

## ЧИСЛЕННЫЕ МЕТОДЫ ДЛЯ ЗАДАЧИ РАСПРЕДЕЛЕНИЯ РЕСУРСОВ В КОМПЬЮТЕРНОЙ СЕТИ<sup>1)</sup>

© 2021 г. Е. А. Воронцова<sup>1</sup>, А. В. Гасников<sup>1,2,3</sup>, П. Е. Двуреченский<sup>3,2</sup>,  
А. С. Иванова<sup>4,\*</sup>, Д. А. Пасечнюк<sup>1</sup>

<sup>1</sup> 141701 Долгопрудный, М.о., Институтский пер., 9, Московский физико-технический институт  
(национальный исследовательский университет), Россия

<sup>2</sup> 127051 Москва, Большой Каретный пер., 19, стр. 1, Институт проблем передачи информации  
им. А.А. Харкевича РАН, Россия

<sup>3</sup> Институт прикладного анализа и стохастики им. К. Вейерштрасса, Берлин, Германия

<sup>4</sup> 109028 Москва, Покровский бульвар, 11, Национальный исследовательский университет  
“Высшая школа экономики”, Россия

\*e-mail: asivanova@hse.ru

Поступила в редакцию 29.11.2019 г.  
Переработанный вариант 10.09.2020 г.  
Принята к публикации 16.09.2020 г.

Рассматривается задача распределения ресурсов в компьютерных сетях с большим числом соединений. Соединения используют для своих целей потребители (пользователи), число которых также может быть очень большим. Для решения двойственной задачи предлагаются следующие численные методы оптимизации: быстрый градиентный метод, стохастический метод проекции субградиента, метод эллипсоидов и метод экстраполяции случайного градиента. Для каждого метода получена оценка скорости сходимости. Также приведены алгоритмы распределенного вычисления шагов рассматриваемых методов при условии приложения их к компьютерным сетям. Отдельное внимание уделено прямо двойственности предложенных алгоритмов. Библиография: 38. Фиг. 1. Табл. 2.

**Ключевые слова:** распределение ресурсов, сети связи, максимизация полезности сети, прямо двойственность, быстрый градиентный метод, стохастический метод проекции субградиента, метод эллипсоидов, метод экстраполяции случайного градиента.

DOI: 10.31857/S0044466921020149

### 1. ВВЕДЕНИЕ

#### 1.1. Мотивация

Рассматривается задача управления современными сетями связи с точки зрения оптимизации и стохастического моделирования. Для решения задач такого рода необходимо представить и проанализировать математическую модель, возникающую при симуляции процесса работы крупномасштабных широкополосных сетей. Ожидается, что в будущих сетях связи появятся приложения, которые смогут изменять свои скорости передачи данных в соответствии с доступной пропускной способностью в сети. В качестве примера такой сети можно привести TSP-трафик через сеть Интернет.

Ключевой вопрос, который мы рассматриваем в этой статье, касается того, как доступная пропускная способность в сети должна быть распределена между конкурирующими потоками. При этом контроль над использованием потребителями доступных пропускных способностей осуществляется посредством корректировки цены на соединение.

Таким образом, в работе рассматривается задача оптимизации распределения ресурсов в компьютерных сетях с большим числом соединений. Соединения используют для своих целей по-

<sup>1)</sup> Работа выполнена при финансовой поддержке РФФИ авторов: А.В. Гасникова поддержана грантами РФФИ 18-31-20005 мол\_a\_вед и 19-31-51001 Научное наставничество, работа П.Е. Двуреченского поддержана грантом РФФИ 18-29-03071 мк. Работа Е.А. Воронцовой была выполнена при поддержке Минобрнауки РФ (госзадание) № 075-00337-20-03, номер проекта 0714-2020-0005.

требители (пользователи), число которых также может быть очень большим. Цель работы состоит в определении механизма управления ресурсами, которые в контексте данной задачи являются доступными пропускными способностями соединений. При этом необходимо обеспечить стабильную работу системы и предотвратить перегрузки. В качестве критерия оптимальности используется сумма полезностей всех пользователей компьютерной сети.

Первоначально стандартные задачи распределения ресурсов, сводящиеся к максимизации совокупной полезности производителей при совместном использовании имеющихся ресурсов, были рассмотрены в [1], также задача распределения ресурсов в компьютерных сетях исследовалась в недавней работе [2]. Предложенные в монографии [3] механизмы децентрализованного распределения ресурсов с тех пор привлекают большое внимание в экономических исследованиях, см., например, [4]–[6] и ссылки в них. В данной работе, следуя [7], [8], мы рассматриваем различные механизмы корректировки цен. Практическую ценность предложенные подходы имеют за счет своей децентрализованности, что означает, что для установления и корректировки цены в отдельном соединении необходима только реакция пользователей, которые используют это соединение, а не реакция всех пользователей сети. При таком механизме корректировки все соединения действуют независимо.

Кроме того, один из предложенных в статье подходов, основанный на использовании стохастического метода проекции субградиента, обходит следующую проблему, возникающую в реальных сетях. Дело в том, что в реальных сетях пакеты с данными, которые поступают от пользователей на соединение, приходят не одновременно, поэтому на практике суммарный трафик на соединении неизвестен. Для решения этой проблемы предлагается использовать стохастические методы, для работы которых требуется не точное значение суммарного трафика, а только его оценка, которую можно получить на основе трафика одного из пользователей. Идея использовать для решения данной проблемы стохастический метод проекции субградиента предложена в работе [2].

### 1.2. Содержание

Статья организована следующим образом. В разд. 2 описаны постановка задачи и построение двойственной к ней. Также приводятся все необходимые предположения для прямой задачи. В разд. 3 рассматривается решение предложенной задачи быстрым градиентным методом Нестерова [9], получена оценка сложности данного метода, которая имеет порядок  $O\left(\frac{1}{\sqrt{\epsilon}}\right)$ . В разд. 4 описано решение данной задачи с использованием стохастического метода проекции субградиента и приведены оценки сложности, которые имеют порядок  $O\left(\frac{1}{\epsilon^2}\right)$ .

В разд. 5 описано решение задачи с использованием метода эллипсоидов, который хорошо подходит для задач небольшой размерности, также приведен алгоритм построения сертификата точности для данного метода. Приведены оценки сложности, которые имеют порядок  $O\left(m^2 \ln \frac{1}{\epsilon}\right)$ , где  $m$  – число соединений в сети. В разд. 6 описана техника регуляризации, которая позволяет восстанавливать решение прямой задачи по решению двойственной для не прямо двойственного метода. В разд. 7 описано решение регуляризованной задачи с использованием метода экстраполяции случайного градиента. Приведены оценки сложности, которые имеют порядок  $O\left(\frac{1}{\sqrt{\epsilon}} \ln \frac{1}{\epsilon}\right)$ , где логарифмический множитель возникает за счет регуляризации двойственной задачи.

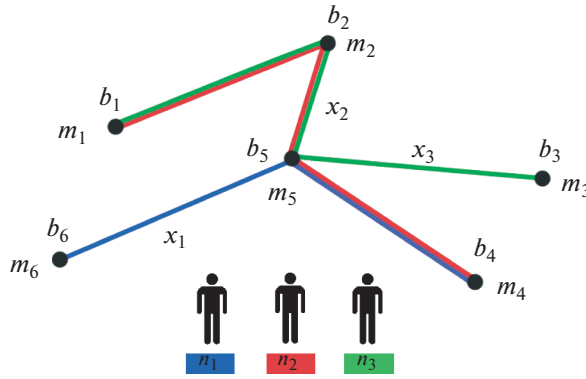
В разд. 8 приведено описание вычислительных экспериментов, подтверждающих на практике теоретические результаты, полученные в предыдущих разделах.

Также для каждого алгоритма описано его распределенное вычисление в контексте рассматриваемой задачи.

## 2. ПОСТАНОВКА ЗАДАЧИ

Рассмотрим компьютерную сеть с  $m$  соединениями и  $n$  пользователями (или узлами), см. фиг. 1.

Пользователи обмениваются пакетами через фиксированный набор соединений. Структура сети задана матрицей маршрутизации  $C = (C_i^j) \in \mathbf{R}^{m \times n}$ . Столбцы матрицы  $C_i \neq 0, i = 1, \dots, n$  – бу-



Фиг. 1. Пример компьютерной сети с  $m = 6$  и  $n = 3$ .

левы  $m$ -мерные векторы такие, что  $C_i^j = 1$  в случае использования узлом  $i$  соединения  $j$ , в противном случае  $C_i^j = 0$ . Ограничения на пропускную способность соединений задаются вектором  $\mathbf{b} \in \mathbb{R}^m$  со строго положительными компонентами.

Пользователи оценивают качество работы сети с помощью функций полезности  $u_k(x_k)$ ,  $k = 1, \dots, n$ , где  $x_k \in \mathbb{R}_+$  – скорость передачи данных  $k$ -го пользователя. За критерий оптимальности системы принята сумма функций полезностей для всех пользователей [1].

Задача максимизации суммарной полезности сети при заданных ограничениях на пропускную способность соединений формулируется следующим образом:

$$\left\{ \begin{array}{l} \max \\ C\mathbf{x} = \sum_{k=1}^n C_k x_k \end{array} \right\} \leq \mathbf{b} \left\{ U(\mathbf{x}) = \sum_{k=1}^n u_k(x_k) \right\}, \tag{1}$$

где  $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}_+^n$ . Решением данной задачи будет оптимальное распределение ресурсов  $\mathbf{x}^*$ .

Рассмотрим стандартный переход к двойственной задаче для (1). Пусть задан вектор двойственных множителей  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}_+^m$ , который можно интерпретировать как вектор цен соединений. Определим двойственную целевую функцию

$$\varphi(\boldsymbol{\lambda}) = \max_{\mathbf{x} \in \mathbb{R}_+^n} \left\{ \sum_{k=1}^n u_k(x_k) + \left\langle \boldsymbol{\lambda}, \mathbf{b} - \sum_{k=1}^n C_k x_k \right\rangle \right\} = \langle \boldsymbol{\lambda}, \mathbf{b} \rangle + \sum_{k=1}^n (u_k(x_k(\boldsymbol{\lambda})) - \langle \boldsymbol{\lambda}, C_k x_k(\boldsymbol{\lambda}) \rangle), \tag{2}$$

причем пользователи выбирают оптимальные скорости передачи информации  $x_k$ , решая следующую задачу оптимизации:

$$x_k(\boldsymbol{\lambda}) = \arg \max_{x_k \in \mathbb{R}_+} \{ u_k(x_k) - x_k \langle \boldsymbol{\lambda}, C_k \rangle \}. \tag{3}$$

Обозначим также за  $\mathbf{x}(\boldsymbol{\lambda})$  вектор с компонентами  $x_k(\boldsymbol{\lambda})$ . Тогда для нахождения оптимальных цен  $\boldsymbol{\lambda}^*$  требуется решить задачу

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}_+^m} \varphi(\boldsymbol{\lambda}). \tag{4}$$

Предположим, что для прямой задачи выполняется условие Слейтера, тогда в силу сильной двойственности и прямая, и двойственная задачи будут иметь решение. Используя условие Слейтера, можно компактифицировать решение двойственной задачи. Будем предполагать, что для решения двойственной задачи верна следующая оценка:

$$\|\boldsymbol{\lambda}^*\|_2 \leq R.$$

При этом значение  $R$  никак не влияет на работу рассматриваемых алгоритмов, а только присутствует в оценках скорости их сходимости.

Основная идея данной работы заключается в применении различных оптимизационных методов для решения двойственной задачи (4) с добавлением прямо двойственного анализа этих методов, позволяющего также восстановить решение прямой задачи (1). В этом смысле мы развиваем подход наших предыдущих работ [10]–[21]. Основное отличие состоит в рассмотрении ограничений–неравенств, а также в анализе стохастических алгоритмов в смысле оценок с большой вероятностью, а не в среднем.

2.1. Сильно вогнутые функции полезности

В некоторых разделах мы будем предполагать, что функции полезности  $u_k(x_k)$ ,  $k = 1, \dots, n$ , являются *сильно вогнутыми* с константой  $\mu$ . В данном подразделе опишем, какими свойствами будет обладать двойственная задача при таком предположении.

**Утверждение 1** (Теорема Демьянова–Данскина–Рубинова, см. [22], [23]). Пусть для любого  $\lambda \in \mathbb{R}_+$  выполняется  $\varphi(\lambda) = \max_{x \in X} F(x, \lambda)$ , где  $F(x, \lambda)$  – выпуклая и гладкая по  $\lambda$  функция и максимум достигается в единственной точке  $x(\lambda)$ . Тогда  $\nabla \varphi(\lambda) = \nabla_\lambda F(x(\lambda), \lambda)$ .

**Утверждение 2** (см. [24]). Пусть для любого  $k = 1, \dots, n$  функции  $u_k(x_k)$  являются  $\mu$ -сильно вогнутыми. Тогда функция (2), где  $x_k(\lambda)$ ,  $k = 1, \dots, n$ , являются решением задачи (3), будет  $nt^2/\mu$ -гладкой, т.е. градиент функции  $\varphi(\lambda)$  будет удовлетворять условию Липшица с константой  $L = nt^2/\mu$ :

$$\|\nabla \varphi(\lambda^2) - \nabla \varphi(\lambda^1)\|_2 \leq L \|\lambda^2 - \lambda^1\|_2.$$

Доказательство утверждения можно найти ниже в Приложении.

2.2. Вогнутые функции полезности

Теперь предположим, что функции полезности  $u_k(x_k)$ ,  $k = 1, \dots, n$ , являются *вогнутыми*, но не *сильно вогнутыми*. Тогда двойственная задача не является гладкой. В данном подразделе описаны некоторые свойства субградиентов двойственной задачи при таких предположениях.

Субградиент двойственной задачи (4) определяется следующим образом:

$$\nabla \varphi(\lambda) = \mathbf{b} - C\mathbf{x}.$$

Учитывая, что  $\mathbf{x}$  – ограниченная скорость передачи информации и вектор  $\mathbf{b}$  также ограничен, получаем, что субградиенты двойственной задачи в данном случае ограничены. Таким образом, существует такая положительная константа  $M$ , что верна следующая оценка:

$$\|\nabla \varphi(\lambda)\|_2 \leq M. \tag{5}$$

В качестве грубой оценки сверху для константы  $M$  из (5) можно использовать  $O(n\sqrt{m})$ . Множитель  $n$  возникает из-за того, что есть  $n$  слагаемых, а  $\sqrt{m}$  используется как оценка зависимости 2-нормы от размерности вектора  $m$ .

3. БЫСТРЫЙ ГРАДИЕНТНЫЙ МЕТОД

В данном разделе мы предполагаем, что функции полезности  $u_k(x_k)$ ,  $k = 1, \dots, n$ , являются *сильно вогнутыми* с константой  $\mu$ , следовательно, двойственная задача будет гладкой.

Для решения двойственной задачи (4) применим быстрый градиентный метод (БГМ) Нестерова в следующем варианте (метод PDFGM, см. алгоритм 1).

**Алгоритм 1.** Прямо двойственный быстрый градиентный метод (PDFGM)

**Вход:**  $u_k(\mathbf{x})$ ,  $k = 1, \dots, n$  – сильно вогнутые функции полезности для каждого пользователя;  $\lambda^0$  – вектор начальных цен,  $\alpha_t := \frac{t+1}{2}$ ,  $A_{-1} := 0$ ,  $A_t := A_{t-1} + \alpha_t = \frac{(t+1)(t+2)}{4}$ ,  $\tau_t := \frac{\alpha_{t+1}}{A_{t+1}} = \frac{2}{t+3}$ ,  $t = 0, 1, \dots, N - 1$ .  
**1:** **for**  $t = 0, 1, \dots, N - 1$

- 2: Вычислить  $\varphi(\lambda^t), \nabla\varphi(\lambda^t)$
  - 3:  $y^t := \left[ \lambda^t - \frac{1}{L} \left( \mathbf{b} - \sum_{k=1}^n C_k x_k(\lambda^t) \right) \right]_+$
  - 4:  $z^t := \left[ \lambda^0 - \frac{1}{L} \sum_{j=0}^t \alpha_j \left( \mathbf{b} - \sum_{k=1}^n C_k x_k(\lambda^j) \right) \right]_+$
  - 5:  $\lambda^{t+1} := \tau_t z^t + (1 - \tau_t) y^t$
  - 6:  $\hat{\mathbf{x}}^{t+1} := \frac{1}{A_{t+1}} \sum_{j=0}^{t+1} \alpha_j \mathbf{x}(\lambda^j)$
  - 7: **end for**
  - 8: **return**  $\lambda^N, \hat{\mathbf{x}}^N$
- 

### 3.1. Распределенный метод

Для решения рассматриваемой задачи можно также применить БГМ в распределенном варианте, что означает, что каждое соединение может вычислять оптимальную скорость передачи данных только исходя из реакции пользователей, которые используют данное соединение, и никак не взаимодействовать с другими соединениями.

Опишем процесс, происходящий на  $t$ -й итерации для соединения  $j$ .

1. На основе информации, полученной от пользователей после предыдущей итерации с номером  $t - 1$  (вектор  $\mathbf{x}^t = \mathbf{x}(\lambda^t)$ ), соединение  $j$  вычисляет

$$y_j^t = \left[ \lambda_j^t - \frac{1}{L} \left( b_j - \sum_{k=1}^n C_k^j x_k^t \right) \right]_+.$$

При этом  $C_k^j \neq 0$  только для тех пользователей, которые используют соединение  $j$ . Поэтому для вычисления этого шага соединению нужна только информация от пользователей, которые используют это соединение.

2. Далее соединение  $j$  аналогично вычисляет

$$z_j^t = \left[ \lambda_j^0 - \frac{\alpha_j}{L} \left( b_j - \sum_{k=1}^n C_k^j x_k^t \right) \right]_+.$$

3. Получив значения на предыдущих двух шагах, соединение  $j$  вычисляет цену для следующей итерации  $t + 1$ :

$$\lambda_j^{t+1} = \tau_t z_j^t + (1 - \tau_t) y_j^t$$

и посылает эту информацию всем пользователям, которые с ним соединены.

