Gasnikov A.* Adaptive Mirror Descent for the Network Utility Maximization Problem. *Arxiv preprint.* 2019. 7 p. Available at: <https://arxiv.org/abs/1911.07354v2>.
22. *Stonyakin F.S.* An analogue of the Hahn–Banach theorem for functionals on abstract convex cones. *Eurasian Math. J.* 2016. V.7, N 3. P. 89–99.

23. *Stonyakin F.S.* A sublinear analogue of the Banach–Mazur theorem in separable convex cones with norm. *Math. Notes.* 2018. V. 104, N 1. P. 118–130.
24. *Stonyakin F.S.* Hahn–Banach type theorems on functional separation for convex ordered normed cones. *Eurasian Math. J.* 2019. V.10, N 1. P. 59–79.
25. *Nesterov Yu.E.* Algorithmic convex optimization. Doctoral thesis. Moscow, MIPT, 2013. 367 p. (in Russian).
26. *Dragomir R.-A., Taylor A., d’Áspremont A., Bolte J.* Optimal Complexity and Certification of Bregman First-Order Methods. Arxiv preprint. 2019. 32 p. Available at: <https://arxiv.org/abs/1911.08510>.
27. *Dvurechensky P., Gasnikov A., Nurminsky E. and Stonyakin F.* Advances in Low-Memory Subgradient Optimization. In: A. M. Bagirov et al. (eds.), *Numerical Nonsmooth Optimization*, Springer Nature Switzerland AG. 2020. 36 p. Available at: <https://arxiv.org/abs/1902.01572v1>.
28. *Stonyakin F.S., Alkousa M.S., Titov A.A., Piskunova V.V.* On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. M. Khachay et al. (eds.). *Mathematical Optimization Theory and Operations Research. MOTOR 2019. Lecture Notes in Computer Science.* 2019. V. 11548. P. 82–96.
29. *Pasechnyuk D.A., Stonyakin F.S.* One method for minimization a convex Lipschitz-continuous function of two variables on a fixed square. *Computer Research and Modeling.* 2019. V. 11, N 3. P. 379–395. (in Russian).

Поступила в редакцию 26.12.2019