

# Распределенная оптимизация в условиях Поляка-Лоясевича\*

И. О. Автор<sup>1</sup>, И. О. Соавтор<sup>2</sup>, И. О. Фамилия<sup>1,2</sup>  
author@site.ru; co-author@site.ru; co-author@site.ru

<sup>1</sup>Организация, адрес; <sup>2</sup>Организация, адрес

В статье рассматривается новый метод децентрализованного распределенного решения больших систем нелинейных уравнений в условиях Поляка-Лоясевича. Суть метода состоит в постановке эквивалентной задачи распределенной оптимизации и последующем ее сведении сперва к задаче ограниченной оптимизации, а затем к задаче композитной оптимизации, но уже без ограничений. Предложенный метод сравнивается с градиентным спуском, ускоренным градиентным спуском, а также с последовательным и параллельным алгоритмом обратного распространения ошибки при обучении многослойной нейронной сети с нелинейной функцией активации нейрона.

**Ключевые слова:** *большие нелинейные системы; распределенная оптимизация; условия Поляка-Лоясевича; многослойные нейронные сети*

**DOI:** 10.21469/22233792

## 1 Введение

Во многих современных прикладных задачах возникают подзадачи, требующие решения больших систем нелинейных уравнений. Примерами таких задач могут послужить задача обучения многослойной нейронной сети или задача об оптимальном распределении мощности между вышками сотовой связи. Возникающие в этих задачах огромные системы уравнений разумно решать распределенно.

Вместо нелинейной системы мы будем рассматривать эквивалентную задачу распределенной оптимизации:  $\min_{x \in \mathbb{R}^p} f(x) = \sum_{i=1}^n f_i(x)$ . В [1] описывается метод сведения этой задачи к задаче ограниченной оптимизации, которая решается прямодвойственным градиентным методом. Ограничения в данной задаче обусловлены необходимостью совпадения решений на различных процессорах при распределенных вычислениях.

Мы, в свою очередь, сводим задачу ограниченной оптимизации к задаче композитной оптимизации, убрав жесткие условия на совпадения решений на различных процессорах. И предлагаем решать полученную задачу аналогами метода подобных треугольников или слайдинга [4].

Наш метод сравнивается с методами градиентного спуска и ускоренного градиентного спуска, описанными в [2] и [5] и имеющими линейную скорость сходимости в условиях Поляка-Лоясевича. Также мы сравнили наш метод с последовательным и параллельным вариантами самого распространенного на данный момент алгоритма обратного распространения ошибки для обучения нейронных сетей с нелинейной функцией активации нейрона, предложенными в [3].

Сравнение производится в ходе вычислительного эксперимента при обучении нейронных сетей с различным количеством слоев и функцией активации нейрона – сигмной. Обучение производится на классических данных (CIFAR, MNIST, IMAGNET).

\*Работа выполнена при финансовой поддержке РФФИ, проекты № 00-00-00000 и 00-00-00001.

## 2 Название раздела

Данный документ демонстрирует оформление статьи, подаваемой в электронную систему подачи статей <http://jmla.org/papers> для публикации в журнале «Машинное обучение и анализ данных». Более подробные инструкции по стилевому файлу `jmla.sty` и использованию издательской системы  $\text{\LaTeX}$  2 $\epsilon$  находятся в документе `authors-guide.pdf`. Работу над статьёй удобно начинать с правки  $\text{\TeX}$ -файла данного документа.

Обращаем внимание, что данный документ должен быть сохранен в кодировке UTF-8 without BOM. Для смены кодировки рекомендуется пользоваться текстовыми редакторами Sublime Text или Notepad++.

### 2.1 Название параграфа

Разделы и параграфы, за исключением списков литературы, нумеруются.

## 3 Заключение

Желательно, чтобы этот раздел был, причём он не должен дословно повторять аннотацию. Обычно здесь отмечают, каких результатов удалось добиться, какие проблемы остались открытыми.

## Литература

- [1] Hamed Karimi, Julie Nutini, and Mark W. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. *CoRR*, abs/1608.04636, 2016.
- [2] Hamed Karimi, Julie Nutini, and Mark W. Schmidt. Linear convergence of gradient and proximal-gradient methods under the polyak-łojasiewicz condition. *CoRR*, abs/1608.04636, 2016.
- [3] G. Sandhya Prafulla. Speaker independent vowel recognition using backpropagation neural network on master-slave architecture jv.s. srinivas,, October 02 2013.
- [4] А. В. Гасников. Универсальный метод для задач стохастической композитной оптимизации. 2016.
- [5] А. В. Гасников. Современные численные методы оптимизации, метод универсального градиентного спуска. 2018.

*Поступила в редакцию 01.01.2017*