4. Далее пользователи вычисляют оптимальные скорости передачи информации  $\hat{\mathbf{x}}^{t+1}$ , в частности, для пользователя  $k$  получаем

$$x_k(\lambda^{t+1}) = \arg \max_{x_k \in \mathbf{R}_+} \left( u_k(x_k) - x_k \sum_{j=1}^m \lambda_j^{t+1} C_k^j \right),$$

где в силу определения матрицы  $C$  пользователю необходима информация только от соединений, которые он использует. Далее пользователь вычисляет оптимальную скорость

$$\hat{x}_k^{t+1} = \frac{A_t \hat{x}_k^t + x_k^{t+1}}{A_{t+1}}.$$

**Замечание 1.** Минусом данного алгоритма является то, что каждому соединению необходимо знать реакции всех пользователей, которые его используют на каждой итерации. К сожалению, в реальных сетях пользователи отправляют данные одновременно, поэтому достаточно сложно собрать данную информацию для соединения. Однако наличие полной информации о пользователях позволяет соединению быстрее устанавливать равновесную цену.

3.2. Оценка скорости сходимости БГМ

Прежде чем привести доказательство сходимости БГМ для рассматриваемой задачи, сформулируем ключевую лемму, необходимую для получения оценок невязки в ограничениях и зазора двойственности после работы прямо двойственного метода PDFGM.

**Лемма 1.** Пусть алгоритм 1 начинает свою работу с начальной точки  $\lambda^0$ , которая лежит в евклидовом шаре радиуса  $R$  с центром в начале координат. Тогда после  $N$  итераций алгоритма 1 будет выполняться следующее неравенство:

$$A_N \varphi(\mathbf{y}^N) - A_N U(\hat{\mathbf{x}}^N) + 2\hat{R}A_N \|(C\hat{\mathbf{x}}^N - \mathbf{b})_+\|_2 \leq \frac{37L\hat{R}^2}{9}, \tag{6}$$

где  $\hat{\mathbf{x}}^N = \frac{1}{A_N} \sum_{t=0}^{N-1} \alpha_t \mathbf{x}(\lambda^t)$  и  $\hat{R} = 3R$ .

Доказательство леммы можно найти в Приложении.

Теперь сформулируем теорему об оценке скорости сходимости алгоритма 1.

**Теорема 1.** Пусть алгоритм 1 начинает свою работу с начальной точки  $\lambda^0$ , которая лежит в евклидовом шаре радиуса  $R$  с центром в начале координат. Тогда после

$$N = \left\lceil \frac{2\hat{R}}{3} \sqrt{\frac{37L}{\varepsilon}} \right\rceil$$

итераций алгоритма 1 будут выполняться следующие неравенства:

$$U(\mathbf{x}^*) - U(\hat{\mathbf{x}}^N) \leq \varepsilon, \quad \|(C\hat{\mathbf{x}}^N - \mathbf{b})_+\|_2 \leq \frac{\varepsilon}{\hat{R}},$$

где  $\hat{\mathbf{x}}^N = \frac{1}{A_N} \sum_{t=0}^{N-1} \alpha_t \mathbf{x}(\lambda^t)$ ,  $\mathbf{x}^*$  – оптимальное решение задачи (1),  $\hat{R} = 3R$ .

**Доказательство.** Обозначим оптимальное значение исходной прямой задачи (1)  $\text{Opt}[P]$ , а оптимальное значение двойственной задачи (4) –  $\text{Opt}[D]$ . В силу слабой двойственности имеем

$$\text{Opt}[D] \geq \text{Opt}[P].$$

Кроме того, для всех  $\mathbf{x} \in \mathbb{R}_+^n$  и оптимального решения двойственной задачи (4)  $\lambda^*$  получаем

$$\text{Opt}[P] \geq U(\mathbf{x}) - \left\langle \lambda^*, \left( \sum_{k=1}^n \mathbf{C}_k x_k - \mathbf{b} \right)_+ \right\rangle \geq U(\mathbf{x}) - \hat{R} \|(C\mathbf{x} - \mathbf{b})_+\|_2. \tag{7}$$

Тогда

$$\begin{aligned} \varphi(\mathbf{y}^N) - U(\hat{\mathbf{x}}^N) &= \varphi(\mathbf{y}^N) - U(\hat{\mathbf{x}}^N) + \text{Opt}[P] - \text{Opt}[P] + \text{Opt}[D] - \text{Opt}[D] = \underbrace{(\text{Opt}[D] - \text{Opt}[P])}_{\geq 0} + \\ &+ (\text{Opt}[P] - U(\hat{\mathbf{x}}^N)) + \underbrace{(\varphi(\mathbf{y}^N) - \text{Opt}[D])}_{\geq 0} \stackrel{(7)}{\geq} -\langle \lambda^*, (\mathbf{b} - C\hat{\mathbf{x}}^N)_+ \rangle \geq -\hat{R} \|(C\hat{\mathbf{x}}^N - \mathbf{b})_+\|_2. \end{aligned}$$

После подстановки последнего неравенства в (6) получим следующую оценку:

$$\hat{R} \|(C\hat{\mathbf{x}}^N - \mathbf{b})_+\|_2 \leq \frac{37L\hat{R}^2}{9A_N}.$$

Следовательно,  $\varphi(\mathbf{y}^N) - U(\hat{\mathbf{x}}^N) \geq -\frac{37L\hat{R}^2}{9A_N}$ . С другой стороны, из (6) следует, что

$\varphi(\mathbf{y}^N) - U(\hat{\mathbf{x}}^N) \leq \frac{37L\hat{R}^2}{9A_N}$ . Поэтому

$$|\varphi(\mathbf{y}^N) - U(\hat{\mathbf{x}}^N)| \leq \frac{37L\hat{R}^2}{9A_N}.$$

Так как  $\varphi(\mathbf{y}^N) \geq \text{Opt}[D] = \varphi(\mathbf{y}^*) \geq \text{Opt}[P] = U(\mathbf{x}^*)$ , выполняется следующее неравенство:

$$U(\mathbf{x}^*) - U(\hat{\mathbf{x}}^N) \leq \frac{37L\hat{R}^2}{9A_N} = \frac{148L\hat{R}^2}{9(N+1)(N+2)} \leq \frac{148L\hat{R}^2}{9N^2} \leq \varepsilon.$$

Выражая из последнего неравенства  $N$ , получаем оценку из условия теоремы.

#### 4. СТОХАСТИЧЕСКИЙ МЕТОД ПРОЕКЦИИ СУБГРАДИЕНТА

Рассмотрим исходную задачу (1), но теперь предположим, что функции полезности  $u_k(x_k)$ ,  $k = 1, \dots, n$ , вогнуты, но не сильно вогнуты. В этом случае двойственная задача (4) становится негладкой. Поэтому для ее решения предлагается применить стохастический метод проекции субградиента. Впервые идея использовать для решения данной проблемы стохастический метод проекции субградиента предложена в работе [2].

Рассмотрим вероятностное пространство  $(\Omega, \mathcal{F}, \mathbb{P})$ . Пусть на нем определена последовательность независимых случайных величин  $\{\xi^t\}_{t=0}^\infty$ , равномерно распределенных на  $\{1, \dots, n\}$ , т.е.

$$\mathbb{P}(\xi^t = i) = \frac{1}{n}, \quad i \in \{1, \dots, n\}.$$

Если доступен оракул, выдающий стохастический субградиент двойственной функции  $\nabla\varphi(\lambda, \xi)$ :

$$\nabla\varphi(\lambda, \xi) = \mathbf{b} - nC_{\xi}x_{\xi}(\lambda),$$

то имеем

$$\mathbb{E}[\mathbf{b} - nC_{\xi^t}x_{\xi^t}(\lambda^t) | \xi^t] = \mathbf{b} - \sum_{k=1}^n C_k x_k(\lambda^t) = \nabla\varphi(\lambda^t)$$

**Алгоритм 2.** Прямо двойственный стохастический метод проекции субградиента (PDSSGM), версия 1

**Вход:**  $u_k(x)$ ,  $k = 1, \dots, n$  – вогнутые функции полезности для каждого пользователя;  $\beta$  – шаг метода.

- 1:  $\lambda^0 := \mathbf{0}$
- 2: **for**  $t = 1, \dots, N - 1$
- 3:   Вычислить  $\nabla\varphi(\lambda^{t-1}, \xi)$
- 4:    $\lambda^t := [\lambda^{t-1} - \beta(\mathbf{b} - nC_{\xi^{t-1}}x_{\xi^{t-1}}(\lambda^{t-1}))]_+$
- 5:    $\hat{\mathbf{x}}^{t+1} := \frac{1}{t+1} \sum_{j=0}^t \mathbf{x}(\lambda^j)$
- 6:    $\hat{\lambda}^{t+1} := \frac{1}{t+1} \sum_{j=0}^t \lambda^j$
- 7: **end for**
- 8: **return**  $\hat{\lambda}^N, \hat{\mathbf{x}}^N$

Следовательно, стохастический субградиент является несмещенной оценкой субградиента.

Оптимальное решение задачи (2) будем искать с помощью прямо двойственного стохастического метода проекции субградиента PDSSGM. Приведем описание двух версий данного метода, см. алгоритм 2 и алгоритм 3. В алгоритме 2 используется полная модель восстановления вектора  $\mathbf{x}(\lambda)$  на каждой итерации. Однако вычисление вектора  $\mathbf{x}(\lambda)$  по сложности практически эквивалентно вычислению полного субградиента  $\varphi(\lambda)$ . Поэтому основным алгоритмом является алгоритм 3, в котором для восстановления вектора  $\mathbf{x}(\lambda)$  используется неполная, стохастическая модель, что означает, что на каждой итерации пересчитывается только одна компонента вектора  $\mathbf{x}(\lambda)$ , остальные остаются без изменений. При доказательстве теоремы сходимости мы сначала

показываем оценку скорости сходимости для алгоритма 2, а затем показываем, что приближенное решение прямой задачи, полученное в алгоритме 3, близко по точности к решению, полученному в алгоритме 2.

**Алгоритм 3.** Прямо двойственный стохастический метод проекции субградиента (PDSSGM), версия 2

**Вход:**  $u_k(\mathbf{x})$ ,  $k = 1, \dots, n$  – вогнутые функции полезности для каждого пользователя;  $\beta$  – шаг метода.

- 1:  $\lambda^0 := \mathbf{0}$
- 2: **for**  $t = 1, \dots, N - 1$
- 3:   Вычислить  $\nabla\varphi(\lambda^{t-1}, \xi)$
- 4:    $\lambda^t := [\lambda^{t-1} - \beta(\mathbf{b} - nC_{\xi^{t-1}x_{\xi^{t-1}}}(\lambda^{t-1}))]_+$
- 5:    $\tilde{\mathbf{x}}_{\xi^t}^{t+1} := \frac{t}{t+1}\tilde{\mathbf{x}}_{\xi^t}^t + \frac{1}{t+1}nx_{\xi^t}(\lambda^t)$ ,  $\tilde{\mathbf{x}}_j^{t+1} := \tilde{\mathbf{x}}_j^t$  при  $j \neq \xi^t$
- 6:    $\hat{\lambda}^{t+1} := \frac{1}{t+1}\sum_{j=0}^t \lambda^j$
- 7: **end for**
- 8: **return**  $\hat{\lambda}^N, \tilde{\mathbf{x}}^N$

#### 4.1. Распределенный метод

Рассмотрим, как можно применить для решения рассматриваемой задачи стохастический метод проекции субградиента в распределенном варианте.

Опишем процесс, происходящий на  $t$ -й итерации для соединения  $j$ :

1. На основе информации, полученной от соединений после предыдущей итерации с номером  $t - 1$ , случайный пользователь  $\xi^t$  передает данные с оптимальной скоростью

$$x_{\xi^t}(\lambda^{t+1}) = \arg \max_{x_{\xi^t} \in \mathbf{R}_+} \left( u_{\xi^t}(x_{\xi^t}) - x_{\xi^t} \sum_{j=1}^m \lambda_j^{t+1} C_{\xi^t}^j \right),$$

где в силу определения матрицы  $C$  пользователю необходима информация только от соединений, которые он использует.

2. Далее соединение  $j$  вычисляет цену на следующую итерацию на основе реакции этого пользователя:

$$\lambda_j^{t+1} = [\lambda_j^t - \beta(b_j - nC_{\xi^t x_{\xi^t}}^j)]_+.$$

При этом  $C_{\xi^t}^j \neq 0$  только для тех пользователей, которые используют соединение  $j$ . Поэтому цена поменяется только для актуальных соединений пользователя, который передал данные.

**Замечание 2.** Главное преимущество данного метода заключается в том, что соединение меняет цену только на основе реакции одного пользователя, что гораздо больше приближает постановку задачи к ситуации в реальных сетях, в которых пользователи отправляют данные не одновременно.

#### 4.2. Оценка скорости сходимости для стохастического метода проекции субградиента

Прежде чем перейти к доказательству основной теоремы об оценках скорости сходимости предложенных методов, приведем необходимые предположения для решаемой задачи. Предположим, что существует положительная константа  $M = O(n\sqrt{m})$  такая, что верна следующая оценка:

$$\|\nabla\varphi(\lambda, \xi)\|_2 \leq M. \tag{8}$$



Данное предположение корректно в силу того, что скорость передачи информации  $\mathbf{x}$  ограничена и вектор пропускных способностей  $\mathbf{b}$  также ограничен в силу физических соображений. Поэтому, исходя из определения стохастического субградиента, он ограничен.

Также предположим, что

$$\mathbb{E} \left[ \exp \left( \frac{\|\nabla \varphi(\lambda, \xi) - \nabla \varphi(\lambda)\|_2^2}{\sigma^2} \right) \right] \leq \exp(1),$$

где  $\sigma$  – положительная числовая константа, порядок зависимости от  $n$  и  $m$  такой же, как у  $M$ .

Для получения оценки скорости сходимости алгоритма 9 необходимо предположение о том, что функции  $u_k(x_k)$ ,  $k = 1, \dots, n$ , являются липшицевыми с константой  $M_{u_k}$ , тогда функция  $U(\mathbf{x})$  будет липшицевой с некоторой константой  $M_U$ :

$$\forall \mathbf{x}, \mathbf{y} \quad |U(\mathbf{x}) - U(\mathbf{y})| \leq M_U \|\mathbf{x} - \mathbf{y}\|_2,$$

при этом  $M_U = O(\sqrt{n})$ . Может оказаться, что функция  $u_k(x_k)$  липшицева везде, кроме, например, точки 0. Примером такой функции является одна из наиболее распространенных функций полезности пользователей  $u_k(x_k) = \ln x_k$ . Но в силу специфики рассматриваемой задачи всегда существуют  $\bar{\varepsilon} > 0$ ,  $\underline{\varepsilon} > 0$  такие, что  $x_k^* \geq \underline{\varepsilon}$ ,  $x_k^* \leq \bar{\varepsilon}$ . Тогда задачу можно решать на компакте  $Q = \{\mathbf{x} : \underline{\varepsilon} \leq x_k \leq \bar{\varepsilon}, k = 1, \dots, n\}$ , и рассматриваемая функция  $u_k(x_k) = \ln x_k$  становится липшицевой на  $Q$ . В общем случае вогнутая функция полезности  $u(x)$  будет липшицевой на компакте, лежащем в относительной внутренней области области определения  $u(x)$ .

Пусть

$$\mathbb{E} \left[ \exp \left( \frac{\|\mathbf{x}(\lambda, \xi) - \mathbf{x}(\lambda)\|_2^2}{\sigma_x^2} \right) \right] \leq \exp(1),$$

где  $\sigma_x = O(\sqrt{n})$  – положительная числовая константа и

$$\mathbf{x}(\lambda, \xi) = (0, \dots, nx_\xi(\lambda), \dots, 0)^T.$$

Сформулируем ключевую лемму, необходимую для получения оценок скорости сходимости невязки в ограничениях и зазора двойственности после работы прямо двойственного метода PDSSGM.

**Лемма 2.** Пусть алгоритм 3 начинает свою работу с начальной точки  $\lambda^0 = 0$  и с шагом  $\beta$ . Тогда после  $N$  итераций алгоритма 3 с вероятностью  $1 - 4\delta$  выполняется неравенство:

$$\begin{aligned} \varphi(\hat{\lambda}^N) - U(\tilde{\mathbf{x}}^N) + 2R \|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+ \leq C_1 \frac{R^2 \sigma \sqrt{g(N)J}}{\sqrt{N}} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2} + \\ + \frac{\sqrt{2} \left(1 + \sqrt{3 \ln \frac{1}{\delta}}\right)}{\sqrt{N}} \left( M_U \sigma_x + 2R \left( \sigma + \sigma_x \sqrt{\lambda_{\max}(C^T C)} \right) \right), \end{aligned}$$

где

$$\hat{\lambda}^N = \frac{1}{N} \sum_{t=0}^{N-1} \lambda^t$$

и

$$\tilde{\mathbf{x}}^N = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(\lambda^t, \xi^t),$$

$C_1$  – положительная числовая константа,  $g(N) = \ln \left( \frac{N}{\delta} \right) + \ln \ln \left( \frac{F}{f} \right)$ ,

$$F = 2\sigma^2 N (2\beta)^N \left( 2R^2 + 2\beta^2 M^2 + \beta R^2 + 24 \ln \frac{N}{\delta} \beta \sigma^2 N \right),$$

$$f = \sigma^2 R^2 u$$

$$J = \max \left\{ 1, \frac{1}{R} \beta C_1 \sqrt{\sigma^2 g(N)} + \sqrt{\frac{1}{R^2} \beta^2 C_1^2 \sigma^2 g(N) + \frac{2R^2 + 2\beta^2 M^2}{R^2}} \right\},$$

а  $R$  определяется условием  $\|\lambda^*\|_2 \leq R$ .

Доказательство леммы приведено в Приложении.

Теперь сформулируем теорему об оценке скорости сходимости алгоритма 3.

