**КОМПЬЮТЕРНЫЕ ИССЛЕДОВАНИЯ
И МОДЕЛИРОВАНИЕ 2020 Т. 12 № 2 С. 263–274**

DOI: 10.20537/2076-7633-2020-12-2-263-274

МАТЕМАТИЧЕСКИЕ ОСНОВЫ И ЧИСЛЕННЫЕ МЕТОДЫ МОДЕЛИРОВАНИЯ

УДК: 519.85

Прямо-двойственный быстрый градиентный метод с моделью

А. И. Тюрин

Национальный исследовательский университет «Высшая школа экономики»,
Россия, 101000, г. Москва, ул. Мясницкая, д. 20

E-mail: alexandertiurin@gmail.com

*Получено 24.06.2019, после доработки — 07.01.2020.
Принято к публикации 18.02.2020.*

В данной работе рассматривается возможность применения концепции (δ, L)-модели функции для оптимизационных задач, в которых посредством решения прямой задачи имеется необходимость восстанавливать решение двойственной задачи. Концепция (δ, L)-модели основана на концепции (δ, L)-оракула, предложенной Деволдером – Глинером – Нестеровым, при этом данные авторы предложили фукнционалы в оптимизационных задачах аппроксимировать сверху выпуклой параболой с некоторым аддитивным шумом δ; таким образом, им удалось получить квадратичные верхние оценки с шумом даже для негладких функционалов. Концепция (δ, L)-модели продолжает эту идею за счет того, что аппроксимация сверху делается не выпуклой параболой, а некоторым более сложным выпуклым функционалом. Возможность восстанавливать решение двойственной задачи хорошо зарекомендовала себя, так как во многих случаях в прямой задаче можно значительно быстрее находить решение, чем в двойственной. Отметим, что прямо-двойственные методы хорошо изучены, но при этом, как правило, каждый метод предлагается под конкретный класс задач. Наша же цель — предложить метод, который бы включал в себя сразу различные методы. Это реализуется за счет использования концепции (δ, L)-модели и адаптивной структуры наших методов. Таким образом, нам удалось получить прямо-двойственный адаптивный градиентный метод и быстрый градиентный метод с (δ, L)-моделью и доказать оценки сходимости для них, причем для некоторых классов задач данные оценки являются оптимальными. Основная идея заключается в том, что нахождение двойственных решений происходит относительно оптимизационной задачи, которая аппроксимируют прямую с помощью концепции (δ, L)-модели и имеет более простую структуру, поэтому находить двойственное решение у нее проще. Стоит отметить, что это происходит на каждом шаге работы оптимизационного метода; таким образом, реализуется принцип «разделяй и властвуй».

Ключевые слова: быстрый градиентный метод, модель функции, прямо-двойственный метод

Работа была поддержана грантом РФФИ 18-31-20005 мол-а-вед в первой части и грантом РНФ 17-11-01027 во второй.

COMPUTER RESEARCH AND MODELING
2020 VOL. 12 NO. 2 P. 263–274
DOI: 10.20537/2076-7633-2020-12-2-263-274

MATHEMATICAL MODELING AND NUMERICAL SIMULATION

UDC: 519.85

Primal-dual fast gradient method with a model

A. I. Tyurin

National Research University Higher School of Economics,
20 Myasnitskaya st., Moscow, 101000, Russia

E-mail: alexandertiurin@gmail.com

Received 24.06.2019, after completion — 07.01.2020.
Accepted for publication 18.02.2020.

In this work we consider a possibility to use the conception of (δ, L)-model of a function for optimization tasks, whereby solving a primal problem there is a necessity to recover a solution of a dual problem. The conception of (δ, L)-model is based on the conception of (δ, L)-oracle which was proposed by Devolder – Glineur – Nesterov, herewith the authors proposed approximate a function with an upper bound using a convex quadratic function with some additive noise δ. They managed to get convex quadratic upper bounds with noise even for nonsmooth functions. The conception of (δ, L)-model continues this idea by using instead of a convex quadratic function a more complex convex function in an upper bound. Possibility to recover the solution of a dual problem gives great benefits in different problems, for instance, in some cases, it is faster to find a solution in a primal problem than in a dual problem. Note that primal-dual methods are well studied, but usually each class of optimization problems has its own primal-dual method. Our goal is to develop a method which can find solutions in different classes of optimization problems. This is realized through the use of the conception of (δ, L)-model and adaptive structure of our methods. Thereby, we developed primal-dual adaptive gradient method and fast gradient method with (δ, L)-model and proved convergence rates of the methods, moreover, for some classes of optimization problems the rates are optimal. The main idea is the following: we find a dual solution to an approximation of a primal problem using the conception of (δ, L)-model. It is much easier to find a solution to an approximated problem, however, we have to do it in each step of our method, thereby the principle of "divide and conquer" is realized.

Keywords: fast gradient method, model of the function, primal-dual method

Citation: *Computer Research and Modeling*, 2020, vol. 12, no. 2, pp. 263–274 (Russian).

This work was supported by RFFI 18-31-20005 mol-a-ved in the first part of the work and by RSCF grant No. 17-11-01027 in the second part of the work.

