

Федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа Прикладной Математики и Информатики
Кафедра интеллектуальных систем

Направление подготовки / специальность: 03.04.01 Прикладная математика и физика


Направленность (профиль) подготовки: Математическая физика, компьютерные технологии и математическое моделирование в экономике

РАСПРЕДЕЛЕННЫЕ МЕТОДЫ ОПТИМИЗАЦИИ И ФЕДЕРАТИВНОЕ ОБУЧЕНИЕ

(магистерская диссертация)

Студент:

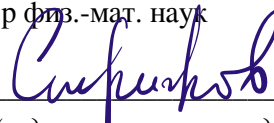
Малиновский Григорий Станиславович



(подпись студента)

Научный руководитель:

Стрижов Вадим Викторович,
д-р физ.-мат. наук



(подпись научного руководителя)

Консультант (при наличии):

Peter Richtarik

(подпись консультанта)

Москва 2021

Аннотация

Данная магистерская диссертация основана на статье "From Local SGD to Local Fixed-Point Methods for Federated Learning" [15] за авторством Григория Малиновского, Лорана Конда, Дмитрия Ковалева, Эльнура Гасанова и Питера Рихтарика.

Большинство алгоритмов для решения задач оптимизации или поиска седловых точек выпукло-вогнутых функций являются методами простой итерации. В этой работе мы рассматриваем общую задачу поиска стационарной точки усредненного оператора или его приближения в распределенной постановке. Наша работа мотивирована потребностями федеративного обучения. В этом контексте каждый локальный оператор моделирует вычисления, выполняемые локально на мобильном устройстве. Мы исследуем две стратегии для достижения такого консенсуса: одна основана на фиксированном количестве локальных шагов, а другая на основе рандомизированных вычислений. В обоих случаях цель состоит в том, чтобы ограничить обмен данными локально вычисляемых переменных, что часто является узким местом в распределенных системах. Мы анализируем скорости сходимости обоих методов и проводим ряд экспериментов, подтверждающих преимущества нашего подхода.

Contents

1	Introduction	4
1.1	Contributions	4
1.2	Mathematical Background	5
2	A Generic Distributed Fixed-Point Method with Local Steps	5
2.1	General Convergence Analysis	7
2.2	Linear Convergence with Contractive Operators	9
3	A Randomized Communication-Efficient Distributed Fixed-Point Method	10
4	Experiments	13
4.1	Local Gradient Descent	15
4.2	Local Cycling Gradient Descent	15
4.3	Results	15
5	Conclusion	16
	References	17
6	Notations and Basic Facts	19
6.1	Notations	19
6.2	Basic Facts	19
6.3	Technical Lemmas	20
7	Analysis of Algorithm 1 in Theorem 2.6	21
7.1	Proof of Lemma 7.1	21
7.2	Proof of Lemma 7.2	24
7.3	Proof of Theorem 2.6	26
7.4	Proof of Corollary 2.7	28
8	Analysis of Algorithm 1: Proof of Theorem 2.14	28
9	Analysis of Algorithm 2	29
9.1	Proof of Lemma 9.1	29
9.2	Proof of Lemma 9.2	32
9.3	Proof of Theorem 3.2	34
9.4	Proof of Corollary 3.3	35

References

- [1] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York, 2nd edition, 2017.
- [2] H. H. Bauschke, R. S. Burachik, P. L. Combettes, V. Elser, D. R. Luke, and H. Wolkowicz, editors. *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*. Springer, 2011.
- [3] S. Chraïbi, Ahmed Khaled, Dmitry Kovalev, Peter Richtárik, Adil Salim, and Martin Takáč. Distributed fixed point methods with compressed iterates. *preprint ArXiv:1912.09925*, 2019.
- [4] P. L. Combettes and Z. C. Woodstock. A fixed point framework for recovering signals from nonlinear transformations. preprint arXiv:2003.01260, 2020.
- [5] P. L. Combettes and I. Yamada. Compositions and convex combinations of averaged nonexpansive operators. *Journal of Mathematical Analysis and Applications*, 425(1): 55–70, 2015.
- [6] L. Condat, D. Kitahara, A. Contreras, and A. Hirabayashi. Proximal splitting algorithms: Relax them all! preprint arXiv:1912.00137, 2019.
- [7] D. Davis and W. Yin. Convergence rate analysis of several splitting schemes. In R. Glowinski, S. J. Osher, and W. Yin, editors, *Splitting Methods in Communication, Imaging, Science, and Engineering*, pages 115–163, Cham, 2016. Springer International Publishing.
- [8] Farzin Haddadpour and Mehrdad Mahdavi. On the convergence of local descent methods in federated learning. *preprint arXiv:1910.14425*, 2019.
- [9] Ahmed Khaled and Peter Richtárik. Gradient descent with compressed iterates. In *NeurIPS Workshop on Federated Learning for Data Privacy and Confidentiality*, 2019.
- [10] Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. First analysis of local GD on heterogeneous data. In *NeurIPS Workshop on Federated Learning for Data Privacy and Confidentiality*, 2019.
- [11] Ahmed Khaled, Konstantin Mishchenko, and Peter Richtárik. Tighter theory for local SGD on identical and heterogeneous data. In *The 23rd International Conference on Artificial Intelligence and Statistics (AISTATS 2020)*, 2020.
- [12] Jakub Konečný, H. Brendan McMahan, Felix X. Yu, Peter Richtárik, Ananda Theertha Suresh, and Dave Bacon. Federated learning: Strategies for improving communication efficiency. In *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
- [13] L. Lessard, B. Recht, and A. Packards. Analysis and design of optimization algorithms via integral quadratic constraints. *SIAM J. Optim.*, 26(1):57–95, 2016.
- [14] Chenxin Ma, Jakub Konečný, Martin Jaggi, Virginia Smith, Michael I Jordan, Peter Richtárik, and Martin Takáč. Distributed optimization with arbitrary local solvers. *Optimization Methods and Software*, 32(4):813–848, 2017.

- [15] Grigory Malinovskiy, Dmitry Kovalev, Elnur Gasanov, Laurent Condat, and Peter Richtarik. From local sgd to local fixed-point methods for federated learning. In *International Conference on Machine Learning*, pages 6692–6701. PMLR, 2020.
- [16] H. Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-efficient learning of deep networks from decentralized data. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS)*, 2017.
- [17] Yurii Nesterov. *Introductory lectures on convex optimization: a basic course*. Kluwer Academic Publishers, 2004.
- [18] J.-C. Pesquet and A. Repetti. A class of randomized primal-dual algorithms for distributed optimization. *J. Nonlinear Convex Anal.*, 12(16), December 2015.
- [19] P. Richtárik and M. Takáč. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Math. Program.*, 144(1–2): 1–38, April 2014.
- [20] Sebastian U. Stich. Local SGD Converges Fast and Communicates Little. In *International Conference on Learning Representations*, 2019.
- [21] Y.-L. Yu. On decomposing the proximal map. In *Proc. of 26th Int. Conf. Neural Information Processing Systems (NIPS)*, pages 91–99, 2013.