**Теорема 2.** Пусть алгоритм 3 начинает свою работу с начальной точки  $\lambda^0 = 0$  и с шагом  $\beta = \frac{R}{M\sqrt{N}}$ . Обозначим

$$A = \sqrt{2} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right) \left( M_U \sigma_x + 2R \left( \sigma + \sigma_x \sqrt{\lambda_{\max}(C^T C)} \right) \right) + 2.5RM.$$

Тогда после

$$N = O \left( \left[ \frac{A^2}{\varepsilon^2} \ln \left( \frac{MR}{\varepsilon \delta} \right) \right] \right)$$

итераций алгоритма 3 с вероятностью  $1 - 4\delta$  будут выполняться следующие неравенства:

$$U(\mathbf{x}^*) - U(\tilde{\mathbf{x}}^N) \leq \varepsilon, \quad \|(C\tilde{\mathbf{x}}^N - \mathbf{b})_+\|_2 \leq \frac{\varepsilon}{R},$$

где  $\tilde{\mathbf{x}}^N = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(\lambda^t, \xi^t)$ ,  $\mathbf{x}^*$  – оптимальное решение задачи (1).

**Доказательство.** Начало доказательства такое же, как в теореме 1, но с использованием оценки из леммы 2. В результате для шага  $\beta = \frac{R}{M\sqrt{N}}$  получим, что

$$\frac{\sqrt{2} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right) \left( M_U \sigma_x + 2R \left( \sigma + \sigma_x \sqrt{\lambda_{\max}(C^T C)} \right) \right) + \frac{5RM}{2\sqrt{N}} + C_1 \frac{R^2 \sigma \sqrt{g(N)J}}{\sqrt{N}}}{\sqrt{N}},$$

при этом с точностью до констант  $g(N) \approx \ln \left( \frac{N}{\delta} \right)$ ,  $J \approx \max \{ 1, \beta \sqrt{g(N)} \}$ . Далее найдем  $N$ , при котором оценка становится меньше  $\varepsilon$ .

Введем следующие обозначения:

$$A = \sqrt{2} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right) \left( M_U \sigma_x + 2R \left( \sigma + \sigma_x \sqrt{\lambda_{\max}(C^T C)} \right) \right) + 2.5RM,$$

$$B = C_1 R^2 \sigma.$$

Необходимо получить минимальную оценку на число итераций  $N$  для достижения заданной точности  $\varepsilon$ . Для  $J = 1$  получаем, что

$$\sqrt{N} = \left\lceil \frac{A + B \sqrt{\ln \left( \frac{N}{\delta} \right)}}{\varepsilon} \right\rceil. \tag{9}$$

Подставляя рекурсивно значение  $N$ , из (9) получим следующую оценку сложности:

$$N = O \left( \left[ \frac{A^2}{\varepsilon^2} \ln \left( \frac{MR}{\varepsilon \delta} \right) \right] \right).$$

Для  $J = \beta\sqrt{g(N)} = \frac{R\sqrt{g(N)}}{M\sqrt{N}}$  потребуем, чтобы

$$\frac{A}{\sqrt{N}} + \frac{Bg(N)R}{MN} = \frac{A}{\sqrt{N}} + \frac{\bar{B}g(N)}{N} \leq \varepsilon.$$

Так как нам нужно найти минимальное  $N$ , то, заменив последнее неравенство на равенство и решая полученное уравнение, получаем, что

$$\sqrt{N} = \left\lceil \frac{A + \sqrt{A^2 + 4\varepsilon\bar{B} \ln\left(\frac{N}{\delta}\right)}}{2\varepsilon} \right\rceil.$$

Аналогично случаю  $J = 1$  из последнего равенства получается следующая оценка:

$$N = O\left(\left\lceil \frac{A^2}{\varepsilon} \ln\left(\frac{MR}{\varepsilon\delta}\right) \right\rceil\right).$$

Худшая из оценок сложности для  $J = 1$  и для  $J = \beta\sqrt{g(N)}$  и будет оценкой из условия теоремы.

## 5. МЕТОД ЭЛЛИПСОИДОВ

В этом разделе для решения исходной задачи (1) предлагается применить метод эллипсоидов [25]. Данный метод можно использовать при небольшой размерности ( $m$ ) двойственной задачи или в случае, когда требуется высокая точность решения. Данный метод является прямо двойственным, т.е. по решению двойственной задачи можно восстановить решение прямой задачи.

Рассмотрим исходную задачу (1) и двойственную к ней (2). Как и в предыдущем разделе, считаем, что функции  $u_k(x_k)$ ,  $k = 1, \dots, n$ , являются вогнутыми, но не сильно вогнутыми. Также предположим, что решение двойственной задачи лежит в евклидовом шаре радиуса  $R$  с центром в начале координат, т.е.  $\|\lambda^*\|_2 \leq R$ . В качестве начальной точки метода возьмем нулевой вектор:  $\lambda^0 = 0$ . Задачу будем решать на множестве:

$$\Lambda_{2R} = \{\lambda \in \mathbb{R}_+^m : \|\lambda\|_2 \leq 2R\}.$$

Приведем описание метода эллипсоидов (алгоритм 4), который будем применять для решения двойственной задачи.

---

### Алгоритм 4. Метод эллипсоидов

---

**Вход:**  $u_k(x_k)$ ,  $k = 1, \dots, n$  – вогнутые функции полезностей

1:  $B_0 := 2R \cdot I_n$ ,  $I_n$  – единичная матрица

2: **for**  $t = 0, \dots, N - 1$

3:   Вычислить  $\nabla\varphi(\lambda^t)$

4:    $\mathbf{q}_t := B_t^T \nabla\varphi(\lambda^t)$

5:    $\mathbf{p}_t := \frac{B_t^T \mathbf{q}_t}{\sqrt{\mathbf{q}_t^T B_t B_t^T \mathbf{q}_t}}$

6:    $B_{t+1} := \frac{m}{\sqrt{m^2 - 1}} B_t + \left(\frac{m}{m+1} - \frac{m}{\sqrt{m^2 - 1}}\right) B_t \mathbf{p}_t \mathbf{p}_t^T$

7:    $\lambda^{t+1} := \lambda^t - \frac{1}{m+1} B_t \mathbf{p}_t$

8: **end for**

9: **return**  $\lambda^N$

---

Для восстановления решения прямой задачи по решению двойственной необходимо определить сертификат точности  $\xi$  для метода эллипсоидов. Напомним, что сертификатом точности называется последовательность  $\xi = \{\xi_t\}_{t=0}^{N-1}$  весов таких, что

$$\xi_t \geq 0, \quad \sum_{t=0}^{N-1} \xi_t = 1.$$

Построение сертификата точности в нашем случае будет осуществляться в процессе работы метода эллипсоидов, см. алгоритм 12, общая схема которого такова [26]:

1. Находим “наиболее узкую полоску”, содержащую эллипсоид  $Q_N$ , остающийся после итерации  $N$ , т.е. такой вектор  $\mathbf{h}$ , что на  $Q_N$  выполняется неравенство:

$$\max_{\lambda \in Q_N} \langle \mathbf{h}, \lambda \rangle - \min_{\lambda \in Q_N} \langle \mathbf{h}, \lambda \rangle \leq 1. \tag{10}$$

Для метода эллипсоидов все  $Q_N$  представлены в виде

$$Q_N = \{B_N \mathbf{z} + \lambda^N : \mathbf{z}^T \mathbf{z} \leq 1\}.$$

Тогда для решения (10) необходимо сделать сингулярное разложение матрицы  $B_N = UDV$ , где  $U$  и  $V$  – ортогональные матрицы и  $D$  – диагональная матрица с положительными элементами на диагонали. Далее искомый вектор  $\mathbf{h}$  определяется следующим образом:  $\mathbf{h} = 1/(2\sigma^{i_*}) \cdot U\mathbf{e}^{i_*}$ , где  $i_*$  – индекс наименьшего диагонального элемента матрицы  $D$ ,  $\sigma^{i_*}$  – соответствующее этому индексу значение элемента матрицы  $D$ ,  $\mathbf{e}^i$  – векторы стандартного базиса.

2. Для векторов  $\mathbf{h}^+ = [\mathbf{h}, -\langle \mathbf{h}, \lambda^N \rangle]$  и  $\mathbf{h}^- = -\mathbf{h}^+$  находим разложения следующего вида:

$$\mathbf{h}^+ = \sum_{t=0}^{N-1} \nu_t [\nabla \varphi(\lambda^t), -\langle \nabla \varphi(\lambda^t), \lambda^t \rangle] + \mathbf{y}^+,$$

$$\mathbf{h}^- = \sum_{t=0}^{N-1} \mu_t [\nabla \varphi(\lambda^t), -\langle \nabla \varphi(\lambda^t), \lambda^t \rangle] + \mathbf{z}^+,$$

существование которых следует из утверждения 4.1 [26]. Этот шаг описывают п. 6–13 рассмотренного ниже алгоритма 5.

3. Из коэффициентов разложения  $\nu_t$  и  $\mu_t$  векторов  $\mathbf{h}^+$  и  $\mathbf{h}^-$ , соответственно, получаем выражения для  $\xi_t$ ,  $t \in I_N$ , где

$$I_N = \{t \leq N - 1 : \lambda^t \in \text{int } \Lambda_{2R}\}.$$

Коэффициенты разложения определяются только для тех точек, получаемых в процессе работы алгоритма 5, которые принадлежат допустимому множеству.

**Алгоритм 5.** Построение сертификата точности для метода эллипсоидов

**Вход:**  $N - 1$  – номер итерации, на которой производится вычисление сертификата точности,

$\{B_t, \lambda^t, \nabla \varphi(\lambda^t)\}_{t=0}^{N-1}$  – протокол работы метода эллипсоидов после  $N$  итераций

- 1: **если**  $\nabla \varphi(\lambda^{N-1}) = 0$ , **то**
- 2:     $\xi_t := 0$  для всех  $t = 0, \dots, N - 2$
- 3:     $\xi_{N-1} := 1$
- 4: **иначе**
- 5:     $\mathbf{h} := 1/(2\sigma^{i_*}) \cdot U\mathbf{e}^{i_*}$
- 6:     $\mathbf{g}_\nu := \mathbf{h}, \mathbf{g}_\mu := -\mathbf{h}$
- 7:    **for**  $t = 0, \dots, N - 1$

```

8:    $\mathbf{q} := B_t^T \nabla \varphi(\lambda^t)$ 
9:    $v_t := [\mathbf{g}_v^T B_t \mathbf{q}]_+ / \|\mathbf{q}\|_2^2$ 
10:   $\mathbf{g}_v := \mathbf{g}_v - v_t \nabla \varphi(\lambda^t)$ 
11:   $\mu_t := [\mathbf{g}_\mu^T B_t \mathbf{q}]_+ / \|\mathbf{q}\|_2^2$ 
12:   $\mathbf{g}_\mu := \mathbf{g}_\mu - \mu_t \nabla \varphi(\lambda^t)$ 
13:  end for
14:   $\xi_t := (v_t + \mu_t) / \sum_{i \in I_N} (v_i + \mu_i), t \in I_N$ 
15: end если
16: return  $\{\xi_t\}_{t=0}^{N-1}$ 

```

---

**Замечание 3.** Отметим, что, в отличие от БГМ и стохастического метода проекции субградиента, в методе эллипсоидов для вычисления шагов 4–6 алгоритма 4 необходима информация о всех компонентах градиента, т.е. нужна информация от всех пользователей. Поэтому необходим общий центр для всех соединений, который будет собирать информацию со всех соединений и выполнять эти вычисления.

Сформулируем теорему об оценке скорости сходимости метода эллипсоидов для решаемой задачи.

**Теорема 3** (см. [26]). Пусть алгоритм 4 начинает свою работу с начальным шаром  $B_0 = \{\lambda \in \mathbb{R}^m : \|\lambda\| \leq 2R\}$  и сертификат точности  $\xi$  формируется в соответствии с алгоритмом 5. Тогда после

$$N = 2m(m+1) \left\lceil \ln \left( \frac{32 \cdot 4MR}{\varepsilon} \right) \right\rceil \quad (11)$$

итераций будут выполняться следующие неравенства:

$$U(\mathbf{x}^*) - U(\hat{\mathbf{x}}^N) \leq \varepsilon, \quad \|\mathbf{C}\hat{\mathbf{x}}^N - \mathbf{b}\|_+ \leq \frac{\varepsilon}{R},$$

где

$$\hat{\mathbf{x}}^N = \sum_{t \in I_N} \xi_t \mathbf{x}(\lambda^t), \quad I_N = \{t \leq N-1 : \lambda^t \in \text{int } \Lambda_{2R}\}.$$

Доказательство теоремы можно найти ниже в Приложении.

## 6. РЕГУЛЯРИЗАЦИЯ ДВОЙСТВЕННОЙ ЗАДАЧИ

В предыдущих разделах мы рассматривали прямо двойственные методы для решения двойственной задачи. Однако существует стандартный подход, который позволяет по решению двойственной задачи не прямо двойственным методом восстанавливать решение прямой задачи. Ключевой идеей данного подхода является регуляризация двойственной задачи, которая гарантирует сильную выпуклость регуляризованной задачи. Далее опишем подробно предлагаемый подход и сформулируем леммы, связывающие решения прямой и двойственной задачи.

Регуляризуем функционал (2) по Тихонову

$$\varphi_\delta(\lambda) = \varphi(\lambda) + \frac{\delta}{2} \|\lambda\|_2^2$$

и вместо задачи (4) будем решать регуляризованную задачу

$$\min_{\lambda \in \mathbb{R}_+^m} \varphi_\delta(\lambda).$$

Параметр  $\delta$  будет оптимально подобран позже. Как и в предыдущем разделе, предполагается, что задача решается на множестве

$$\Lambda_{2R} = \{\lambda \in \mathbb{R}_+^m : \|\lambda\|_2 \leq 2R\}.$$

Для полученной регуляризованной функции сформулируем следующую лемму о гладкости регуляризованной задачи.

**Лемма 3.** Пусть функция  $\varphi(\lambda)$  –  $L$ -гладкая, тогда регуляризованная функция  $\varphi_\delta(\lambda)$  является  $(L + \delta)$ -гладкой, т.е. для любых  $\lambda^1, \lambda^2 \in \mathbf{R}_+^m$

$$\|\nabla\varphi_\delta(\lambda^1) - \nabla\varphi_\delta(\lambda^2)\|_2 \leq (L + \delta)\|\lambda^1 - \lambda^2\|_2. \tag{12}$$

**Доказательство.** Градиент регуляризованной функции

$$\nabla\varphi_\delta(\lambda) = \nabla\varphi(\lambda) + \delta\lambda.$$

Следовательно, имеем оценку

$$\|\nabla\varphi_\delta(\lambda^1) - \nabla\varphi_\delta(\lambda^2)\|_2 = \|\nabla\varphi(\lambda^1) - \nabla\varphi(\lambda^2) + \delta(\lambda^1 - \lambda^2)\|_2 \leq \|\nabla\varphi(\lambda^1) - \nabla\varphi(\lambda^2)\|_2 + \delta\|\lambda^1 - \lambda^2\|_2.$$

Откуда в силу утверждения 2 следует (12).

Также для оценки сходимости алгоритма для прямой задачи нам понадобится следующая вспомогательная лемма о связи между оценкой на градиент для двойственной задачи и оценками на сходимость по функции и невязку в ограничении для прямой задачи.

**Лемма 4** (см. [10]). Пусть  $\mathbf{x}^*$  – решение прямой задачи (1). Тогда выполняются следующие неравенства:

$$\|C\mathbf{x}(\lambda) - \mathbf{b}\|_2 \leq \|\nabla\varphi_\delta(\lambda)\|_2 + \delta\|\lambda\|_2, \tag{13}$$

$$U(\mathbf{x}^*) - U(\mathbf{x}(\lambda)) \leq \|\nabla\varphi_\delta(\lambda)\|_2 \cdot \|\lambda\|_2 + \delta\|\lambda\|_2^2, \tag{14}$$

где  $\mathbf{x}(\lambda)$  определяется в (3).

**Доказательство.** В силу (3) имеем

$$U(\mathbf{x}(\lambda)) + \langle \lambda, \mathbf{b} - C\mathbf{x}(\lambda) \rangle \geq U(\mathbf{x}^*) + \langle \lambda, \mathbf{b} - C\mathbf{x}^* \rangle \geq U(\mathbf{x}^*).$$

Откуда

$$U(\mathbf{x}(\lambda)) \geq U(\mathbf{x}^*) - \langle \lambda, \mathbf{b} - C\mathbf{x}(\lambda) \rangle = U(\mathbf{x}^*) - \langle \lambda, \nabla\varphi(\lambda) \rangle.$$

Так как  $\varphi(\lambda) = \varphi_\delta(\lambda) - \frac{\delta}{2}\|\lambda\|_2^2$ , имеем

$$\|\nabla\varphi(\lambda)\|_2 = \|\nabla\varphi_\delta(\lambda) - \delta\lambda\|_2 \leq \|\nabla\varphi_\delta(\lambda)\|_2 + \delta\|\lambda\|_2.$$

Из последнего неравенства, учитывая соотношение  $\nabla\varphi(\lambda) = \mathbf{b} - C\mathbf{x}(\lambda)$ , получаем (13).

Далее, оценка (14) следует из

$$\begin{aligned} U(\mathbf{x}^*) - U(\mathbf{x}(\lambda)) &\leq \langle \lambda, \nabla\varphi(\lambda) \rangle \leq \|\nabla\varphi(\lambda)\|_2 \cdot \|\lambda\|_2 \leq \\ &\leq \|\lambda\|_2 \cdot (\|\nabla\varphi_\delta(\lambda)\|_2 + \delta\|\lambda\|_2) \leq \|\nabla\varphi_\delta(\lambda)\|_2 \cdot \|\lambda\|_2 + \delta\|\lambda\|_2^2. \end{aligned}$$

Кроме этого, понадобится лемма о сходимости по градиенту для регуляризованной функции.