Введение

Методы оптимизации играют большую роль в решении различных задач. Важным свойством некоторых оптимизационных методов является их прямо-двойственность [Аникин и др., 2017; Boyd, 2004; Nesterov, 2018; Nesterov, 2009]: это возможность восстанавливать достаточно эффективно решение двойственной задачи по прямой (или наоборот). Данный подход хорошо себя зарекомендовал в транспортных задачах [Баймурзина и др., 2018; Гасников и др., 2018; Гасников, 2016], задаче машинного обучения SVM и многих других [Гасников, 2019]. В данной работе мы предлагаем прямо-двойственный адаптивный градиентный и быстрый градиентный метод, использующий концепцию (δ, L)-модели функции [Гасников, 2019; Гасников, Тюрин, 2019], которая в свою очередь основана на концепции (δ, L)-оракула [Devolder et al., 2014; Devolder et al., 2013a; Devolder et al., 2013b; Devolder, 2013]. Как и в ранних работах по (δ, L)-модели, методы из текущей работы включают в себя классический градиентный метод [Нестеров, 2010], универсальный метод [Nesterov, 2015], метод Франка – Вульфа [Nemirovski, 2015], композитную оптимизацию [Nesterov, 2013]. Более того, концепция (δ, L)-модели позволяет решать эффективно достаточно нетривиальные постановки задач [Stonyakin et al., 2019; Гасников, Тюрин, 2019; Гасников, 2019]. Для многих из них предложенные нами методы являются оптимальными [Немировский, Юдин, 1979; Гасников, Тюрин, 2019].

Прямо-двойственный метод

Рассмотрим общую задачу оптимизации [Нестеров, 2010; Васильев, 2011]:

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

Функция $f(x)$ определена на некотором множестве Q, которое принадлежит линейному пространству $\mathbb{R}^n$:

$$f(x)\colon Q \to \mathbb{R}, \quad Q \subset \mathbb{R}^n.$$

Далее и везде будем считать, что функция $f(x)$ выпуклая и на множестве Q имеет хотя бы одну точку минимума, принадлежащую множество Q. Более того, будем предполагать, что множество Q имеет следующий вид:

$$Q = \{x \mid x \in \widetilde{Q},\ f_i(x) \leq 0\ \forall i \in [1, m]\},$$

где для любого i функция $f_i(x)\colon \widetilde{Q} \to \mathbb{R}$ — выпуклая функция, и множество $\widetilde{Q}$ является выпуклым. Введем следующее обозначение:

$$F(x) = [f_1(x), \ldots, f_m(x)]^T,$$

таким образом, получаем следующую задачу оптимизации:

$$f(x) \to \min_{x \in \widetilde{Q},\ F(x) \leq 0}. \tag{2}$$

Далее нам понадобится понятие *прокс-функции* и *дивергенции Брегмана* [Bregman, 1967], [Nemirovski, 2015, с. 327].

Определение 1. Функция $d(x)\colon Q \to \mathbb{R}$ называется прокс-функцией, если $d(x)$ непрерывно дифференцируемая на int Q и $d(x)$ является 1-сильно выпуклой относительно нормы $\|\cdot\|$ на множестве int Q.

Определение 2. Функция

$$V(x, y) = d(x) - d(y) - \langle \nabla d(y), x - y \rangle \tag{3}$$

называется дивергенцией Брегмана, где $d(x)$ — произвольная прокс-функция.

Из 1-сильной выпуклости прокс-функции моментально следует [Nemirovski, 2015, с. 328], что

$$V(x, y) \geq \frac{1}{2}\|x - y\|^2. \tag{4}$$

Введем понятие (δ, L)-модели функции.

Определение 3. Пусть функция $\psi_\delta(x, y)$ выпуклая на множестве Q и выполняется условие $\psi_\delta(x, x) = 0$ для всех $x \in Q$. Будем говорить, что $\psi_\delta(x, y)$ есть (δ, L)-модель функции f в точке y относительно нормы $\|\cdot\|$, если для любого $x \in Q$ неравенство

$$0 \leq f(x) - (f_\delta(y) + \psi_\delta(x, y)) \leq \frac{L}{2}\|x - y\|^2 + \delta \tag{5}$$

выполнено для некоторых $L, \delta > 0$.

Данное определение было введено и ранее в работах [Гасников, 2019; Гасников, Тюрин, 2019; Stonyakin et al., 2019] и базируется на концепции (δ, L)-оракула [Devolder et al., 2013a; Devolder et al., 2014; Devolder, 2013].

Найдем двойственную задачу [Boyd, 2004, с. 215] для задачи (2), для этого выпишем следующее равенство:

$$\min_{x \in \widetilde{Q},\ F(x) \leq 0} f(x) = \min_{x \in \widetilde{Q}} \max_{z \in \mathbb{R}^m_+} [f(x) + \langle z, F(x) \rangle],$$

где $\mathbb{R}^m_+ = \{x \mid x \in \mathbb{R}^n,\ x_i \geq 0\ \forall i \in [1, m]\}$. В силу слабой двойственности [Boyd, 2004, с. 225] будет верно неравенство

$$\min_{x \in \widetilde{Q},\ F(x) \leq 0} f(x) \geq \max_{z \in \mathbb{R}^m_+} \min_{x \in \widetilde{Q}} [f(x) + \langle z, F(x) \rangle].$$

Пусть

$$g(z) = \max_{x \in \widetilde{Q}} [-f(x) - \langle z, F(x) \rangle], \tag{6}$$

тогда

$$\min_{x \in \widetilde{Q},\ F(x) \leq 0} f(x) \geq -\min_{z \in \mathbb{R}^m_+} g(z). \tag{7}$$

Выражение слева называется прямой задачей, а справа — двойственной задачей, определим ее отдельно:

$$g(z) \to \min_{z \in \mathbb{R}^m_+}. \tag{8}$$

Далее будем предполагать, что выполнены условия сильной двойственности [Boyd, 2004, с. 226], одним следствием этого является то, что неравенство в (7) переходит в равенство. Для решений прямых и двойственных задач введем следующее обозначение.

Определение 4. Пусть x_* — произвольное решение прямой задачи

$$p(x) \to \min_{x \in \widetilde{Q},\ G(x) \leq 0}. \tag{9}$$

Точка z_* — произвольное решение двойственной задачи

$$h(z) \to \min_{z \in \mathbb{R}^m_+},$$

для (9), где z — это двойственные переменные, соответствующие ограничениям $G(x) \leq 0$. Введем оператор argdual, зависящий от функций $p(x)$ и $G(x)$ и возвращающий x_* и z_*:

$$(x_*, z_*) := \operatorname*{argdual}_{x \in \widetilde{Q}}(p(x), G(x)).$$

Пусть x_* и z_* — произвольные решения прямой и двойственной задачи из (7); таким образом,

$$(x_*, z_*) := \operatorname*{argdual}_{x \in \widetilde{Q}}(f(x), F(x)).$$

Прежде чем доказывать теорему, рассмотрим алгоритм 1. Этот алгоритм является комбинацией градиентого спуска в модельной общности из работы [Гасников, 2019] с прямо-двойственным субградиентным методом из работы [Nesterov, 2009]. На вход алгоритму подаются начальная точка x_0, произвольная константа $L_0 > 0$ и последовательность $\{\delta_k\}_{k \geq 0}$. Будем предполагать далее, что для δ_k и точки x_k всегда найдется некоторая константа $L_{k+1} > 0$ такая, что существует (δ_k, L_{k+1})-модель в точке x_k. Будем также считать, что данное требование выполнено и для алгоритма 2. Отметим еще, что i_k в шаге 3 алгоритма 1 находится обычным перебором от 0 до бесконечности, но из условия о существовании (δ_k, L_{k+1})-модели в точке x_k следует, что это этот процесс конечен; более того, несложно показать, что в среднем минимальное целое число i_k, для которого выполнено (10), равно 1 (см. [Nesterov, 2015, с. 7–8]).

Алгоритм 1. Прямо-двойственный градиентный метод с моделью функции

Input: x_0 — начальная точка, $L_0 > 0$ и $\{\delta_k\}_{k \geq 0}$

1: $A_0 := 0$

2: **for** $k \geq 0$ **do**

3: Найти минимальное целое число $i_k \geq 0$ такое, что

$$f_{\delta_k}(x_{k+1}) \leq f_{\delta_k}(x_k) + \psi_{\delta_k}(x_{k+1}, x_k) + \frac{L_{k+1}}{2}\|x_{k+1} - x_k\|^2 + \delta_k, \tag{10}$$

где $L_{k+1} := 2^{i_k - 1} L_k$, $A_{k+1} := A_k + \dfrac{1}{L_{k+1}}$

$$\phi_{k+1}(x) := \psi_{\delta_k}(x, x_k) + L_{k+1} V(x, x_k), \quad (x_{k+1}, z_{k+1}) := \operatorname*{argdual}_{x \in \widetilde{Q}}(\phi_{k+1}(x), F(x)) \tag{11}$$

4: **end for**

Output: $\overline{x}_N = \dfrac{1}{A_N} \sum\limits_{k=0}^{N-1} \dfrac{x_{k+1}}{L_{k+1}}$, $\overline{z}_N = \dfrac{1}{A_N} \sum\limits_{k=0}^{N-1} \dfrac{z_{k+1}}{L_{k+1}}$

Лемма 1. *Пусть $\psi(x)$ — выпуклая функция и*

$$y = \operatorname*{argmin}_{x \in Q}\{\psi(x) + V(x, u)\}.$$

Тогда выполнено неравенство

$$\psi(x) + V(x, u) \geq \psi(y) + V(y, u) + V(x, y) \quad \forall x \in Q.$$

Доказательство представлено в работе [Гасников, Тюрин, 2019, лемма 1].

Следствие 1. *Пусть $\psi(x)$ — выпуклая функция и*

$$(y, z) := \operatorname*{argdual}_{x \in \widetilde{Q}}(\psi(x) + V(x, u), F(x)). \tag{12}$$

Тогда выполнено неравенство

$$\psi(x) + \langle z, F(x) \rangle + V(x, u) \geq \psi(y) + V(y, u) + V(x, y) \quad \forall x \in \widetilde{Q}.$$

Доказательство. Из (12) и сильной двойственности следует, что

$$y = \operatorname*{argmin}_{x\in\widetilde{Q}}\{\psi(x) + \langle z, F(x)\rangle + V(x,u)\}. \tag{13}$$

Используя лемму 1 для (13), получаем неравенство

$$\psi(x) + \langle z, F(x)\rangle + V(x,u) \geq \psi(y) + \langle z, F(y)\rangle + V(y,u) + V(x,y) \quad \forall x \in \widetilde{Q}.$$

Из условия дополняющей нежесткости [Boyd, 2004, с. 242] верно, что $\langle z, F(y)\rangle = 0$. Следствие доказано. □

Докажем теорему, которая дает оценки сходимости алгоритма 1.