**Лемма 5.** Пусть  $\lambda_\delta^*$  – решение регуляризованной двойственной задачи. Имеет место следующее неравенство:

$$\|\nabla\varphi_\delta(\lambda^N)\|_2 \leq (L + \delta)\|\lambda^N - \lambda_\delta^*\|_2.$$

**Доказательство** немедленно следует из леммы 3 и равенства

$$\nabla\varphi_\delta(\lambda_\delta^*) = 0.$$

Мы сформулировали необходимые леммы для регуляризованной задачи и в следующем разделе рассмотрим на примере применение этого подхода.

## 7. МЕТОД ЭКСТРАПОЛЯЦИИ СЛУЧАЙНОГО ГРАДИЕНТА

Рассмотрим метод экстраполяции случайного градиента [27]. Отметим, что данный метод не требует пересчета градиента на каждой итерации, необходимо пересчитывать только одну его компоненту на каждой итерации, что значительно сокращает вычисления, особенно для задач

большой размерности. Так как данный метод не является прямо двойственным, то необходимо применять алгоритм 6 к регуляризованной задаче.

Параметры  $\alpha$ ,  $\eta$ ,  $\tau$  и  $\theta_t$  выбираются следующим образом:

$$\bar{\alpha} = 1 - \frac{1}{n + \sqrt{n^2 + 16nL/\delta}}, \quad (15)$$

$$\alpha = n\bar{\alpha}, \quad \eta = \frac{\delta\bar{\alpha}}{1 - \bar{\alpha}}, \quad \tau = \frac{1}{n(1 - \bar{\alpha})} - 1, \quad \theta_t = \bar{\alpha}^{-t}. \quad (16)$$

### 7.1. Распределенный метод

В данной секции опишем распределенную версию рассмотренного метода. Для начала отметим, что векторы  $\underline{\lambda}_1^0, \dots, \underline{\lambda}_n^0$  хранятся у соответствующих пользователей и влияют на формирование оптимального значения трафика для соответствующего пользователя. Как было отмечено при описании распределенного БГМ, на определение оптимального трафика пользователя влияют только цены соединений, через которые данный пользователь обменивается пакетами. Поэтому можно считать, что в векторе  $\underline{\lambda}_k^t$  ненулевые только те компоненты, номера которых совпадают с номерами используемых соединений.

---

#### Алгоритм 6. Метод экстраполяции случайного градиента (RGEM)

---

**Вход:** Параметры  $\alpha$ ,  $\eta$ ,  $\tau$ ,  $\{\theta_t\}_{t=1}^N$

1:  $\lambda^0 := \mathbf{0}$

2:  $\lambda_i^0 := \lambda^0, i = 1, \dots, n$

3:  $y_{-1} = y_0 = \mathbf{0}$

4: **for**  $t = 1, \dots, N$

5:   Выбрать случайным образом  $k_t$  из множества  $\{1, \dots, n\}$  равномерно по всем значениям

6:    $\tilde{y}_k^t := \mathbf{y}_k^{t-1} + \alpha(\mathbf{y}_k^{t-1} - \mathbf{y}_k^{t-2}), k = 1, \dots, n$

7:    $\lambda^t := \left[ \eta \lambda^{t-1} - \frac{1}{n} \sum_{k=1}^n \tilde{y}_k^t \right]_+ / (\delta + \eta)$

8:

9:    $\lambda_{k_t}^t := (\lambda^t + \tau \lambda_{k_t}^{t-1}) / (1 + \tau)$

10:    $\lambda_k^t := \lambda_k^{t-1}, k \in \{1, \dots, n\} \setminus \{k_t\}$

11:

12:    $\mathbf{y}_{k_t}^t := \mathbf{b} - nC_{k_t} x_{k_t}(\lambda_{k_t}^t)$

13:    $\mathbf{y}_k^t := \mathbf{y}_k^{t-1}, k \in \{1, \dots, n\} \setminus \{k_t\}$

14: **end for**

15:    $\bar{\lambda}^N := \left( \sum_{t=0}^{N-1} \theta_t \lambda^t \right) / \sum_{t=1}^N \theta_t$

16: **return**  $\bar{\lambda}^N$

---

Опишем распределенный алгоритм на  $t$ -й итерации.

1. Используя информацию, собранную с пользователей на предыдущей итерации, соединение  $j$  вычисляет

$$\tilde{y}_{k,j}^t := y_{k,j}^{t-1} + \alpha(y_{k,j}^{t-1} - y_{k,j}^{t-2}) = b_j - nC_k^j x_k(\lambda_{k_t}^{t-1}) + \alpha(nC_k^j x_k(\lambda_{k_t}^{t-2}) - nC_k^j x_k(\lambda_{k_t}^{t-1})).$$

Заметим, что в силу определения матрицы  $C$  соединению  $j$  нужна информация только от пользователей, которые обмениваются пакетами через это соединение.

2. Далее соединение  $j$  меняет цену по следующему правилу:

$$\lambda_j^t = \left[ \eta \lambda_j^{t-1} - \frac{1}{n} \sum_{k=1}^n \tilde{y}_{k,j}^t \right]_+ / (\delta + \eta).$$

3. Один из пользователей  $k_t$  реагирует на изменение цен и в качестве локального вектора цен сохраняет

$$\underline{\lambda}_{k_t}^t = (\lambda^t + \tau \underline{\lambda}_{k_t}^{t-1}) / (1 + \tau),$$

остальные пользователи никак не изменяют локальные цены, т.е.  $\underline{\lambda}_{k_t}^t = \underline{\lambda}_{k_t}^{t-1}$ .

4. Пользователь  $k_t$  вычисляет

$$x_{k_t}^t(\underline{\lambda}_{k_t,j}^t) = \arg \max_{x_k \in \mathbf{R}_+} \left( u_k(x_k) - x_k \sum_{j=1}^m \lambda_{k_t,j}^t C_k^j \right)$$

и отправляет эту информацию соединениям, которые использует.

5. Соединение  $j$  обновляет у себя информацию для  $k_t$  пользователя

$$y_{k_t,j}^t = b_j - n C_{k_t}^j x_{k_t}(\underline{\lambda}_{k_t}^t).$$

Данную информацию соединение будет обновлять только в случае, когда через него обмениваются пакетами пользователь  $k_t$ .

### 7.2. Оценка скорости сходимости метода экстраполяции случайного градиента

В данном разделе мы, так же, как в разд. 3, рассматриваем задачу (2) с  $\mu$ -сильно вогнутыми функциями затрат  $u_k(x_k)$ ,  $k = 1, \dots, n$ . Напомним, что в силу сильной вогнутости функций затрат двойственная задача (4) будет гладкой с константой Липшица  $L = \frac{nm^2}{\mu}$ .

Для оценки скорости сходимости метода необходима следующая оценка для невязки по аргументу из теоремы 2.1 [27]:

$$\mathbb{E} \left[ \left\| \lambda_\delta^* - \lambda^N \right\|_2^2 \right] \leq \frac{4\Delta(\bar{\alpha})^N}{\delta}, \tag{17}$$

где  $\Delta = \delta \left\| \lambda_\delta^* - \lambda^0 \right\|_2^2 + \frac{B}{n\delta} + \varphi_\delta(\lambda^0) - \varphi_\delta(\lambda_\delta^*)$ ,  $B = \|\mathbf{b}\|_2^2$ .

Используя (17), можно доказать теорему об оценке скорости сходимости метода непосредственно для задачи (11).

**Теорема 4.** Пусть для решения регуляризованной двойственной задачи (11) применяется метод *RGEM* с параметрами (15), (16) и  $\delta = \frac{\varepsilon}{8R^2}$  для

$$N = \left\lceil 2 \left( n + \sqrt{n^2 + \frac{128nLR^2}{\varepsilon}} \right) \ln \left( \frac{4RA}{\varepsilon} \right) \right\rceil$$

итераций, где  $A = 2 \left( LR + \frac{\varepsilon}{8R} \right) \sqrt{6 + \frac{16LR^2n + 8B}{n\varepsilon}}$ . Тогда

$$\mathbb{E}[U(\mathbf{x}^*) - U(\mathbf{x}(\lambda^N))] \leq \varepsilon, \quad \mathbb{E} \left[ \left\| C\mathbf{x}(\lambda^N) - \mathbf{b} \right\|_2 \right] \leq \frac{\varepsilon}{2R}.$$

**Доказательство.** Из леммы 4 имеем оценки для невязки по ограничениям (13) и по целевой функции (14). Ввиду предположения  $\lambda \in \Lambda_{2R}$ , выполняются следующие неравенства:

$$\left\| C\mathbf{x}^N - \mathbf{b} \right\|_2 \leq \left\| \nabla \varphi_\delta(\lambda^N) \right\|_2 + 2\delta R, \tag{18}$$



$$U(\mathbf{x}^*) - U(\mathbf{x}^N) \leq 2R \|\nabla \varphi_\delta(\lambda^N)\|_2 + 4\delta R^2, \quad (19)$$

где  $\mathbf{x}^N = \mathbf{x}(\lambda^N)$ . Из леммы 5 и неравенства (17) получаем следующую оценку для  $\|\nabla \varphi_\delta(\lambda^N)\|_2$ :

$$\mathbb{E} \left[ \|\nabla \varphi_\delta(\lambda^N)\|_2 \right] \leq 2(L + \delta) \sqrt{\frac{\Delta}{\delta}} (\bar{\alpha})^{N/2}.$$

Оценим  $\Delta$ . Для функции  $\varphi_\delta$  с липшицевым градиентом выполняется неравенство

$$\varphi_\delta(\lambda^0) - \varphi_\delta(\lambda_\delta^*) \leq \langle \nabla \varphi_\delta(\lambda_\delta^*), \lambda^0 - \lambda^* \rangle + \frac{L + \delta}{2} \|\lambda_\delta^* - \lambda^0\|_2^2.$$

Учитывая, что  $\nabla \varphi_\delta(\lambda_\delta^*) = 0$ , получаем

$$\Delta \leq \delta \|\lambda_\delta^* - \lambda^0\|_2^2 + \frac{B}{n\delta} + \frac{L + \delta}{2} \|\lambda_\delta^* - \lambda^0\|_2^2 \leq (6\delta + 2L)R^2 + \frac{B}{n\delta}.$$

Пусть  $\delta$  подбирается так, чтобы выполнялось  $4\delta R^2 = \frac{\varepsilon}{2}$ , тогда  $\delta = \frac{\varepsilon}{8R^2}$ . Отсюда имеем

$$4\delta R^2 = \frac{\varepsilon}{2}, \quad 2\delta R = \frac{\varepsilon}{4R}.$$

Потребуем выполнения неравенства  $U(\mathbf{x}^*) - U(\mathbf{x}(\lambda^N)) \leq \varepsilon$ . Тогда, в силу (18), (19) должно выполняться неравенство:

$$\|\nabla \varphi_\delta(\lambda^N)\|_2 \leq \frac{\varepsilon}{4R}.$$

Откуда имеем

$$2(L + \delta) \sqrt{\frac{\Delta}{\delta}} (\bar{\alpha})^{N/2} \leq \frac{\varepsilon}{4R}.$$

Учитывая, что

$$2(L + \delta) \sqrt{\frac{\Delta}{\delta}} \leq 2 \left( LR + \frac{\varepsilon}{8R} \right) \sqrt{6 + \frac{16LR^2n + 8B}{n\varepsilon}},$$

получаем следующую оценку на число итераций:

$$N = \left\lceil 2 \left( n + \sqrt{n^2 + \frac{128nLR^2}{\varepsilon}} \right) \ln \left( \frac{4RA}{\varepsilon} \right) \right\rceil,$$

где  $A = 2 \left( LR + \frac{\varepsilon}{8R} \right) \sqrt{6 + \frac{16LR^2n + 8B}{n\varepsilon}}$ .

**Замечание 3.** Отметим, что оценку сложности алгоритма 6 можно также представить в виде  $O\left(\max\left\{n, \sqrt{nLR^2/\varepsilon}\right\} \ln\left(\frac{1}{\varepsilon}\right)\right)$ , где логарифмический множитель появляется за счет необходимости регуляризации двойственной задачи. При этом на каждой итерации вычисляется только одна компонента вектора реакции пользователя на изменение цены, соответственно, арифметическая сложность операции лучше, чем при вычислении всех компонент этих векторов. Для БГМ предположения для целевой функции аналогичны, но за счет того, что на каждой итерации необходимо вычислять полный градиент, стоимость работы алгоритма, соответственно,  $O\left(n\sqrt{LR^2/\varepsilon}\right)$ . Таким образом, несмотря на то, что теоретическая оценка скорости сходимости для RGEM имеет тот же порядок, что и для БГМ, на практике выигрыш получается за счет более дешевых вычислений в рамках одной итерации.

## 8. ВЫЧИСЛИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Программный код для численных экспериментов был написан на языках программирования Python версии 3.6 и C++ стандарта C++14. Исходный код для экспериментов и рассмотренных в статье методов опубликован на GitHub и доступен по ссылке: <https://github.com/dmivilen-sky/network-resource-allocation>. Измерение времени работы производилось на компьютере с

**Таблица 1.** Сравнение количества итераций и времени работы БГМ и метода экстраполяции случайного градиента для сильно выпуклых (квадратичных) функций полезности

Сеть	FGM		RGEM	
	Итерации	Время	Итерации	Время
$m = 2, n = 1500, \epsilon = 10^{-2}$	350	24.5 с	3000	21.1 с
$m = 5, n = 1500, \epsilon = 10^{-2}$	380	42.7 с	6700	36.9 с
$m = 70, n = 5000, \epsilon = 10^{-2}$	400	150.0 с	7800	132.6 с
$m = 70, n = 5000, \epsilon = 10^{-3}$	1070	374.5 с	9180	283.7 с
$m = 100, n = 5000, \epsilon = 10^{-2}$	417	175.1 с	8200	164.0 с
$m = 70, n = 7000, \epsilon = 10^{-2}$	421	218.9 с	8600	206.4 с
$m = 100, n = 7000, \epsilon = 10^{-2}$	427	290.3 с	9200	276.0 с
$m = 100, n = 7000, \epsilon = 10^{-3}$	1120	761.6 с	10130	638.2 с

2-ядерным процессором Intel Core i5-5250U с тактовой частотой 1.6 ГГц на каждое ядро, ОЗУ компьютера составляла 8 Гб.

*8.1. Сильно выпуклые (квадратичные) функции полезности*

Рассмотрим задачу (1) для функций полезности следующего вида:

$$u_k(x_k) = a_k x_k - \frac{\sigma n}{2} x_k^2, \quad a_k \sim \mathcal{U}(0,100), \quad \sigma = 0.1,$$

где  $a_k$  – независимые и одинаково распределенные случайные величины. Тогда задачу (3) можно решить явно:

$$x(\lambda) = \frac{[a - C\lambda]_+}{n\sigma}.$$

Для малого числа пользователей ( $n = 1500$ ) пропускные способности соединений выбираются одинаковыми (в данном случае  $\mathbf{b} = (5, \dots, 5)^T$ ), а спрос на передачу информации равномерен ( $c_{ij} = 1$  для любых  $i, j$ ). Для большего числа пользователей вектор пропускных способностей генерируется случайно, так что  $b_i \sim \mathcal{U}(1,6)$ . Также случайно и независимо выбираются элементы матрицы спроса, так что  $c_{ij} = 1$  с вероятностью  $p = 0.5$  и  $c_{ij} = 0$  с вероятностью  $q = 0.5$ .

В табл. 1 представлены число итераций и время работы быстрого градиентного метода (FGM) и метода экстраполяции случайного градиента (RGEM) для различных конфигураций сети (с числом соединений  $m$ ), числа пользователей  $n$  и требуемой точности  $\epsilon$ . В таблице выделены те случаи, в которых метод экстраполяции случайного градиента сходится к решению за меньшее время, чем быстрый градиентный метод, несмотря на большее число итераций по сравнению с БГМ. Действительно, для  $n \gg 0$  число требующихся запросов оптимального решения  $x_k(\lambda)$  от пользователей, в случае метода экстраполяции случайного градиента, будет меньшим, чем при использовании других алгоритмов, так как за одну итерацию метода экстраполяции случайного градиента запрос отправляется только к одному случайному пользователю.

*8.2. Выпуклые (логарифмические) функции полезности*

Рассмотрим работу стохастического субградиентного метода (алгоритм 9) и метода эллипсоидов (алгоритм 4) для функции полезности следующего вида:

$$u_k(x_k) = \ln x_k.$$

В таком случае явное решение задачи (3) выглядит следующим образом (операция  $1 / \cdot$  применительно к вектору понимается поэлементно):

$$x(\lambda) = \frac{1}{C\lambda}.$$

**Таблица 2.** Сравнение количества итераций и времени работы стохастического субградиентного метода и метода эллипсоидов для выпуклых (логарифмических) функций полезности

Сеть	Метод эллипсоидов		SGM	
	Итерации	Время	Итерации	Время
$m = 2, n = 1500, \varepsilon = 10^{-2}$	40	0.02 с	2000	0.2 с
$m = 5, n = 1500, \varepsilon = 10^{-2}$	85	0.06 с	2500	0.3 с
$m = 70, n = 5000, \varepsilon = 10^{-2}$	120	1.9 с	4000	1.3 с
$m = 70, n = 5000, \varepsilon = 10^{-3}$	800	5.4 с	9020	2.4 с
$m = 100, n = 5000, \varepsilon = 10^{-2}$	300	9.0 с	5000	3.1 с
$m = 70, n = 7000, \varepsilon = 10^{-2}$	250	8.7 с	5590	5.5 с
$m = 100, n = 7000, \varepsilon = 10^{-2}$	380	19.0 с	6480	10.8 с
$m = 100, n = 7000, \varepsilon = 10^{-3}$	1830	91.5 с	17970	30.6 с

Для малого числа пользователей ( $n = 1500$ ) пропускные способности соединений выбираются одинаковыми (в данном случае  $\mathbf{b} = (5, \dots, 5)^T$ ), а спрос на передачу информации равномерен ( $c_{ij} = 1$  для любых  $i, j$ ). Для большего числа пользователей вектор пропускных способностей генерируется случайно, так что  $b_i \sim \mathcal{U}(1, 6)$ . Также случайно и независимо выбираются элементы матрицы спроса, так что  $c_{ij} = 1$  с вероятностью  $p = 0.5$  и  $c_{ij} = 0$  с вероятностью  $q = 0.5$ .