Теорема 1. *Будем предполагать, что f — выпуклая функция и для δ_k и точки x_k из алгоритма 1 всегда найдется некоторая константа $L_{k+1} > 0$ такая, что существует (δ_k, L_{k+1})-модель в точке x_k. Пусть x_0 — начальная точка; $(\overline{x}_N, \overline{z}_N)$ — точки, полученные в результате работы алгоритма 1; $x(\overline{z}_N)$ — точка, в которой достигается максимум в* (6) *при $z = \overline{z}_N$; тогда будет верно неравенство*

$$f(\overline{x}_N) \leq \min_{x\in\widetilde{Q}}\left[\frac{1}{A_N}\sum_{k=0}^{N-1}\frac{1}{L_{k+1}}(f_{\delta_k}(x_k) + \psi_{\delta_k}(x,x_k)) + \langle\overline{z}_N, F(x)\rangle + \frac{V(x,x_0)}{A_N}\right] + \frac{1}{A_N}\sum_{k=0}^{N-1}\frac{2\delta_k}{L_{k+1}}$$

и

$$f(\overline{x}_N) + g(\overline{z}_N) \leq \frac{V(x(\overline{z}_N), x_0)}{A_N} + \frac{1}{A_N}\sum_{k=0}^{N-1}\frac{2\delta_k}{L_{k+1}}.$$

Доказательство. Отметим, что из требования существования (δ_k, L_{k+1})-модели следует, что процедура подбора параметра i_k конечна, неравенство (10) для достаточно большого i_k выполнится. Рассмотрим следующую цепочку неравенств:

$$\begin{aligned}
f(x_{k+1}) &\overset{(5)}{\leq} f_{\delta_k}(x_{k+1}) + \delta_k \overset{(10)}{\leq}\\
&\overset{(10)}{\leq} f_{\delta_k}(x_k) + \psi_{\delta_k}(x_{k+1}, x_k) + \frac{L_{k+1}}{2}\|x_{k+1} - x_k\|^2 + 2\delta_k \overset{(4)}{\leq}\\
&\overset{(4)}{\leq} f_{\delta_k}(x_k) + \psi_{\delta_k}(x_{k+1}, x_k) + L_{k+1}V(x_{k+1}, x_k) + 2\delta_k \overset{\text{С. 1}}{\leq}\\
&\overset{\text{С. 1}}{\leq} f_{\delta_k}(x_k) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, x_k) + L_{k+1}V(x, x_k) - L_{k+1}V(x, x_{k+1}) + 2\delta_k.
\end{aligned}$$

Поделим неравенства на L_{k+1}, тогда получим

$$\frac{1}{L_{k+1}}f(x_{k+1}) \leq \frac{1}{L_{k+1}}(f_{\delta_k}(x_k) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, x_k) + 2\delta_k) + V(x, x_k) - V(x, x_{k+1}).$$

Если вычислить сумму неравенств по k от 0 до $N - 1$ и поделить на A_N, то будет верно

$$\frac{1}{A_N}\sum_{k=0}^{N-1}\frac{1}{L_{k+1}}f(x_{k+1}) \leq \frac{1}{A_N}\sum_{k=0}^{N-1}\frac{1}{L_{k+1}}(f_{\delta_k}(x_k) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, x_k) + 2\delta_k) + \frac{1}{A_N}(V(x, x_0) - V(x, x_N)).$$

Воспользуемся выпуклостью функции $f(x)$ и $V(x, x_N) \geq 0$, тогда получим

$$f(\overline{x}_N) \leq \frac{1}{A_N}\sum_{k=0}^{N-1}\frac{1}{L_{k+1}}(f_{\delta_k}(x_k) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, x_k) + 2\delta_k) + \frac{V(x, x_0)}{A_N}.$$

Данное неравенство верно для любого $x \in \widetilde{Q}$, поэтому

$$f(\overline{x}_N) \leq \min_{x\in\widetilde{Q}}\left[\frac{1}{A_N}\sum_{k=0}^{N-1}\frac{1}{L_{k+1}}(f_{\delta_k}(x_k) + \psi_{\delta_k}(x,x_k)) + \langle\overline{z}_N, F(x)\rangle + \frac{V(x,x_0)}{A_N}\right] + \frac{1}{A_N}\sum_{k=0}^{N-1}\frac{2\delta_k}{L_{k+1}}.$$

Используя (5), получаем неравенство

$$f(\overline{x}_N) \le \min_{x \in \widetilde{Q}} \left[f(x) + \langle \overline{z}_N, F(x) \rangle + \frac{V(x, x_0)}{A_N} \right] + \frac{1}{A_N} \sum_{k=0}^{N-1} \frac{2\delta_k}{L_{k+1}}.$$

По условию теоремы, $x(\overline{z}_N) = \mathrm{argmax}_{x \in \widetilde{Q}} (-f(x) - \langle \overline{z}_N, F(x) \rangle)$; тогда по определению (6) получаем, что

$$f(\overline{x}_N) + g(\overline{z}_N) \le \frac{V(x(\overline{z}_N), x_0)}{A_N} + \frac{1}{A_N} \sum_{k=0}^{N-1} \frac{2\delta_k}{L_{k+1}}.$$

□

Отметим, что в теореме 1 оценка скорости сходимости зависит от расстояния Брегмана между точками $x(\overline{z}_N)$ и x_0, в то время как обычно скорость сходимости зависит от расстояния Брегмана между x_* и x_0 [Гасников, Тюрин, 2019; Stonyakin et al., 2019; Nesterov, 2015]. Сложность нахождения $x(\overline{z}_N)$ сопоставима со сложностью нахождения x_*. Один из способов оценивать $V(x(\overline{z}_N), x_0)$ или $V(x_*, x_0)$ — это использовать диаметр множества $\widetilde{Q}$, то есть найти такую константу R^2, что $V(x, x_0) \le R^2$ для любого $x \in \widetilde{Q}$. Данное замечание верно для теоремы 2 из следующего раздела.

Быстрый прямо-двойственный метод

В этом параграфе рассмотрим быстрый вариант градиентного метода. Этот алгоритм является комбинацией быстрого градиентого метода в модельной общности из работы [Гасников, Тюрин, 2019, раздел 3] с прямо-двойственным субградиентным методом из работы [Nesterov, 2009]. Сформулируем следующую теорему.

Алгоритм 2. Быстрый прямо-двойственный градиентный метод с моделью функции

Input: x_0 — начальная точка, $L_0 > 0$ и $\{\delta_k\}_{k \ge 0}$

1: $y_0 := x_0$, $u_0 := x_0$, $\alpha_0 := 0$, $A_0 := \alpha_0$
2: **for** $k \ge 0$ **do**
3: Найти минимальное целое число $i_k \ge 0$ такое, что

$$f_{\delta_k}(x_{k+1}) \le f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_{k+1}, y_{k+1}) + \frac{L_{k+1}}{2} \|x_{k+1} - y_{k+1}\|^2 + \delta_k, \tag{14}$$

где $L_{k+1} = 2^{i_k - 1} L_k$

$$\alpha_{k+1} := \frac{1 + \sqrt{1 + 4L_{k+1}A_k}}{2L_{k+1}}, \quad A_{k+1} := A_k + \alpha_{k+1} \tag{15}$$

$$y_{k+1} := \frac{\alpha_{k+1} u_k + A_k x_k}{A_{k+1}} \tag{16}$$

$$\phi_{k+1}(x) = V(x, u_k) + \alpha_{k+1} \psi_{\delta_k}(x, y_{k+1}), \quad (u_{k+1}, z_{k+1}) := \operatorname*{argdual}_{x \in \widetilde{Q}} (\phi_{k+1}(x), F(x)) \tag{17}$$

$$x_{k+1} := \frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}} \tag{18}$$

4: **end for**

Output: x_N, $\overline{z}_N = \frac{1}{A_N} \sum_{k=0}^{N-1} \alpha_{k+1} z_{k+1}$

Теорема 2. *Будем предполагать, что f — выпуклая функция и для δ_k и точки y_{k+1} из алгоритма 2 всегда найдется некоторая константа $L_{k+1} > 0$ такая, что существует (δ_k, L_{k+1})-модель в точке y_{k+1}. Пусть x_0 — начальная точка; $(x_N, \overline{z}_N)$ — точки, полученные в результате работы алгоритма 2; $x(\overline{z}_N)$ — точка, в которой достигается максимум в (6) при $z = \overline{z}_N$; тогда будет верно неравенство*

$$f(x_N) \leq \min_{x \in \widetilde{Q}} \left[\frac{1}{A_N} \sum_{k=0}^{N-1} \alpha_{k+1} \Big(f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x, y_{k+1}) \Big) + \langle \overline{z}_N, F(x) \rangle + \frac{V(x, u_0)}{A_N} \right] + \frac{2}{A_N} \sum_{k=0}^{N-1} A_{k+1} \delta_k$$

и

$$f(x_N) + g(\overline{z}_N) \leq \frac{V(x(\overline{z}_N), x_0)}{A_N} + \frac{2}{A_N} \sum_{k=0}^{N-1} A_{k+1} \delta_k.$$

Доказательство. Отметим, что из требования существования (δ_k, L_{k+1})-модели следует, что процедура подбора параметра i_k конечна, неравенство (14) для достаточно большого i_k выполнится. Рассмотрим следующую цепочку неравенств:

$$\begin{aligned}
f(x_{k+1}) &\overset{(5)}{\leq} f_{\delta_k}(x_{k+1}) + \delta_k \overset{(14)}{\leq} \\
&\overset{(14)}{\leq} f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_{k+1}, y_{k+1}) + \frac{L_{k+1}}{2} \|x_{k+1} - y_{k+1}\|^2 + 2\delta_k \overset{(18)}{=} \\
&\overset{(18)}{=} f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}\left(\frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}}, y_{k+1} \right) + \frac{L_{k+1}}{2} \left\| \frac{\alpha_{k+1} u_{k+1} + A_k x_k}{A_{k+1}} - y_{k+1} \right\|^2 + 2\delta_k \overset{(5),\,(16)}{\leq} \\
&\overset{(5),\,(16)}{\leq} f_{\delta_k}(y_{k+1}) + \frac{\alpha_{k+1}}{A_{k+1}} \psi_{\delta_k}(u_{k+1}, y_{k+1}) + \frac{A_k}{A_{k+1}} \psi_{\delta_k}(x_k, y_{k+1}) + \frac{L_{k+1} \alpha_{k+1}^2}{2A_{k+1}^2} \|u_{k+1} - u_k\|^2 + 2\delta_k = \\
&= \frac{A_k}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_k, y_{k+1})) + \frac{\alpha_{k+1}}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(u_{k+1}, y_{k+1})) + \\
&+ \frac{L_{k+1} \alpha_{k+1}^2}{2A_{k+1}^2} \|u_{k+1} - u_k\|^2 + 2\delta_k.
\end{aligned}$$