В табл. 2 представлены число итераций и время работы стохастического субградиентного метода (SGM) и метода эллипсоидов для различных конфигураций сети, числа пользователей и требуемой точности. В таблице выделены те случаи, в которых стохастический субградиентный метод сходится к решению за меньшее время, чем метод эллипсоидов.

Отметим, что аналогично методу экстраполяции случайного градиента, стохастический субградиентный метод требует вычисления лишь одной компоненты вектора  $\mathbf{x}(\lambda)$  реакций пользователей на установившиеся цены на каждой итерации. Таким образом, при большем числе итераций метода число вычисляемых компонент  $x_k(\lambda')$  будет меньшим по сравнению с другими алгоритмами, например, с методом эллипсоидов, также как и коммуникационная сложность в случае распределенной реализации.

## 9. ЗАКЛЮЧЕНИЕ

В заключение отметим некоторые возможные развития данной работы и кратко опишем методы без подробного анализа оценки скорости сходимости.

В разд. 5 мы рассматривали метод эллипсоидов для задач небольшой размерности, который является прямо двойственным. Существуют и другие методы, которые дают высокую точность и хорошо подходят для задач небольшой размерности. Одним из таких методов является метод Вайды [28]. Однако для восстановления решения прямой задачи при решении двойственной задачи методом Вайды необходимо, чтобы была сходимостью по норме градиента для двойственной задачи. Для этого требуется гладкость двойственной задачи, что порождается сильной выпуклостью целевой функции прямой задачи (утверждение 2). Если сильной выпуклости в прямой задаче нет, то ее можно регуляризовать, как было описано в разд. 6, но при этом оценка сходимости логарифмически ухудшится.

Также для решения двойственной задачи можно применить методы высокого порядка [29], [30], если двойственная задача достаточно гладкая. При этом шаги данных методов можно будет считать распределенно, так как в данной задаче рассматривается централизованная архитектура с точки зрения взаимодействия соединения и использующих его пользователей. Однако отметим, что оптимальные методы высокого порядка, требующие одномерного поиска и не обладающие прямо двойственностью, необходимо применять только после регуляризации двойственной задачи.

Еще одно направление — методы типа редукции дисперсии, например, [31], [32], которые являются промежуточными между стохастическим градиентным методом и БГМ. Однако данные

методы тоже не прямо двойственные, поэтому их необходимо применять к предварительно регуляризованной двойственной задаче.

Особый интерес для авторов представляют метод Hoggwild! [33] и методы с использованием мини-батчинга. Заметим, что при этом одновременно данные передают не все пользователи, но и не только один пользователь, как это получается при использовании стохастических методов. Выбирая в качестве размера батча количество пользователей, которые отправляют данные в один момент времени, можно учесть специфику работы реальных сетей.

ПРИЛОЖЕНИЕ

Вспомогательные результаты

Приведем некоторые леммы из других работ, на которые мы будем ссылаться в доказательствах. Также приведем доказательства утверждений о свойствах двойственной функции, которые используются при доказательстве основных теорем.

**Лемма 6** (см. [34, лемма 2]). Для случайного вектора  $\xi \in \mathbf{R}^n$  следующие утверждения эквивалентны с точностью до константного множителя у  $\sigma$ .

1. Хвосты:  $\mathbb{P}\{\|\xi\|_2 \geq \gamma\} \leq 2 \exp\left(-\frac{\gamma^2}{2\sigma^2}\right) \forall \gamma \geq 0$ .
2. Моменты:  $(\mathbb{E}\{\xi^p\})^{1/p} \leq \sigma \sqrt{p}$  для любого положительного целого  $p$ .
3. Предположение легких хвостов:  $\mathbb{E}\left[\exp\left(\frac{\|\xi\|_2^2}{\sigma^2}\right)\right] \leq \exp(1)$ .

**Лемма 7** (см. [34, следствие 8]). Пусть  $\{\xi^k\}_{k=1}^N$  – последовательность случайных векторов из  $\mathbf{R}^n$  таких, что для  $k = 1, \dots, N$  и для любых  $\gamma \geq 0$

$$\mathbb{E}[\xi^k | \xi^1, \dots, \xi^{k-1}] = 0, \quad \mathbb{E}\left[\|\xi^k\|_2 \geq \gamma | \xi^1, \dots, \xi^{k-1}\right] \leq \exp\left(-\frac{\gamma^2}{2\sigma_k^2}\right) \text{ почти наверное,}$$

где  $\sigma_k^2$  принадлежит  $\sigma(\xi^1, \dots, \xi^{k-1})$  для всех  $k = 1, \dots, N$ . Пусть  $S_N = \sum_{k=1}^N \xi^k$ . Тогда существует константа  $C_1$  такая, что для любых фиксированных  $\delta > 0$  и  $B > b > 0$  с вероятностью  $1 - \delta$  выполняется:

$$\text{либо } \sum_{k=1}^N \sigma_k^2 \geq B,$$

$$\text{либо } \|S_N\|_2 \leq C_1 \sqrt{\max\left\{\sum_{k=1}^N \sigma_k^2, b\right\} \left(\ln \frac{2n}{\delta} + \ln \ln \frac{B}{b}\right)}.$$

**Лемма 8** (см. [35, следствие из теор. 2.1, случай (ii)]). Пусть  $\{\xi^k\}_{k=1}^N$  – последовательность случайных векторов из  $\mathbf{R}^n$  удовлетворяет условию

$$\mathbb{E}[\xi^k | \xi^1, \dots, \xi^{k-1}] = 0 \text{ почти наверное, } k = 1, \dots, N$$

и пусть  $S_N = \sum_{k=1}^N \xi^k$ . Пусть последовательность  $\{\xi^k\}_{k=1}^N$  удовлетворяет “light-tail”-условию:

$$\mathbb{E}\left[\exp\left(\frac{\|\xi^k\|_2^2}{\sigma_k^2}\right) | \xi^1, \dots, \xi^{k-1}\right] \leq \exp(1) \text{ почти наверное, } k = 1, \dots, N,$$

где  $\sigma_1, \dots, \sigma_N$  – положительные числа. Тогда для всех  $\gamma \geq 0$  выполняется

$$\mathbb{P}\left\{\|S_N\| \geq (\sqrt{2} + \sqrt{2\gamma}) \sqrt{\sum_{k=1}^N \sigma_k^2}\right\} \leq \exp\left(-\frac{\gamma^2}{3}\right).$$

**Доказательство утверждения 2.** Представим двойственную функцию в следующем виде:

$$\varphi(\lambda) = \sum_{k=1}^n \left\{ u_k(x_k(\lambda)) - \langle \lambda, \mathbf{C}_k \rangle x_k(\lambda) + \frac{1}{n} \langle \lambda, \mathbf{b} \rangle \right\} = \sum_{k=1}^n \varphi_k(\lambda),$$

при этом из утверждения 1 получаем, что

$$\nabla \varphi(\lambda) = \sum_{k=1}^n \nabla \varphi_k(\lambda) = \sum_{k=1}^n \left( \frac{1}{n} \mathbf{b} - \mathbf{C}_k x_k(\lambda) \right).$$

Определим

$$x_k(\lambda^1) = \arg \max_{x_k \in \mathbf{R}_+} \left\{ u_k(x_k) - x_k \langle \lambda^1, \mathbf{C}_k \rangle \right\},$$

$$x_k(\lambda^2) = \arg \max_{x_k \in \mathbf{R}_+} \left\{ u_k(x_k) - x_k \langle \lambda^2, \mathbf{C}_k \rangle \right\}.$$

Запишем необходимые условия максимума первого порядка

$$\langle \nabla u_k(x_k(\lambda^1)) - \langle \lambda^1, \mathbf{C}_k \rangle, x_k(\lambda^1) - x_k(\lambda^2) \rangle \geq 0,$$

$$\langle \nabla u_k(x_k(\lambda^2)) - \langle \lambda^2, \mathbf{C}_k \rangle, x_k(\lambda^2) - x_k(\lambda^1) \rangle \geq 0.$$

Складывая эти неравенства, получаем

$$\langle \nabla u_k(x_k(\lambda^2)) - \nabla u_k(x_k(\lambda^1)), x_k(\lambda^1) - x_k(\lambda^2) \rangle \leq \langle \langle \lambda^2, \mathbf{C}_k \rangle - \langle \lambda^1, \mathbf{C}_k \rangle, x_k(\lambda^1) - x_k(\lambda^2) \rangle.$$

В силу сильной вогнутости  $u_k(x_k)$  для любых  $x_k^1$  и  $x_k^2$ ,  $k = 1, \dots, n$ , выполняется

$$\langle \nabla u_k(x_k^2) - \nabla u_k(x_k^1), x_k^1 - x_k^2 \rangle \geq \mu \|x_k^1 - x_k^2\|_2^2.$$

Отсюда получаем, что

$$\mu \|x_k(\lambda^1) - x_k(\lambda^2)\|_2^2 \leq \langle \langle \lambda^2, \mathbf{C}_k \rangle - \langle \lambda^1, \mathbf{C}_k \rangle, x_k(\lambda^1) - x_k(\lambda^2) \rangle \leq \|\mathbf{C}_k\|_2 \cdot \|\lambda^1 - \lambda^2\|_2 \cdot \|x_k(\lambda^1) - x_k(\lambda^2)\|_2.$$

Тогда можно получить следующую оценку для всех  $\nabla \varphi_k$  градиента

$$\|\nabla \varphi_k(\lambda^1) - \nabla \varphi_k(\lambda^2)\|_2 \leq \|\mathbf{C}_k\|_2 \cdot \|x_k(\lambda^1) - x_k(\lambda^2)\|_2 \leq \frac{1}{\mu} \|\mathbf{C}_k\|_2 \cdot \|\lambda^1 - \lambda^2\|_2.$$

Для матрицы  $C$  с учетом ее структуры верна оценка  $\|\mathbf{C}_k\|_2 \leq m$ . Тогда для градиента двойственной функции

$$\|\nabla \varphi(\lambda^1) - \nabla \varphi(\lambda^2)\|_2 \leq \sum_{k=1}^n \|\nabla \varphi_k(\lambda^1) - \nabla \varphi_k(\lambda^2)\|_2 \leq \frac{m^2 n}{\mu} \|\lambda^1 - \lambda^2\|_2.$$

**Доказательство леммы 1.** Для доказательства леммы 1 сначала сформулируем и докажем одну техническую лемму.

Обозначим  $d_L(\lambda) = \frac{L}{2} \|\lambda - \lambda^0\|_2^2$  и рассмотрим последовательности

$$l_t(\lambda) = \sum_{j=0}^t \alpha_j \left[ \varphi(\lambda^j) + \langle \nabla \varphi(\lambda^j), \lambda - \lambda^j \rangle \right]$$

и

$$\psi_t(\lambda) = l_t(\lambda) + d_L(\lambda), \quad t = 0, 1, \dots,$$

где  $\{\lambda^j\}_{j \geq 0}$  — последовательность точек, генерируемых алгоритмом 1.

**Лемма 9.** После  $N$  шагов алгоритма 1 выполняется следующее неравенство:

$$A_N \varphi(\mathbf{y}^N) \leq \min_{\lambda \in \mathbf{R}_+^m} \psi_N(\lambda) = \psi_N(\mathbf{z}^N). \quad (\text{П.20})$$

**Доказательство леммы 9.** Докажем по индукции, что (П.20) верно. При  $t = 0$  неравенство (П.20) выполняется. Действительно,

$$\begin{aligned} \psi_0 &= \min_{\lambda \in \mathbb{R}_+^m} \left\{ \alpha_0 \left[ \varphi(\lambda^0) + \langle \nabla \varphi(\lambda^0), \lambda - \lambda^0 \rangle \right] + \frac{L}{2} \|\lambda - \lambda^0\|_2^2 \right\}^{\textcircled{1}} \\ &\stackrel{\textcircled{1}}{\geq} \alpha_0 \min_{\lambda \in \mathbb{R}_+^m} \left\{ \varphi(\lambda^0) + \langle \nabla \varphi(\lambda^0), \lambda - \lambda^0 \rangle + \frac{L}{2} \|\lambda - \lambda^0\|_2^2 \right\} \stackrel{\textcircled{2}}{\geq} \alpha_0 \varphi(y_0), \end{aligned}$$

где  $\textcircled{1}$  выполняется, так как  $\alpha_0 = 1/2 \leq 1$ , а  $\textcircled{2}$  – в силу того, что функция  $\varphi(\lambda)$  имеет липшицев градиент (см. утверждение 2 и [36, лемма 1.2.3]). Итак,  $A_0 \varphi(y^0) = \frac{1}{2} \varphi(y^0) \leq \psi_0$ .

Пусть (П.20) верно при  $t$ :

$$A_t \varphi(y^t) \leq \psi_t(z^t). \tag{П.21}$$

Докажем, что (П.20) верно при  $t + 1$ . Действительно, имеем,

$$\begin{aligned} \psi_{t+1}(z^{t+1}) &= \min_{\lambda \in \mathbb{R}_+^m} \left\{ \psi_t(\lambda) + \alpha_{t+1} \left[ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right] \right\}^{\textcircled{1}} \\ &\stackrel{\textcircled{1}}{\geq} \min_{\lambda \in \mathbb{R}_+^m} \left\{ \psi_t(z^t) + \frac{L}{2} \|\lambda - z^t\|_2^2 + \alpha_{t+1} \left[ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right] \right\}^{\textcircled{2}} \\ &\stackrel{\textcircled{2}}{\geq} \min_{\lambda \in \mathbb{R}_+^m} \left\{ A_t \varphi(y^t) + \frac{L}{2} \|\lambda - z^t\|_2^2 + \alpha_{t+1} \left[ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right] \right\}^{\textcircled{3}} \\ &\stackrel{\textcircled{3}}{\geq} \min_{\lambda \in \mathbb{R}_+^m} \left\{ A_t \left( \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right) + \frac{L}{2} \|\lambda - z^t\|_2^2 + \alpha_{t+1} \left[ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right] \right\}, \end{aligned}$$

где  $\textcircled{1}$  выполняется в силу сильной выпуклости прокс-функции  $\frac{1}{2} \|\lambda - \lambda^0\|_2^2$  и свойств экстремума в точке  $z^t$ ,  $\textcircled{2}$  следует из (П.21),  $\textcircled{3}$  – в силу выпуклости функции  $\varphi(\lambda)$ .

Так как между коэффициентами  $A_t$  и  $\alpha_t$  БГМ есть следующая зависимость:  $A_{t+1} = \sum_{j=0}^{t+1} \alpha_j = A_t + \alpha_{t+1}$  и  $\tau_t = \alpha_{t+1}/A_{t+1}$ , соотношение  $\lambda^{t+1} = \tau_t z^t + (1 - \tau_t) y^t$  из алгоритма 1 можно переписать в виде:

$$A_{t+1} \lambda^{t+1} = \alpha_{t+1} z^t + A_t y^t.$$

Используя последние соотношения, можно сделать следующие преобразования:

$$\begin{aligned} A_t \langle \nabla \varphi(\lambda^{t+1}), y^t - \lambda^{t+1} \rangle + \alpha_{t+1} \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle &= -A_{t+1} \langle \nabla \varphi(\lambda^{t+1}), \lambda^{t+1} \rangle + \\ &+ \alpha_{t+1} \langle \nabla \varphi(\lambda^{t+1}), \lambda \rangle + A_t \langle \nabla \varphi(\lambda^{t+1}), y^t \rangle = \alpha_{t+1} \langle \nabla \varphi(\lambda^{t+1}), \lambda - z^t \rangle. \end{aligned}$$

Тогда имеем

$$\begin{aligned} A_t \left( \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), y^t - \lambda^{t+1} \rangle \right) + \frac{L}{2} \|\lambda - z^t\|_2^2 + \alpha_{t+1} \left[ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle \right] &= \\ = A_{t+1} \varphi(\lambda^{t+1}) + \frac{L}{2} \|\lambda - z^t\|_2^2 + \alpha_{t+1} \langle \nabla \varphi(\lambda^{t+1}), \lambda - z^t \rangle. \end{aligned} \tag{П.23}$$

После замены последнего выражения в (П.22) на (П.23) можно воспользоваться расширенным вариантом неравенства Фенхеля для сопряженных функций [37]:

$$\langle \mathbf{g}, \mathbf{s} \rangle + \frac{\xi}{2} \|\mathbf{s}\|^2 \geq -\frac{1}{2\xi} \|\mathbf{g}\|_*^2, \quad \mathbf{g} \in \mathbb{E}^*, \quad \mathbf{s} \in \mathbb{E},$$

где  $\mathbb{E}$  – конечномерное вещественное векторное пространство,  $\mathbb{E}^*$  – пространство линейных функций на  $\mathbb{E}$  (двойственное пространство), норма в двойственном пространстве  $\|\mathbf{g}\|_* = \max_{\mathbf{x}} \{\langle \mathbf{g}, \mathbf{x} \rangle \mid \|\mathbf{x}\|_{\mathbb{E}} = 1\}$ . В нашем случае  $\mathbf{g} = \nabla \varphi(\lambda^{t+1})$ ,  $\mathbf{s} = \lambda - \mathbf{z}^t$ ,  $\xi = \frac{L}{\alpha_{t+1}}$ . Следовательно,

$$\psi_{t+1}(\mathbf{z}^{t+1}) \geq A_{t+1}\varphi(\lambda^{t+1}) - \frac{\alpha_{t+1}^2}{2L} \|\nabla \varphi(\lambda^{t+1})\|_2^2. \quad (\text{П.24})$$

Для завершения доказательства леммы требуется показать, что  $A_{t+1}\varphi(\mathbf{y}^{t+1})$  меньше, чем правая часть неравенства в (П.24).