Из (15) следует, что $A_{k+1} = L_{k+1} \alpha_{k+1}^2$, поэтому

$$\begin{aligned}
f(x_{k+1}) &= \frac{A_k}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_k, y_{k+1})) + \\
&+ \frac{\alpha_{k+1}}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(u_{k+1}, y_{k+1}) + \frac{1}{2\alpha_{k+1}} \|u_{k+1} - u_k\|^2) + 2\delta_k \overset{(4)}{\leq} \\
&\overset{(4)}{\leq} \frac{A_k}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x_k, y_{k+1})) + \\
&+ \frac{\alpha_{k+1}}{A_{k+1}} (f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(u_{k+1}, y_{k+1}) + \frac{1}{\alpha_{k+1}} V(u_{k+1}, u_k)) + 2\delta_k \overset{\text{С. 1, (5)}}{\leq} \\
&\overset{\text{С. 1, (5)}}{\leq} \frac{A_k}{A_{k+1}} f(x_k) + \frac{\alpha_{k+1}}{A_{k+1}} \Big(f_{\delta_k}(y_{k+1}) + \langle z_{k+1}, F(x) \rangle + \psi_{\delta_k}(x, y_{k+1}) + \\
&+ \frac{1}{\alpha_{k+1}} V(x, u_k) - \frac{1}{\alpha_{k+1}} V(x, u_{k+1}) \Big) + 2\delta_k.
\end{aligned}$$

Умножим неравенство

$$\begin{aligned}
f(x_{k+1}) \leq \frac{A_k}{A_{k+1}} f(x_k) + \frac{\alpha_{k+1}}{A_{k+1}} \Big(& f_{\delta_k}(y_{k+1}) + \langle z_{k+1}, F(x) \rangle + \psi_{\delta_k}(x, y_{k+1}) + \\
&+ \frac{1}{\alpha_{k+1}} V(x, u_k) - \frac{1}{\alpha_{k+1}} V(x, u_{k+1}) \Big) + 2\delta_k
\end{aligned}$$

на A_{k+1}, тогда верно

$$A_{k+1}f(x_{k+1}) \le A_k f(x_k) + \alpha_{k+1}\Big(f_{\delta_k}(y_{k+1}) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, y_{k+1})\Big) + \\ + V(x, u_k) - V(x, u_{k+1}) + 2A_{k+1}\delta_k.$$

Если вычислить сумму неравенств по k от 0 до $N-1$ и поделить на A_N, то будет верно

$$f(x_N) \le \frac{1}{A_N}\sum_{k=0}^{N-1}\alpha_{k+1}\Big(f_{\delta_k}(y_{k+1}) + \langle z_{k+1}, F(x)\rangle + \psi_{\delta_k}(x, y_{k+1})\Big) + \\ + \frac{1}{A_N}(V(x, u_0) - V(x, u_N)) + \frac{2}{A_N}\sum_{k=0}^{N-1}A_{k+1}\delta_k.$$

Данное неравенство верно для любого $x \in \widetilde{Q}$. Кроме того, $V(x, u_N) \ge 0$, поэтому

$$f(x_N) \le \min_{x\in\widetilde{Q}}\left[\frac{1}{A_N}\sum_{k=0}^{N-1}\alpha_{k+1}\Big(f_{\delta_k}(y_{k+1}) + \psi_{\delta_k}(x, y_{k+1})\Big) + \langle \overline{z}_N, F(x)\rangle + \frac{V(x, u_0)}{A_N}\right] + \frac{2}{A_N}\sum_{k=0}^{N-1}A_{k+1}\delta_k.$$

Из (5) получаем неравенство

$$f(x_N) \le \min_{x\in\widetilde{Q}}\left[f(x) + \langle \overline{z}_N, F(x)\rangle + \frac{V(x, u_0)}{A_N}\right] + \frac{2}{A_N}\sum_{k=0}^{N-1}A_{k+1}\delta_k.$$

По условию теоремы, $x(\overline{z}_N) = \operatorname{argmax}_{x\in\widetilde{Q}}(-f(x) - \langle \overline{z}_N, F(x)\rangle)$, тогда по определению (6) получаем, что

$$f(x_N) + g(\overline{z}_N) \le \frac{V(x(\overline{z}_N), u_0)}{A_N} + \frac{2}{A_N}\sum_{k=0}^{N-1}A_{k+1}\delta_k.$$

□

Анализ полученных результатов

Для простоты будем далее считать, что $\delta_k \le \varepsilon/2$ в случае алгоритма 1 и $\delta_k \le \varepsilon\alpha_{k+1}/(2A_{k+1})$ в случае алгоритма 2. Таким образом, мы получим, что слагаемое

$$\frac{1}{A_N}\sum_{k=0}^{N-1}\frac{2\delta_k}{L_{k+1}}$$

из теоремы 1 и слагаемое

$$\frac{2}{A_N}\sum_{k=0}^{N-1}A_{k+1}\delta_k$$

из теоремы 2 будут меньше или равны ε. Стоит отметить, что подобное ограничение на δ_k несильно сужает класс задач; в частности, в универсальных методах [Nesterov, 2015; Гасников, 2019; Гасников, Тюрин, 2019] δ_k имеет именно такие порядки.