В силу  $L$ -гладкости функции  $\varphi(\lambda)$  (см. утверждение 2)

$$\begin{aligned} \varphi(\mathbf{y}^{t+1}) &\leq \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \mathbf{y}^{t+1} - \lambda^{t+1} \rangle + \frac{L}{2} \|\mathbf{y}^{t+1} - \lambda^{t+1}\|_2^2 = \\ &= \min_{\lambda} \left\{ \varphi(\lambda^{t+1}) + \langle \nabla \varphi(\lambda^{t+1}), \lambda - \lambda^{t+1} \rangle + \frac{L}{2} \|\lambda - \lambda^{t+1}\|_2^2 \right\} = \varphi(\lambda^{t+1}) - \frac{1}{2L} \|\nabla \varphi(\lambda^{t+1})\|_2^2. \end{aligned}$$

После умножения обеих частей полученного неравенства на  $A_{t+1}$ :

$$A_{t+1}\varphi(\mathbf{y}^{t+1}) \leq A_{t+1}\varphi(\lambda^{t+1}) - \frac{A_{t+1}}{2L} \|\nabla \varphi(\lambda^{t+1})\|_2^2.$$

В силу того, что для коэффициентов БГМ  $\alpha_{t+1}^2 \leq A_{t+1}$ , получаем

$$A_{t+1}\varphi(\mathbf{y}^{t+1}) \leq A_{t+1}\varphi(\lambda^{t+1}) - \frac{\alpha_{t+1}^2}{2L} \|\nabla \varphi(\lambda^{t+1})\|_2^2. \quad (\text{П.25})$$

Следовательно, в силу (П.24) и (П.25)  $A_{t+1}\varphi(\mathbf{y}^{t+1}) \leq \psi_{t+1}(\mathbf{z}^{t+1})$ , что и требовалось доказать.

**Доказательство леммы 1.** Определим следующее множество:

$$\Lambda_{2\hat{R}} = \{\lambda \in \mathbb{R}_+^m : \|\lambda\|_2 \leq 2\hat{R}\}.$$

$\hat{R}$  определяется в силу следующих неравенств:

$$\|\lambda^0 - \lambda^*\|_2 + \|\lambda^0\|_2 \leq \|\lambda^*\|_2 + 2\|\lambda^0\|_2 \leq 3R = \hat{R}.$$

При этом все  $\lambda^t$  будут принадлежать  $\Lambda_{2\hat{R}}$ , так как

$$\|\lambda^t\|_2 \leq \|\lambda^t - \lambda^*\|_2 + \|\lambda^* - \lambda^0\|_2 + \|\lambda^0\|_2 \leq 2\|\lambda^* - \lambda^0\|_2 + \|\lambda^0\|_2 \leq 2\|\lambda^*\|_2 + 3\|\lambda^0\|_2 \leq 5R \leq 2\hat{R},$$

где для второго неравенства учитывалось, что  $\|\lambda^t - \lambda^*\|_2 \leq \|\lambda^* - \lambda^0\|_2$ ,  $t = 0, 1, \dots$ .

Последнее неравенство можно доказать следующим образом. Для любого  $\lambda \in \mathbf{R}_+^m$  в силу леммы 9 и сильной выпуклости функции  $\psi_t(\lambda)$  с константой  $L$  верно

$$\begin{aligned} A_t\varphi(\mathbf{y}^t) + \frac{L}{2} \|\lambda - \mathbf{z}^t\|_2^2 &\leq \psi_t(\mathbf{z}^t) + \frac{L}{2} \|\lambda - \mathbf{z}^t\|_2^2 \leq \psi_t(\lambda) = \\ &= \sum_{j=0}^t \alpha_j \left[ \varphi(\lambda^j) + \langle \nabla \varphi(\lambda^j), \lambda - \lambda^j \rangle \right] + \frac{L}{2} \|\lambda - \lambda^0\|_2^2. \end{aligned} \quad (\text{П.26})$$

Последнее выражение в (П.26) в силу выпуклости функции  $\varphi(\lambda)$  можно оценить сверху как  $A_t\varphi(\lambda) + \frac{L}{2} \|\lambda - \lambda^0\|_2^2$ . Тогда при  $\lambda = \lambda^*$

$$\frac{L}{2} \|\lambda^* - \mathbf{z}^t\|_2^2 \leq A_t(\varphi(\mathbf{y}^t) - \varphi(\lambda^*)) + \frac{L}{2} \|\lambda^* - \mathbf{z}^t\|_2^2 \leq \frac{L}{2} \|\lambda^* - \lambda^0\|_2^2.$$

Следовательно,

$$\|\lambda^* - \mathbf{z}^t\|_2 \leq \|\lambda^* - \lambda^0\|_2. \quad (\text{П.27})$$

Поскольку  $\mathbf{y}^t$  в алгоритме 1 определяется с помощью шага метода проекции градиента для выпуклой функции  $\varphi(\lambda)$ , последовательность генерируемых алгоритмом точек  $\mathbf{y}^t$ ,  $t = 0, 1, \dots$  также будет ограничена (доказательство этого факта см., например, в [38, лемма 9.17, с. 183] или в [28, с. 265]):

$$\|\lambda^* - \mathbf{y}^t\|_2 \leq \|\lambda^* - \lambda^0\|_2. \tag{П.28}$$

Далее

$$\|\lambda^{t+1} - \lambda^*\|_2 = \|\tau_t(\mathbf{z}^t - \lambda^*) + (1 - \tau_t)(\mathbf{y}^t - \lambda^*)\|_2 \leq \tau_t \|\mathbf{z}^t - \lambda^*\|_2 + (1 - \tau_t) \|\mathbf{y}^t - \lambda^*\|_2.$$

Из последнего неравенства с помощью (П.27), (П.28) получаем нужный результат:

$$\|\lambda^{t+1} - \lambda^*\|_2 \leq \|\lambda^* - \lambda^0\|_2, \quad t = -1, 0, 1, \dots$$

В силу леммы 9

$$\begin{aligned} A_N \varphi(\mathbf{y}^N) &\leq \min_{\lambda \in \mathbf{R}_+^n} \left\{ \frac{L}{2} \|\lambda - \lambda^0\|_2^2 + \sum_{t=0}^N \alpha_t [\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle] \right\} \leq \\ &\leq \min_{\lambda \in \Lambda_{2\hat{R}}} \left\{ \frac{L}{2} \|\lambda - \lambda^0\|_2^2 + \sum_{t=0}^N \alpha_t [\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle] \right\} \leq \\ &\stackrel{\textcircled{1}}{\leq} \min_{\lambda \in \Lambda_{2\hat{R}}} \left\{ \sum_{t=0}^N \alpha_t [\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle] \right\} + \frac{37L\hat{R}^2}{9}, \end{aligned}$$

где  $\textcircled{1}$  выполняется, так как

$$\|\lambda - \lambda^0\|_2^2 \leq 2\|\lambda\|_2^2 + 2\|\lambda^0\|_2^2 \leq 8\hat{R}^2 + \frac{2}{9}\hat{R}^2 = \frac{74}{9}\hat{R}^2. \tag{П.29}$$

После подстановки определений двойственной целевой функции  $\varphi(\lambda^t)$  (2) и ее градиента  $\nabla \varphi(\lambda^t)$  (см. утверждение 1):

$$\begin{aligned} &\sum_{t=0}^N \alpha_t [\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle] = \\ &= \sum_{t=0}^N \alpha_t \left( \langle \lambda^t, \mathbf{b} \rangle + \sum_{k=1}^n (u_k(x_k^t(\lambda^t)) - \langle \lambda^t, \mathbf{C}_k x_k^t(\lambda^t) \rangle) + \left\langle \mathbf{b} - \sum_{k=1}^n \mathbf{C}_k x_k^t(\lambda^t), \lambda - \lambda^t \right\rangle \right) = \\ &= \sum_{t=0}^N \alpha_t \left( \sum_{k=1}^n u_k(x_k^t(\lambda^t)) + \left\langle \lambda, \mathbf{b} - \sum_{k=1}^n \mathbf{C}_k x_k^t(\lambda^t) \right\rangle \right) \leq A_N (U(\hat{\mathbf{x}}^N) + \langle \lambda, \mathbf{b} - \mathbf{C}\hat{\mathbf{x}}^N \rangle), \end{aligned}$$

где последнее неравенство выполняется в силу вогнутости функций полезности.

Итак,

$$\begin{aligned} A_N \varphi(\mathbf{y}^N) &\leq A_N U(\hat{\mathbf{x}}^N) + \frac{37L\hat{R}^2}{9} + A_N \min_{\lambda \in \Lambda_{2\hat{R}}} \{ \langle \lambda, \mathbf{b} - \mathbf{C}\hat{\mathbf{x}}^N \rangle \} = A_N U(\hat{\mathbf{x}}^N) + \frac{37L\hat{R}^2}{9} - \\ &- A_N \max_{\lambda \in \Lambda_{2\hat{R}}} \{ \langle \lambda, \mathbf{C}\hat{\mathbf{x}}^N - \mathbf{b} \rangle \} = A_N U(\hat{\mathbf{x}}^N) + \frac{37L\hat{R}^2}{9} - 2\hat{R}A_N \|(C\hat{\mathbf{x}}^N - \mathbf{b})_+\|_2. \end{aligned}$$

Из этого получаем оценку (6).

**Доказательство леммы 2.** Для доказательства леммы 2 сначала приведем доказательства нескольких вспомогательных технических лемм.

**Лемма 10.** Пусть  $A, B$  и  $\{r_l\}_{l=0}^N$  – неотрицательные числа такие, что для любого  $l = 1, \dots, N$  выполняется неравенство

$$\frac{1}{2} r_l^2 \leq A r_0^2 + B r_0 \sqrt{\sum_{t=0}^{l-1} r_t^2}. \tag{П.30}$$



Тогда верно следующее неравенство:

$$r_l \leq Cr_0, \tag{П.31}$$

где  $C$  – положительная константа, для которой выполняется  $C^2 \geq \max\{1, 2A + 2BC\sqrt{N}\}$ , т.е., в частности, можно выбрать

$$C = \max\left\{1, B\sqrt{N} + \sqrt{B^2N + 2A}\right\}.$$

**Доказательство.** Докажем (П.31) по индукции. Для  $l = 0$  неравенство выполнено, так как  $C \geq 1$ . Пусть (П.31) выполнено для всех  $l < N$ . Докажем, что оно выполнено и для  $l + 1$ . Действительно,

$$r_{l+1} \stackrel{(П.30)}{\leq} \sqrt{2} \sqrt{Ar_0^2 + Br_0 \sqrt{\sum_{t=0}^l r_t^2}} \stackrel{(П.31)}{\leq} r_0 \sqrt{2} \sqrt{A + BC\sqrt{N}} = r_0 \underbrace{\sqrt{2A + 2BC\sqrt{N}}}_{\leq C} \leq Cr_0.$$

**Лемма 11.** Пусть для последовательностей неотрицательных коэффициентов  $\{R_t\}_{t \geq 0}$  и случайных векторов  $\{\boldsymbol{\eta}^t\}_{t \geq 0}$ ,  $\{\mathbf{a}^t\}_{t \geq 0}$  для всех  $l = 1, \dots, N$  выполняется неравенство

$$\frac{1}{2} R_l^2 \leq A + u \sum_{t=0}^{l-1} \langle \boldsymbol{\eta}^t, \mathbf{a}^t \rangle, \tag{П.32}$$

где  $A$  – неотрицательная константа,  $d \geq 1$  – положительная константа,  $\|\mathbf{a}^t\|_2 \leq \tilde{R}_t d$  и  $\tilde{R}_t = \max\{\tilde{R}_{t-1}, R_t\}$  для всех  $t \geq 1$ ,  $\tilde{R}_0 = R_0$ ,  $\tilde{R}_t$  зависит только от  $\boldsymbol{\eta}^0, \dots, \boldsymbol{\eta}^t$ . Пусть также вектор  $\mathbf{a}^t$  – это функция от  $\boldsymbol{\eta}^0, \dots, \boldsymbol{\eta}^{t-1} \forall t \geq 1$ ,  $\mathbf{a}^0$  – постоянный вектор и для любого  $t \geq 0$

$$\mathbb{E}[\boldsymbol{\eta}^t | \{\boldsymbol{\eta}^k\}_{k=0}^{t-1}] = \mathbf{0}, \quad \mathbb{E}\left[\exp\left(\|\boldsymbol{\eta}^t\|_2^2 \sigma^{-2}\right) | \{\boldsymbol{\eta}^k\}_{k=0}^{t-1}\right] \leq \exp(1).$$

Тогда с вероятностью  $1 - 2\delta$  выполняются следующие неравенства:

$$\tilde{R}_l \leq JR_0 \quad \text{и} \quad A + u \sum_{t=0}^{l-1} \langle \boldsymbol{\eta}^t, \mathbf{a}^t \rangle \leq A + udD\sqrt{\sigma^2 g(N)NJ} \tilde{R}_0^2$$

$\forall l = 1, \dots, N$  одновременно, где  $D$  – положительная константа,

$$F = 2\sigma^2 d^2 N (2ud)^N \left(2A + ud\tilde{R}_0^2 + 12ud \ln \frac{N}{\delta} \sigma^2 N\right),$$

$$f = d^2 \sigma^2 \tilde{R}_0^2, \quad g(N) = \ln\left(\frac{N}{\delta}\right) + \ln \ln\left(\frac{F}{f}\right) \text{ и}$$

$$J = \max\left\{1, \frac{1}{\tilde{R}_0} udD\sqrt{\sigma^2 g(N)} + \sqrt{\frac{1}{\tilde{R}_0^2} u^2 d^2 C_1^2 \sigma^2 g(N) + \frac{2A}{\tilde{R}_0^2}}\right\}.$$

**Доказательство.** Применим ко второму слагаемому из правой части (П.32) неравенство Коши–Буняковского:

$$\frac{1}{2} R_l^2 \leq A + ud \sum_{t=0}^{l-1} \|\boldsymbol{\eta}^t\|_2 \tilde{R}_t \leq A + \frac{ud}{2} \sum_{t=0}^{l-1} \tilde{R}_t^2 + \frac{ud}{2} \sum_{t=0}^{l-1} \|\boldsymbol{\eta}^t\|_2^2. \tag{П.33}$$

По теореме 2.1 из [35]

$$(\forall N \geq 1, \forall \gamma \geq 0): \quad \mathbb{P}\left\{\left\|\sum_{t=0}^{N-1} \boldsymbol{\eta}^t\right\|_2 \geq (\sqrt{2} + \sqrt{2}\gamma) \sqrt{\sum_{t=0}^{N-1} \sigma_t^2}\right\} \leq \exp\left(-\frac{\gamma^2}{3}\right). \tag{П.34}$$

Тогда с вероятностью не меньшей, чем

$$1 - \frac{\delta}{N} = 1 - \exp\left(-\frac{\gamma^2}{3}\right) \tag{П.35}$$

выполняется следующее неравенство:

$$\|\boldsymbol{\eta}^t\|_2 \leq \sqrt{2} \left(1 + \sqrt{3 \ln \frac{N}{\delta}}\right) \sigma \leq 2\sqrt{6 \ln \frac{N}{\delta}} \sigma. \tag{П.36}$$

Действительно, выражая из (П.35)  $\gamma$ , получаем, что  $\gamma = \sqrt{3 \ln \frac{N}{\delta}}$ . После подстановки данного выражения в (П.34) и выбора единого  $\sigma \in \mathbf{R}_+$  вместо последовательности  $\sigma_t, t = 0, \dots, N - 1$ , получаем оценку (П.36).

Объединяя полученные неравенства, получаем, что с вероятностью, большей или равной  $1 - \delta$ , неравенство

$$\frac{1}{2} R_l^2 \leq A + \frac{ud}{2} \sum_{t=0}^{l-1} \tilde{R}_t^2 + 12ud \ln \frac{N}{\delta} \sigma^2 l$$

выполняется для всех  $l = 1, \dots, N$  одновременно. Заметим, что последнее слагаемое в полученной оценке – неубывающая функция от  $l$ . Определим  $\hat{l}$  как наибольшее целое число, для которого выполнено  $\hat{l} \leq l$  и  $\tilde{R}_{\hat{l}} = \tilde{R}_l = \tilde{R}_{\hat{l}+1} = \dots = \tilde{R}_l$ , и, следовательно, с вероятностью  $\geq 1 - \delta$  имеем

$$\frac{1}{2} \tilde{R}_l^2 \leq A + \frac{ud}{2} \sum_{t=0}^{\hat{l}-1} \tilde{R}_t^2 + 12ud \ln \frac{N}{\delta} \sigma^2 \hat{l} \leq A + \frac{ud}{2} \sum_{t=0}^{l-1} \tilde{R}_t^2 + 12ud \ln \frac{N}{\delta} \sigma^2 l \quad \forall l = 1, \dots, N.$$

Получаем, что с вероятностью  $\geq 1 - \delta$  верна следующая оценка:

$$\begin{aligned} \tilde{R}_l^2 &\leq 2A + ud \sum_{t=0}^{l-1} \tilde{R}_k^2 + 24ud \ln \frac{N}{\delta} \sigma^2 l \leq 2A \underbrace{(1 + ud)}_{\leq 2ud} + \underbrace{(ud + u^2 d^2)}_{\leq 2u^2 d^2} \sum_{t=0}^{l-2} \tilde{R}_t^2 + \\ &+ 24ud \ln \frac{N}{\delta} \sigma^2 \underbrace{(l + ud(l-1))}_{\leq 2udl} \leq 2ud \left( 2A + ud \sum_{t=0}^{l-2} \tilde{R}_t^2 + 24ud \ln \frac{N}{\delta} \sigma^2 l \right) \quad \forall l = 1, \dots, N. \end{aligned}$$

Применяя данную оценку рекурсивно, получаем, что с вероятностью  $\geq 1 - \delta$  верно

$$\tilde{R}_\delta^2 \leq (2ud)^l \left( 2A + ud \tilde{R}_0^2 + 24ud \ln \frac{N}{\delta} \sigma^2 l \right).$$

Далее рассмотрим последовательность случайных величин  $\xi^t = \langle \boldsymbol{\eta}^t, \mathbf{a}^t \rangle$ . Заметим, что  $\mathbb{E}[\xi^t | \xi^0, \dots, \xi^{t-1}] = \langle \mathbb{E}[\boldsymbol{\eta}^t | \boldsymbol{\eta}^0, \dots, \boldsymbol{\eta}^{t-1}], \mathbf{a}^t \rangle = 0$ , тогда, используя неравенство Коши–Буняковского, получаем,

$$\begin{aligned} \mathbb{E} \left[ \exp \left( \frac{(\xi^t)^2}{\sigma^2 d^2 \tilde{R}_t^2} \right) \middle| \xi^0, \dots, \xi^{t-1} \right] &\leq \mathbb{E} \left[ \exp \left( \frac{\|\boldsymbol{\eta}^t\|_2^2 d^2 \tilde{R}_t^2}{\sigma^2 d^2 \tilde{R}_t^2} \right) \middle| \boldsymbol{\eta}^0, \dots, \boldsymbol{\eta}^{t-1} \right] = \\ &= \mathbb{E} \left[ \exp \left( \frac{\|\boldsymbol{\eta}^t\|_2^2}{\sigma^2} \right) \middle| \boldsymbol{\eta}^0, \dots, \boldsymbol{\eta}^{t-1} \right] \leq \exp(1). \end{aligned}$$

Определим  $\hat{\sigma}_t^2 = \sigma^2 d^2 \tilde{R}_t^2$ , тогда с вероятностью  $\geq 1 - \delta$  выполняется

$$\begin{aligned} \sum_{t=0}^{l-1} \hat{\sigma}_t^2 &\leq \sigma^2 d^2 l (2ud)^l \left( 2A + ud \tilde{R}_0^2 + 24ud \ln \frac{N}{\delta} \sigma^2 l \right) \leq \\ &\leq \sigma^2 d^2 N (2ud)^N \left( 2A + ud \tilde{R}_0^2 + 24ud \ln \frac{N}{\delta} \sigma^2 N \right) := \frac{F}{2} \end{aligned}$$

для всех  $l = 1, \dots, N$  одновременно, где

$$F = 2\sigma^2 d^2 N (2ud)^N \left( 2A + ud\tilde{R}_0^2 + 24ud \ln \frac{N}{\delta} \sigma^2 N \right).$$

Используя следствие 8 из [34] для  $b = \hat{\sigma}_0^2$ , получаем для любого  $l = 1, \dots, N$  с вероятностью  $\geq 1 - \frac{\delta}{N}$  следующую оценку:

$$\text{либо } \sum_{t=0}^{l-1} \hat{\sigma}_t^2 \geq F, \quad \text{либо } \left| \sum_{t=0}^{l-1} \xi^t \right| \leq C_1 \sqrt{\sum_{t=0}^{l-1} \hat{\sigma}_t^2 \left( \ln \left( \frac{N}{\delta} \right) + \ln \ln \left( \frac{F}{f} \right) \right)}, \quad (\text{П.37})$$

где  $C_1 > 0$  – константа, которая не зависит от  $F$  и  $f$ .

Далее, объединяя полученные оценки, получаем, что оценка (П.37) с вероятностью  $\geq 1 - \delta$  верна для всех  $l = 1, \dots, N$  одновременно.

Учитывая выбор  $F$ , получаем, что с вероятностью  $\geq 1 - 2\delta$

$$\left| \sum_{t=0}^{l-1} \xi^t \right| \leq C_1 \sqrt{\sum_{t=0}^{l-1} \hat{\sigma}_t^2 \left( \ln \left( \frac{N}{\delta} \right) + \ln \ln \left( \frac{F}{f} \right) \right)}$$

для всех  $l = 1, \dots, N$  одновременно.

Для удобства дальнейших рассуждений обозначим  $g(N) := \ln \left( \frac{N}{\delta} \right) + \ln \ln \left( \frac{F}{f} \right) \approx \ln \left( \frac{N}{\delta} \right)$ , пренебрегая константой. Используя  $\hat{\sigma}_t^2 = \sigma^2 d^2 \tilde{R}_t^2$ , получаем, что с вероятностью  $\geq 1 - 2\delta$  справедлива следующая оценка:

$$\frac{1}{2} \tilde{R}_l^2 \leq A + u \sum_{t=0}^{l-1} \underbrace{\langle \boldsymbol{\eta}^t, \mathbf{a}^t \rangle}_{\xi^t} \leq A + udD \sqrt{\sigma^2 g(N)} \sqrt{\sum_{t=0}^{l-1} \tilde{R}_t^2} \quad (\text{П.38})$$

для всех  $l = 1, \dots, N$  одновременно. Выбирая в качестве  $A = \frac{A}{\tilde{R}_0^2}$ ,  $B = \frac{1}{\tilde{R}_0} udC_1 \sqrt{\sigma^2 g(N)}$ ,  $r_t = \tilde{R}_t$ , из леммы 10 получаем, что с вероятностью  $1 - 2\delta$  выполняется

$$\tilde{R}_l \leq J R_0$$

для всех  $l = 1, \dots, N$  одновременно, где

$$J = \max \left\{ 1, \frac{1}{R_0} udC_1 \sqrt{\sigma^2 g(N)} + \sqrt{\frac{1}{\tilde{R}_0^2} u^2 d^2 C_1^2 \sigma^2 g(N) + \frac{2A}{R_0^2}} \right\}.$$

Отсюда получаем, что с вероятностью  $1 - 2\delta$  оценка

$$A + u \sum_{t=0}^{l-1} \langle \boldsymbol{\eta}^t, \mathbf{a}^t \rangle \leq A + udC_1 \sqrt{\sigma^2 g(N)} J \tilde{R}_0^2 \leq A + udC_1 \sqrt{\sigma^2 g(N)} N J \tilde{R}_0^2$$

верна для всех  $l = 1, \dots, N$  одновременно.

**Доказательство леммы 2.** Для  $\boldsymbol{\lambda} \in \mathbf{R}^m$

$$\|\boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}\|_2^2 \|\boldsymbol{\lambda}^t - \beta \nabla \varphi(\boldsymbol{\lambda}^t, \boldsymbol{\xi}^t)\|_+ - \boldsymbol{\lambda}\|_2^2 \leq \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}\|_2^2 - 2\beta \langle \nabla \varphi(\boldsymbol{\lambda}^t, \boldsymbol{\xi}^t), \boldsymbol{\lambda}^t - \boldsymbol{\lambda} \rangle + \beta^2 \|\nabla \varphi(\boldsymbol{\lambda}^t, \boldsymbol{\xi}^t)\|_2^2,$$

т.е.

$$0 \leq \frac{1}{2\beta} \left( \|\boldsymbol{\lambda}^t - \boldsymbol{\lambda}\|_2^2 - \|\boldsymbol{\lambda}^{t+1} - \boldsymbol{\lambda}\|_2^2 \right) + \langle \nabla \varphi(\boldsymbol{\lambda}^t, \boldsymbol{\xi}^t), \boldsymbol{\lambda} - \boldsymbol{\lambda}^t \rangle + \frac{\beta}{2} \|\nabla \varphi(\boldsymbol{\lambda}^t, \boldsymbol{\xi}^t)\|_2^2. \quad (\text{П.39})$$

После прибавления к обеим сторонам неравенства (П.39)  $\varphi(\lambda^t)$ , умножения на  $N$  и суммирования от 0 до  $N - 1$ :

$$\frac{1}{N} \sum_{t=0}^{N-1} \varphi(\lambda^t) \leq \frac{1}{N} \sum_{t=0}^{N-1} \left\{ \varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle + \frac{\beta}{2} \|\nabla \varphi(\lambda^t, \xi^t)\|_2^2 + \frac{1}{2\beta} (\|\lambda^t - \lambda\|_2^2 - \|\lambda^{t+1} - \lambda\|_2^2) \right\}. \quad (\text{П.40})$$

В силу выпуклости  $\varphi(\lambda)$  для  $\hat{\lambda}^N = \frac{1}{N} \sum_{t=0}^{N-1} \lambda^t$  получаем, что

$$N\varphi(\hat{\lambda}^N) \leq \sum_{t=0}^{N-1} \left\{ \varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle \right\} + \frac{\beta}{2} \|\nabla \varphi(\lambda^t, \xi^t)\|_2^2 + \frac{1}{2\beta} (\|\lambda^0 - \lambda\|_2^2 - \|\lambda^N - \lambda\|_2^2). \quad (\text{П.41})$$

Выбираем  $\lambda = \lambda^*$  и прибавляем и вычитаем справа  $\sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle$ . Получаем

$$N\varphi(\hat{\lambda}^N) \leq \sum_{t=0}^{N-1} \left\{ \varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle \right\} + \frac{\beta}{2} \|\nabla \varphi(\lambda^t, \xi^t)\|_2^2 + \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle + \frac{1}{2\beta} (\|\lambda^0 - \lambda^*\|_2^2 - \|\lambda^N - \lambda^*\|_2^2). \quad (\text{П.42})$$

Из выпуклости  $\varphi(\lambda)$  имеем

$$\sum_{t=0}^{N-1} \left\{ \varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle \right\} \leq \sum_{t=0}^{N-1} \left\{ \varphi(\lambda^t) + \varphi(\lambda^*) - \varphi(\lambda^t) \right\} \leq \sum_{t=0}^{N-1} \varphi(\lambda^*) \leq N\varphi(\lambda^*).$$

Подставляя полученную оценку в (П.42), получаем

$$\frac{1}{2\beta} \|\lambda^N - \lambda^*\|_2^2 \leq \frac{1}{2\beta} \|\lambda^0 - \lambda^*\|_2^2 + \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle + \frac{\beta}{2} \|\nabla \varphi(\lambda^t, \xi^t)\|_2^2. \quad (\text{П.43})$$

Определим  $R_t = \|\lambda^t - \lambda^*\|_2$  и  $\tilde{R}_t = \max\{\tilde{R}_{t-1}, R_t\}$ , причем  $R_0 = \tilde{R}_0$  и, так как  $\lambda^0 = \mathbf{0}$  и  $\|\lambda^*\|_2 \leq R$ , то  $R_0 = R$ . При этом по построению получаем, что  $\lambda^t \in B_{\tilde{R}_t}(\lambda^*)$ . Так же определим  $\|\mathbf{a}^t\|_2 = \|\lambda^t - \lambda^*\|_2 \leq \tilde{R}_t$ . Тогда (П.43) можно переписать в следующем виде:

$$\frac{1}{2\beta} \tilde{R}_N^2 \leq \frac{1}{2\beta} \tilde{R}_0^2 + \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \mathbf{a}^t \rangle + \frac{\beta}{2} \|\nabla \varphi(\lambda^t, \xi^t)\|_2^2.$$

Обозначим  $\eta^t = \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t)$ . По теореме 2.1 из [35] имеем

$$\mathbb{P} \left\{ \left\| \sum_{t=0}^{N-1} \eta^t \right\|_2 \geq (\sqrt{2} + \sqrt{2\gamma}) \sqrt{\sum_{t=0}^{N-1} \sigma_t^2 |\xi^t|^{N-1}} \right\} \leq \exp\left(-\frac{\gamma^2}{3}\right). \quad (\text{П.44})$$

Используя лемму 2 из [34], получаем, что

$$\mathbb{E} \left[ \exp\left(\frac{\|\eta^t\|_2^2}{\sigma^2}\right) \mid \{\xi^k\}_{k=0}^{t-1} \right] \leq \exp(1),$$

при этом  $\eta^t$  зависит только от  $\xi^{t-1}, \dots, \xi^0$ . Используя новые обозначения и (8), имеем

$$\tilde{R}_N^2 \leq \tilde{R}_0^2 + 2\beta \sum_{t=0}^{N-1} \langle \eta^t, \mathbf{a}^t \rangle + \beta^2 M^2.$$

Тогда из леммы 11 с константами  $A = \tilde{R}_0^2 + \beta^2 M^2$ ,  $d = 1$  и  $u = \beta$ , получаем, что с вероятностью  $1 - 2\delta$ , где  $\frac{\delta}{N} = \exp\left(-\frac{\gamma^2}{3}\right)$ , верна следующая оценка:

$$\tilde{R}_l \leq JR_0 \quad \text{и} \quad \sum_{t=0}^{l-1} \langle \eta^t, \mathbf{a}^t \rangle \leq D\sqrt{\sigma^2 g(N)NJ\tilde{R}_0^2} \quad (\text{П.45})$$

$\forall l = 1, \dots, N$  одновременно, где  $D$  – положительная константа,

$$F = 2\sigma^2 N(2\beta)^N \left(2A + \beta\tilde{R}_0^2 + 24 \ln \frac{N}{\delta} \beta\sigma^2 N\right),$$

$f = \sigma^2 \tilde{R}_0^2$ ,  $g(N) = \ln\left(\frac{N}{\delta}\right) + \ln\ln\left(\frac{F}{f}\right)$  и

$$J = \max \left\{ 1, \frac{1}{R_0} \beta C_1 \sqrt{\sigma^2 g(N)} + \sqrt{\frac{1}{R_0^2} \beta^2 C_1^2 \sigma^2 g(N) + \frac{2A}{R_0^2}} \right\}.$$

Чтобы оценить зазор двойственности, используем (П.41), также отметим, что данная оценка верна для любого  $\lambda \in \mathbf{R}_+^m$ . Поэтому, беря минимум по всем  $\lambda$  из множества  $\Lambda_{2R} = \{\lambda \in \mathbf{R}_+^m : \|\lambda\|_2 \leq 2R\}$ , получаем

$$N\varphi(\hat{\lambda}^N) \leq \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle) + \frac{1}{2\beta} \|\lambda^0 - \lambda\|_2^2 \right\} + \frac{N\beta M^2}{2},$$

где для оценки последнего слагаемого использовалось предположение (8). Также учитывалось, что  $\|\lambda^N - \lambda\|_2^2 \geq 0$ . В силу (П.29) получаем следующую оценку:

$$\varphi(\hat{\lambda}^N) \leq \frac{1}{N} \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle) \right\} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2}.$$

Прибавим и вычтем из выражения под минимумом  $\sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle$ . Тогда получаем

$$\begin{aligned} \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle) \right\} &\leq \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle) \right\} + \\ &+ \max_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda \rangle \right\} + \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), -\lambda^t \rangle. \end{aligned}$$

Заметим, что  $-\lambda^* \in \Lambda_{2R}$ . Тогда имеем

$$\begin{aligned} \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), -\lambda^t \rangle &= \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle + \\ &+ \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), -\lambda^* \rangle \leq \max_{\lambda \in \Lambda_{2R}} \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda \rangle + \\ &+ \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle. \end{aligned}$$

Отсюда получаем следующую оценку:

$$\begin{aligned} \varphi(\hat{\lambda}^N) &\leq \frac{1}{N} \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t, \xi^t), \lambda - \lambda^t \rangle) \right\} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2} \leq \\ &\leq \frac{1}{N} \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle) \right\} + \frac{1}{N} \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle + \\ &+ \frac{2}{N} \max_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda \rangle \right\} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2}. \end{aligned}$$

Из определения нормы получаем, что

$$\max_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda \rangle \right\} \leq 2R \left\| \sum_{t=0}^{N-1} (\nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t)) \right\|_2.$$

Используя (П.44), получаем, что с вероятностью  $1 - \delta$  выполняется

$$\left\| \sum_{t=0}^{N-1} (\nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t)) \right\|_2 \leq \sigma \sqrt{2N} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right). \quad (\text{П.47})$$

Подставим в выражение  $\sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle)$  из (П.46) значения  $\varphi(\lambda^t)$  и  $\nabla \varphi(\lambda^t)$  и получим

$$\begin{aligned} \sum_{t=0}^{N-1} \left( \langle \lambda^t, \mathbf{b} \rangle + \sum_{k=1}^n (u_k(x_k(\lambda^t)) - \langle \lambda^t, \mathbf{C}_k x_k(\lambda^t) \rangle + \langle \mathbf{b} - \mathbf{C}x^t(\lambda^t), \lambda - \lambda^t \rangle) \right) = \\ = \sum_{t=0}^{N-1} \left( \sum_{k=1}^n (u_k(x_k(\lambda^t)) + \langle \mathbf{b} - \mathbf{C}x^t(\lambda^t), \lambda \rangle) \right). \end{aligned}$$

Тогда в силу вогнутости функций  $u_k(x_k)$

$$\frac{1}{N} \min_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} (\varphi(\lambda^t) + \langle \nabla \varphi(\lambda^t), \lambda - \lambda^t \rangle) \right\} \leq U(\hat{\mathbf{x}}^N) - \frac{1}{N} \max_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} \langle \mathbf{C}x^t(\lambda^t) - \mathbf{b}, \lambda \rangle \right\}.$$

Учитывая последнее неравенство, из (П.46) получаем

$$\begin{aligned} \varphi(\hat{\lambda}^N) \leq U(\hat{\mathbf{x}}^N) - \frac{1}{N} \max_{\lambda \in \Lambda_{2R}} \left\{ \sum_{t=0}^{N-1} \langle \mathbf{C}x^t(\lambda^t) - \mathbf{b}, \lambda \rangle \right\} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2} + \\ + \frac{2R}{N} \left\| \sum_{t=0}^{N-1} (\nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t)) \right\|_2 + \frac{1}{N} \sum_{t=0}^{N-1} \langle \nabla \varphi(\lambda^t, \xi^t) - \nabla \varphi(\lambda^t), \lambda^* - \lambda^t \rangle. \end{aligned}$$

Отсюда, учитывая оценку (П.47) и результат (П.44), получаем, что с вероятностью  $1 - 3\delta$

$$\varphi(\hat{\lambda}^N) - U(\hat{\mathbf{x}}^N) + 2R \left\| \mathbf{C}\hat{\mathbf{x}}^N - \mathbf{b} \right\|_2 \leq \frac{2R\sigma\sqrt{2} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right)}{\sqrt{N}} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2} + C_1 \frac{\sigma\sqrt{g(N)JR^2}}{\sqrt{N}}. \quad (\text{П.48})$$