В работе [Гасников, Тюрин, 2019] были получены различные оценки на A_N в зависимости от дополнительной информации о функции $f(x)$, в данной работе все они сохраняются. В частности, когда $f(x)$ гладкая, с L-липшицевым градиентом в норме $\|\cdot\|$ (см. [Гасников, Тюрин, 2019, теорема 1, лемма 3]), в случае теоремы 1

$$A_N \ge \frac{N}{2L},$$

а в случае теоремы 2

$$A_N \ge \frac{(N+1)^2}{8L}.$$

Таким образом, алгоритмы 1 и 2 представляют из себя градиентный метод и быстрый градиентный метод соответственно. Отметим, что благодаря адаптивной структуре полученных методов они также являются универсальными [Nesterov, 2015; Гасников, 2019; Гасников, Тюрин, 2019]. Более подробная информация о порядках A_N содержится в работе [Гасников, Тюрин, 2019].

В теореме 2 была получена скорость сходимости зазора двойственности $f(x_N)+g(\overline{z}_N)$. Если взять N таким, чтобы $V(x(\overline{z}_N), x_0)/A_N \leq \varepsilon$, то мы получим неравенство

$$0 \leq f(x_N) + g(\overline{z}_N) \leq 2\varepsilon.$$

Из неравенства (7) мы получаем, что верны два других неравенства:

$$0 \leq f(x_N) - f(x_*) \leq 2\varepsilon, \quad 0 \leq g(\overline{z}_N) - g(z_*) \leq 2\varepsilon,$$

где x_* — оптимальное решение (2), а z_* — оптимальное решение (8). Таким образом, нам удалось доказать, что алгоритм 2 генерирует ε-решение как для прямой, так и для двойственной задачи. Все рассуждения о зазоре двойственности верны и для теоремы 1.

Заключение

В данный работе представлены два метода: прямо-двойственный градиентный и быстрый градиентный метод и доказаны их оценки скорости сходимости. В тех случаях, когда имеются функциональные ограничения в оптимизационной задаче, данные методы являются обобщением алгоритмов, представленных в работе [Гасников, Тюрин, 2019]; более того, все примеры задач из [Гасников, Тюрин, 2019; Stonyakin et al., 2019], которые допускают функциональные ограничения, также применимы и здесь. Данные алгоритмы применимы для задач, где важно одновременно находить решение прямой и двойственной задач, решая только одну из этих задач; например, это используется в транспортных задачах [Баймурзина и др., 2018; Гасников и др., 2018; Гасников, 2016]. Таким образом, предложенные методы и соответствующие теоремы обобщают результаты из [Гасников, Тюрин, 2019; Stonyakin et al., 2019] для задач с функциональными ограничениями за счет введения прямо-двойственности, что позволяет одновременно решать прямые и двойственные задачи в модельной общности, в то время как методы из [Гасников, Тюрин, 2019; Stonyakin et al., 2019] позволяют решать либо прямые, либо двойственные оптимизационные задачи.

Список литературы (References)

Аникин А. С., Гасников А. В., Двуреченский П. Е., Тюрин А. И., Чернов А. В. Двойственные подходы к задачам минимизации сильно выпуклых функционалов простой структуры при аффинных ограничениях // Журнал выч. математики и мат. физики. — 2017. — Т. 57, № 8. — С. 1270–1284.

Anikin A. S., Gasnikov A. V., Dvurechensky P. E., Tyurin A. I., Chernov A. V. Dvoistvenii podhodi k zadacham minimizacii silno vipuclich functionalov prostoi structuri pri affinich ogranicheniyach [Dual approaches to the minimization of strongly convex functionals with a simple structure under affine constraints] // ZhVM & MF [Comp. Math. & Math. Phys.]. — 2017. — Vol. 58, No. 8. — P. 1270–1284 (in Russian).

Баймурзина Д. Р., Гасников А. В., Гасникова Е. В., Двуреченский П. Е., Ершов Е. И., Кубентаева М. Б., Лагуновская А. А. Универсальный метод поиска равновесий и стохастических равновесий в транспортных сетях // Ж. вычисл. матем. и матем. физ. — 2019. — Т. 59, № 1. — С. 21–36.

Baimurzina D. R., Gasnikov A. V., Gasnikova E. V., Dvurechensky P. E., Ershov E. I., Kubentaeva M. B., Lagunovskaya A. A. Universalnyi metod poiska ravnovesii i stohasticheskih ravnovesii v transportnyh setyah [Universal similar triangulars method for searching equilibriums in traffic flow distribution models] // ZhVM & MF [Comp. Math. & Math. Phys.]. — 2019. — Vol. 59, No. 1. — P. 21–36 (in Russian).

Васильев Ф. П. Методы оптимизации. — Т. 2. — М.: МЦНМО, 2011. — 433 с.

Vasiliev F. P. Metody optimizatsii [Methods of Optimization]. — Vol. 2. — Moscow: MTSNMO, 2011. — 433 p. (in Russian).

Гасников А. В. Современные численные методы оптимизации. Метод универсального градиентного спуска // e-print, 2019. — URL: https://arxiv.org/pdf/1711.00394.pdf

Gasnikov A. V. Sovremenii chislenii methodi optimizacii [Universal gradient descent] // e-print, 2019. — URL: https://arxiv.org/pdf/1711.00394.pdf (in Russian).

Гасников А. В. Эффективные численные методы поиска равновесий в больших транспортных сетях: диссертация на соискание ученой степени д. ф.-м. н. по специальности 05.13.18 «Математическое моделирование, численные методы, комплексы программ». — М.: МФТИ, 2016. — 487 с.

Gasnikov A. V. Effektivnyye chislennyye metody poiska ravnovesiya v bol'shikh transportnykh setyakh: dissertatsiya na soiskaniye uchenoy stepeni d. f.-m. n. po spetsial'nosti 05.13.18 [Effective numerical methods for finding equilibrium in large transport networks: thesis for PhD on the specialty 05.13.18] "Matematicheskoye modelirovaniye, chislennyye metody, kompleksy programm" ["Mathematical modeling, numerical methods, program complexes"]. — Moscow: MFTI, 2016. — 487 p. (in Russian).

Гасников А. В., Гасникова Е. В., Нестеров Ю. Е. Двойственные методы поиска равновесий в смешанных моделях распределения потоков в больших транспортных сетях // Журнал выч. математики и мат. физики. — 2018. — Т. 58, № 9. — С. 1447–1454.

Gasnikov A. V., Gasnikova E. V., Nesterov Yu. E. Dvoistvenii metodi poiska ravnovesii v smechanich modelach raspredeliniya potokov v bolshich transpotnich setyach [Dual Methods for Finding Equilibriums in Mixed Models of Flow Distribution in Large Transportation Networks] // ZhVM & MF [Comp. Math. & Math. Phys.]. — 2018. — Vol. 58, No. 9. — P. 1447–1454 (in Russian).

Гасников А. В., Тюрин А. И. Быстрый градиентный спуск для задач выпуклой минимизации с оракулом, выдающим (δ, L)-модель функции в запрошенной точке // Журнал выч. математики и мат. физики. — 2019. — Т. 59, № 7. — С. 1137–1150.

Gasnikov A. V., Tyurin A. I. Bistriy gradientniy spusk dli zadachi vipucloy minimizacii s oraculom, viduachim (δ, L)-model funtcii v zaprochenoy tochke [Fast gradient descent method for convex optimization problems with an oracle that generates a (δ, L)-model of a function in a requested point] // ZhVM & MF [Comp. Math. & Math. Phys.]. — 2019. — Vol. 59, No. 7. — P. 1137–1150 (in Russian).

Немировский А. С., Юдин Д. Б. Сложность задач и эффективность методов оптимизации. — М.: Наука, 1979. — 384 с.

Nemirovsky A. S., Yudin D. B. Slozhnost' zadach i effektivnost' metodov optimizatsii [Problem complexity and method efficiency in optimization]. — Moscow: Nauka, 1979. — 384 p. (in Russian).

Нестеров Ю. Е. Введение в выпуклую оптимизацию. — М.: МЦНМО, 2010. — 262 с.

Nesterov Yu. E. Vvedeniye v vypukluyu optimizatsiyu [Introductory lectures on convex optimization]. — Moscow: MCCME, 2010. — 262 p. (in Russian).

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

Polyak B. T. Vvedeniye v optimizatsiyu [Introductory lectures on convex optimization]. — Moscow: Nauka, 1983. — 384 p. (in Russian).

Boyd S., Vandenberghe L. Convex optimization. — Cambridge University Press, 2004.

Bregman L. The relaxation method of finding the common points of convex sets and its application to the solution of problems in convex programming // USSR Computational Mathematics and Mathematical Physics. — 1967. — Vol. 7, No. 3. — P. 200–217.

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.

Devolder O. Exactness, inexactness and stochasticity in first-order methods for large-scale convex optimization. — PhD thesis, CORE UCL, 2013.

Devolder O., Glineur F, Nesterov Yu. First-order methods with inexact oracle: the strongly convex case // CORE Discussion Papers 2013/16. — 2013. — URL: https://www.uclouvain.be/cps/ucl/doc/core/documents/coredp2013_16web.pdf

Devolder O., Glineur F., Nesterov Yu. Intermediate gradient methods for smooth convex problems with inexact oracle // CORE Discussion Papers 2013/17. — 2013. — URL: https://cdn.uclouvain.be/public/Exports%20reddot/core/documents/coredp2013_17web.pdf

Nemirovski A. Lectures on modern convex optimization analysis, algorithms, and engineering applications. — Philadelphia: SIAM, 2015. — URL: http://www2.isye.gatech.edu/ñemirovs/Lect_ModConvOpt.pdf

Nesterov Yu. Complexity bounds for primal-dual methods minimizing the model of objective function // Mathematical Programming. — 2018. — Vol. 171, No. 1–2. — P. 311–330.

Nesterov Yu. Gradient methods for minimizing composite functions // Mathematical Programming. — 2013. — Vol. 140, No. 1. — P. 125–161.

Nesterov Yu. Primal-dual subgradient methods for convex problems // Mathematical Programming. — 2009. — Vol. 120, No. 1. — P. 221–259.

Nesterov Yu. Universal gradient methods for convex optimization problems // Mathematical Programming. — 2015. — Vol. 152, No. 1–2. — P. 381–404.

Stonyakin F., 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 // Mathematical Optimization Theory and Operations Research. 18th International Conference. — Ekaterinburg, Russia, 2019. — P. 97–114.