По теореме 2.1 из [35] для всех  $\gamma > 0$  имеем

$$P \left\{ \left\| \sum_{t=0}^{N-1} (\mathbf{x}(\lambda^t, \xi^t) - \mathbf{x}(\lambda^t)) \right\|_2 \geq (\sqrt{2} + \sqrt{2}\gamma) \sqrt{\sum_{t=0}^{N-1} \sigma_x^2 \{ \xi^t \}_{t=0}^{N-1}} \right\} \leq \exp \left( -\frac{\gamma^2}{3} \right).$$

Выбирая  $\gamma = \sqrt{3 \ln \frac{1}{\delta}}$ , получаем, что с вероятностью  $1 - \delta$

$$\left\| \tilde{\mathbf{x}}^N - \hat{\mathbf{x}}^N \right\|_2 = \frac{1}{N} \left\| \sum_{t=0}^{N-1} (\mathbf{x}(\lambda^t, \xi^t) - \mathbf{x}(\lambda^t)) \right\|_2 \leq \sigma_x \sqrt{\frac{2}{N}} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right).$$

Тогда с вероятностью  $1 - \delta$  выполняется следующее неравенство:

$$\left\| \mathbf{C}\tilde{\mathbf{x}}^N - \mathbf{C}\hat{\mathbf{x}}^N \right\|_2 \leq \|\mathbf{C}\|_2 \cdot \left\| \tilde{\mathbf{x}}^N - \hat{\mathbf{x}}^N \right\|_2 \leq \sigma_x \sqrt{\frac{2\lambda_{\max}(\mathbf{C}^T \mathbf{C})}{N}} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right).$$

Заметим, что верно следующее:

$$\begin{aligned} 2R\|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+ &= \max_{\lambda \in \Lambda_{2R}} \left\{ \langle C\tilde{\mathbf{x}}^N - \mathbf{b}, \lambda \rangle + \langle C\hat{\mathbf{x}}^N - C\tilde{\mathbf{x}}^N - \mathbf{b} + \mathbf{b}, \lambda \rangle \right\} \leq \\ &\leq \max_{\lambda \in \Lambda_{2R}} \left\{ \langle C\hat{\mathbf{x}}^N - \mathbf{b}, \lambda \rangle \right\} + \max_{\lambda \in \Lambda_{2R}} \left\{ \langle C\tilde{\mathbf{x}}^N - C\hat{\mathbf{x}}^N, \lambda \rangle \right\} \leq 2R\|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+ + 2R\|C\tilde{\mathbf{x}}^N - C\hat{\mathbf{x}}^N\|_2 \leq \\ &\leq 2R\|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+ + 2R\sigma_x \sqrt{\frac{2\lambda_{\max}(C^T C)}{N}} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right). \end{aligned} \quad (\text{П.49})$$

В силу липшицевости функции  $U$  получаем

$$|U(\tilde{\mathbf{x}}^N) - U(\hat{\mathbf{x}}^N)| \leq M_U \|\tilde{\mathbf{x}}^N - \hat{\mathbf{x}}^N\|_2 \leq M_U \sigma_x \sqrt{\frac{2}{N}} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right).$$

Тогда имеем

$$U(\hat{\mathbf{x}}^N) = U(\tilde{\mathbf{x}}^N) + (U(\hat{\mathbf{x}}^N) - U(\tilde{\mathbf{x}}^N)) \geq U(\tilde{\mathbf{x}}^N) - M_U \sigma_x \sqrt{\frac{2}{N}} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right). \quad (\text{П.50})$$

Подставляя (П.49) и (П.50) в (П.48), получаем, что с вероятностью  $1 - 4\delta$  выполняется

$$\begin{aligned} \varphi(\hat{\lambda}^N) - U(\tilde{\mathbf{x}}^N) + 2R\|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+ &\leq C_1 \frac{\sigma \sqrt{g(N)} J R^2}{\sqrt{N}} + \frac{2R^2}{\beta N} + \frac{\beta M^2}{2} + \\ &+ \frac{\sqrt{2} \left( 1 + \sqrt{3 \ln \frac{1}{\delta}} \right)}{\sqrt{N}} \left( M_U \sigma_x + 2R \left( \sigma + \sigma_x \sqrt{\lambda_{\max}(C^T C)} \right) \right). \end{aligned}$$

**Доказательство теоремы 3.** Так как  $\|\nabla \varphi(\lambda)\|_2 \leq M$  для любых  $\lambda \in \Lambda_{2R}$  (см. (5)), то справедлива следующая оценка:

$$\sup_{\lambda^1, \lambda^2 \in \Lambda_{2R}} \langle \nabla \varphi(\lambda^1), \lambda^2 - \lambda^1 \rangle \leq M \cdot 4R.$$

Из теоремы 4.1 [26] получаем

$$\max_{\lambda \in \Lambda_{2R}} \sum_{t=1}^N \xi^t \langle \nabla \varphi(\lambda^t), \lambda^t - \lambda \rangle \leq \varepsilon_N,$$

где  $\varepsilon_N = 32 \times 4MR \exp\left\{-\frac{N}{2m(m+1)}\right\}$ . Тогда имеем

$$\forall \lambda \in \Lambda_{2R} \sum_{t \in I_N} \xi^t \langle \nabla \varphi(\lambda^t), \lambda^t - \lambda \rangle \leq \sum_{t=1}^N \xi^t \langle \nabla \varphi(\lambda^t), \lambda^t - \lambda \rangle \leq \varepsilon_N.$$

Отсюда получаем, что верна следующая оценка:

$$\sum_{t \in I_N} \xi^t \langle \mathbf{b} - C\mathbf{x}^t, \lambda^t \rangle + \max_{\lambda \in \Lambda_R} \left\langle -\sum_{t \in I_N} \xi^t (\mathbf{b} - C\mathbf{x}^t), \lambda \right\rangle \leq \varepsilon_N,$$

которую можно переписать в следующем виде:

$$\sum_{t \in I_N} \xi^t \langle \mathbf{b} - C\mathbf{x}^t, \lambda^t \rangle \leq \varepsilon_N - 2R\|C\tilde{\mathbf{x}}^N - \mathbf{b}\|_+. \quad (\text{П.51})$$

Далее, в силу (3), для каждого  $\mathbf{x} \geq 0$  и  $t \in I_N$  выполнено

$$U(\mathbf{x}^t(\lambda^t)) - \langle C\mathbf{x}^t(\lambda^t) - \mathbf{b}, \lambda^t \rangle \geq U(\mathbf{x}) - \langle C\mathbf{x} - \mathbf{b}, \lambda^t \rangle.$$

Умножая  $t$ -е неравенство на  $\xi^t$ , суммируя по всем индексам из  $I_N$  и учитывая, что  $\sum_{t \in I_N} \xi^t U(\mathbf{x}^t) \leq U(\hat{\mathbf{x}}^N)$ , в силу вогнутости функций  $u_k(x_k)$ ,  $k = 1, \dots, N$ , получаем

$$U(\mathbf{x}) - U(\hat{\mathbf{x}}^N) + \langle \mathbf{b} - C\mathbf{x}, \hat{\lambda}^N \rangle \leq \sum_{t \in I_N} \xi^t \langle \mathbf{b} - C\mathbf{x}^t, \lambda^t \rangle,$$

где  $\hat{\lambda}^N = \sum_{t \in I_N} \xi^t \lambda^t$ . Используя оценку (П.51), получаем

$$2R \left\| C\hat{\mathbf{x}}^N - \mathbf{b} \right\|_+ + U(\mathbf{x}^*) - U(\hat{\mathbf{x}}^N) + \langle \mathbf{b} - C\mathbf{x}^*, \hat{\lambda}^N \rangle \leq \varepsilon_N. \quad (\text{П.52})$$

Поскольку  $\hat{\lambda}^N \in \Lambda_{2R}$ , и, следовательно,  $\hat{\lambda}^N \geq 0$ , откуда  $\langle \mathbf{b} - C\mathbf{x}^*, \hat{\lambda}^N \rangle \geq 0$ , из (П.51) следует  $U(\mathbf{x}^*) - U(\hat{\mathbf{x}}^N) \leq \varepsilon_N$ . Далее, так как для всех  $\mathbf{x} \geq 0$ , в силу определения  $\lambda^*$ , выполняется  $U(\mathbf{x}^*) \geq U(\mathbf{x}) - \langle \lambda^*, C\mathbf{x} - \mathbf{b} \rangle$ , получаем

$$\begin{aligned} U(\hat{\mathbf{x}}^N) &\leq U(\mathbf{x}^*) - \langle \lambda^*, \mathbf{b} - C\hat{\mathbf{x}}^N \rangle \leq U(\mathbf{x}^*) - \min_{\lambda \in \Lambda_R} \{ \langle \lambda, \mathbf{b} - C\hat{\mathbf{x}}^N \rangle \} = \\ &= U(\mathbf{x}^*) + \max_{\lambda \in \Lambda_R} \{ \langle \lambda, C\hat{\mathbf{x}}^N - \mathbf{b} \rangle \} \leq U(\mathbf{x}^*) + R \left\| C\hat{\mathbf{x}}^N - \mathbf{b} \right\|_+. \end{aligned}$$

Отсюда вместе с (П.52) получаем  $R \left\| C\hat{\mathbf{x}}^N - \mathbf{b} \right\|_+ \leq \varepsilon_N$ . Оценка (11) на количество итераций метода следует из следующей выкладки:

$$\varepsilon_N = 32 \times 4MR \exp \left\{ -\frac{N}{2m(m+1)} \right\} \leq \varepsilon \Rightarrow -\frac{N}{2m(m+1)} \leq \ln \left( \frac{\varepsilon}{32 \times 4MR} \right) \Rightarrow N \geq 2m(m+1) \ln \left( \frac{32 \times 4MR}{\varepsilon} \right).$$

### СПИСОК ЛИТЕРАТУРЫ

1. Kelly F.P., Maulloo A.K., Tan D.K.H. Rate control for communication networks: shadow prices, proportional fairness and stability // J. of the Operational Research Society. 1998. V. 49. 3. P. 237–252.
2. Рохлин Д.Б. Распределение ресурсов в сетях связи с большим числом пользователей: стохастический метод градиентного спуска // Теория вероятностей и ее применения (в печати). 2019.
3. Arrow K.J., Hurwicz L. Decentralization and computation in resource allocation. Stanford University, Department of Economics, 1958.
4. Kakhbod A. Resource allocation in decentralized systems with strategic agents: an implementation theory approach. Springer Science & BusinessMedia, 2013.
5. Campbell D.E. Resource allocation mechanisms. Cambridge University Press, 1987.
6. Friedman E.J., Oren S.S. The complexity of resource allocation and price mechanisms under bounded rationality // Economic Theory. 1995. V. 6. 2. P. 225–250.
7. Nesterov Yu., Shikhman V. Dual subgradient method with averaging for optimal resource allocation // European Journal of Operational Research. 2018. V. 270. 3. P. 907–916.
8. Ivanova A., Dvurechensky P., Gasnikov A., Kamzolov D. Composite optimization for the resource allocation problem // arXiv preprint arXiv:1810.00595. 2018.
9. Нестеров Ю.Е. Метод минимизации выпуклых функций со скоростью сходимости  $O(1/k^2)$  // Докл. АН СССР. 1983. Т. 269. 39. С. 543–547.
10. Gasnikov A.V., Gasnikova E.V., Nesterov Yu.E., Chernov A.V. Efficient numerical methods for entropy-linear programming problems // Comput. Math. and Math. Phys. 2016. V. 56. 4. P. 514–524.
11. Chernov A., Dvurechensky P., Gasnikov A. Fast Primal-Dual Gradient Method for Strongly Convex Minimization Problems with Linear Constraints // Discrete Optimization and Operations Research: 9th International Conference, DOOR 2016, Vladivostok, Russia, September 19–23, 2016 Proceedings. Springer International Publishing, 2016. P. 391–403.
12. Dvurechensky P., Gasnikov A., Gasnikova E., Matsievsky S., Rodomanov A., Usik I. Primal-Dual Method for Searching Equilibrium in Hierarchical Congestion Population Games // Supplementary Proceedings of the 9th International Conference on Discrete Optimization and Operations Research and Scientific School (DOOR 2016) Vladivostok, Russia, September 19–23, 2016. P. 584–595. arXiv:1606.08988.
13. Anikin A., Gasnikov A., Turin A., Chernov A. Dual approaches to the minimization of strongly convex functionals with asimple structure under affine constraints // Comput. Math. and Math. Phys. 2017. V. 57. № 8. P. 1262–1276.
14. Dvurechensky P., Gasnikov A., Kroshnin A. Computational Optimal Transport: Complexity by Accelerated Gradient Descent Is Better Than by Sinkhorn’s Algorithm // Proceedings of the 35th International Conference on Machine Learning. 2018. V. 80. P. 1367–1376. arXiv:1802.04367.
15. Nesterov Yu., Gasnikov A., Guminov S., Dvurechensky P. Primal-dual accelerated gradient methods with small-dimensional relaxation oracle // arXiv:1809.05895. 2018.
16. Guminov S., Dvurechensky P., Gasnikov A. On Accelerated Alternating Minimization // arXiv:1906.03622. 2019.



17. *Guminov S.V., Nesterov Yu.E., Dvurechensky P.E., Gasnikov A.V.* Accelerated Primal-Dual Gradient Descent with Linesearch for Convex, Nonconvex, and Nonsmooth Optimization Problems // *Doklady Mathematics*. 2019. V. 99. P. 125–128.
18. *Kroshnin A., Tupitsa N., Dvinskikh D., Dvurechensky P., Gasnikov A., Uribe C.A.* On the Complexity of Approximating Wasserstein Barycenters // *Proceedings of the 36th International Conference on Machine Learning*. 2019. V. 97. Eds. K. Chaudhuri, R. Salakhutdinov. California, USA: PMLR, 2019. P. 3530–3540. arXiv:1901.08686.
19. *Uribe C.A., Dvinskikh D., Dvurechensky P., Gasnikov A., Nedich A.* Distributed Computation of Wasserstein Barycenters Over Networks // *2018 IEEE Conference on Decision and Control (CDC)*. 2018. P. 6544–6549. arXiv:1803.02933.
20. *Dvinskikh D., Gorbunov E., Gasnikov A., Dvurechensky P., Uribe C.A.* On Primal and Dual Approaches for Distributed Stochastic Convex Optimization over Networks // *2019 IEEE Conference on Decision and Control (CDC)*. 2019. arXiv:1903.09844.
21. *Dvurechensky P., Dvinskikh D., Gasnikov A., Uribe C.A., Nedic A.* Decentralize and Randomize: Faster Algorithm for Wasserstein Barycenters // *Advances in Neural Information Processing Systems*. 2018. V. 31. P. 10783–10793. arXiv:1806.03915.
22. *Данскин Д.М.* Теория максимина. М.: Советское радио, 1970.
23. *Демьянов В.Ф., Малоземов В. Н.* Введение в минимакс. М.: Наука, 1972.
24. *Nesterov Yu.* Smooth minimization of non-smooth functions // *Math. Programming*. 2005. V. 103. P. 127–152.
25. *Юдин Д.Б., Немировский А.С.* Информационная сложность и эффективные методы решения выпуклых экстремальных задач // *Экономика и матем. методы*. 1976. 2. С. 357–369.
26. *Nemirovski A., Onn S., Rothblum U.G.* Accuracy certificates for computational problems with convex structure // *Math. of Operations Research*. 2010. V. 35. 1. P. 52–78.
27. *Lan G., Zhou Y.* Random gradient extrapolation for distributed and stochastic optimization // *SIAM Journal on Optimization*. 2018. V. 28. 4. P. 2753–2782.
28. *Bubeck S.* Convex optimization: Algorithms and complexity // *Foundations and Trends in Machine Learning*. 2015. V. 8. № 3/4. P. 231–357.
29. *Nesterov Yu.* Implementable tensor methods in unconstrained convex optimization: tech. rep. Universite catholique de Louvain, Center for Operations Research and Econometrics (CORE). 2018.
30. *Gasnikov A., Dvurechensky P., Gorbunov E., Vorontsova E., Selikhanovych D., Uribe C.A., Jiang B., Wang H., Zhang S., Bubeck S., Jiang Q., Lee Y.T., Li Y., Sidford A.* Near Optimal Methods for Minimizing Convex Functions with Lipschitz  $p$ -th Derivatives // *Proceedings of the Thirty-Second Conference on Learning Theory*. 2019. V. 99. P. 1392–1393. URL: <http://proceedings.mlr.press/v99/gasnikov19b.html>. arXiv:1809.00382.
31. *Zhou K., Shang F., Cheng J.* A simple stochastic variance reduced algorithm with fast convergence rates // *arXiv preprint arXiv:1806.11027*. 2018.
32. *Zhou K.* Direct acceleration of SAGA using sampled negative momentum // *arXiv preprint arXiv:1806.11048*. 2018.
33. *Niu F., Recht B., Re C., Wright S.J.* Hogwild: A lock-free approach to parallelizing stochastic gradient descent // *Advances in Neural Information Processing Systems*. 2011. P. 693–701.
34. *Jin C., Netrapalli P., Ge R., Kakade S.M., Jordan M.I.* A short note on concentration inequalities for random vectors with subgaussian norm // *arXiv preprint arXiv:1902.03736*. 2019.
35. *Juditsky A., Nemirovski A.* Large deviations of vector-valued martin-gales in 2-smooth normed spaces: tech. rep. HAL: hal-00318071. 2008. URL: <http://hal.archives-ouvertes.fr/hal-00318071/>. arXiv:0809.0813.
36. *Nesterov Yu.* *Lectures on Convex Optimization*. 2nd ed. Springer, 2018.
37. *Нестеров Ю.Е.* Алгоритмическая выпуклая оптимизация: дисс. д.ф.-м.н.: 01.01.07. М.: МФТИ, 2013.
38. *Beck A.* *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*. SIAM, Philadelphia, 2014.