Московский физико-технический институт
(Государственный университет)

Физтех-школа Прикладной Математики и Информатики (ФПМИ)
Кафедра интеллектуальных систем

Ковалев Дмитрий Александрович

О стохастическом экстраградиентном методе для вариационных неравенств

Магистерская диссертация

Направление подготовки 03.04.01 «Прикладные математика и физика»
Магистерская программа «Математическая физика, компьютерные технологии и математическое моделирование в экономике»

Студент:
Ковалев Дмитрий Александрович
группа М05-904б

Научный руководитель:
Гасников Александр Владимирович
д-р физ.-мат. наук, доц

Москва, 2021

Аннотация

В данной работе исправлена фундаментальная проблема стохастического экстраградиентного метода с помощью новой стратегии семплирования, мотивированной аппроксимацией неявного градиентного метода. Так как существующий стохастический экстраградиентный метод Mirror-Prox [Juditsky et al., 2011] расходится на простой билинейной задаче, когда область определения неограничена, в данной работе доказываются гарантии сходимости нового метода для более общих постановок, чем в существующих результатах. Численные эксперименты в данной работе показывают, что предложенный вариант экстраградиентного метода сходится на билинейных седловых задачах быстрее, чем многие другие методы. Также в работе рассматривается применение экстраградиентного метода для обучения генеративно-состязательных нейронных сетей и показывается с помощью численных экспериментов, что предложенный подход имеет преимущество по количеству проходов по обучающей выборке, в то время как более высокая стоимость итераций метода уменьшает это преимущество.

Данная работа основана на статье «Revisiting Stochastic Extragradient» [Mishchenko et al., 2020], написанной в соавторстве с Константином Мищенко, Егором Шульгиным, Питером Рихтариком и Юрой Малицким.

Abstract

We fix a fundamental issue in the stochastic extragradient method by providing a new sampling strategy that is motivated by approximating implicit updates. Since the existing stochastic extragradient algorithm, called Mirror-Prox, of [Juditsky et al., 2011] diverges on a simple bilinear problem when the domain is not bounded, we prove guarantees for solving variational inequality that go beyond existing settings. Furthermore, we illustrate numerically that the proposed variant converges faster than many other methods on bilinear saddle-point problems. We also discuss how extragradient can be applied to training Generative Adversarial Networks (GANs) and how it compares to other methods. Our experiments on GANs demonstrate that the introduced approach may make the training faster in terms of data passes, while its higher iteration complexity makes the advantage smaller.

This work is based on a paper «Revisiting Stochastic Extragradient» [Mishchenko et al., 2020] written in collaboration with Konstantin Mishchenko, Egor Shulgin, Peter Richtárik, and Yura Malitsky.

Contents

References

Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223, 2017.

Tatjana Chavdarova, Gauthier Gidel, François Fleuret, and Simon Lacoste-Julien. Reducing noise in gan training with variance reduced extragradient. In *Advances in Neural Information Processing Systems 32*, pages 391–401. Curran Associates, Inc., 2019.

Constantinos Daskalakis and Ioannis Panageas. Last-iterate convergence: Zero-sum games and constrained min-max optimization. *Innovations in Theoretical Computer Science*, 2019.

Constantinos Daskalakis, Andrew Ilyas, Vasilis Syrgkanis, and Haoyang Zeng. Training GANs with optimism. In *International Conference on Learning Representations (ICLR 2018)*, 2018.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Gauthier Gidel, Hugo Berard, Gaëtan Vignoud, Pascal Vincent, and Simon Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*, 2019a. URL https://openreview.net/forum?id=r1laEnA5Ym.

Gauthier Gidel, Reyhane Askari Hemmat, Mohammad Pezeshki, Rémi Le Priol, Gabriel Huang, Simon Lacoste-Julien, and Ioannis Mitliagkas. Negative momentum for improved game dynamics. In Kamalika Chaudhuri and Masashi Sugiyama, editors, *Proceedings of Machine Learning Research*, volume 89 of *Proceedings of Machine Learning Research*, pages 1802–1811. PMLR, 16–18 Apr 2019b.

Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

Anatoli Juditsky, Arkadi Nemirovski, and Claire Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 12 2014.

G. M. Korpelevich. Extragradient method for finding saddle points and other problems. *Matekon*, 13(4):35–49, 1977.

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

Luke Metz, Ben Poole, David Pfau, and Jascha Sohl-Dickstein. Unrolled generative adversarial networks. *arXiv preprint arXiv:1611.02163*, 2016.

Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *CoRR*, abs/1411.1784, 2014.

Konstantin Mishchenko and Peter Richtárik. A stochastic decoupling method for minimizing the sum of smooth and non-smooth functions. *arXiv preprint arXiv:1905.11535*, 2019.

Konstantin Mishchenko, Dmitry Kovalev, Egor Shulgin, Peter Richtárik, and Yura Malitsky. Revisiting stochastic extragradient. In *International Conference on Artificial Intelligence and Statistics*, pages 4573–4582. PMLR, 2020.

Aryan Mokhtari, Asuman Ozdaglar, and Sarath Pattathil. A unified analysis of extra-gradient and optimistic gradient methods for saddle point problems: Proximal point approach. *arXiv preprint arXiv:1901.08511*, 2019.

Arkadi Nemirovski. Prox-method with rate of convergence $o(1/t)$ for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

Yurii Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2-3):319–344, 2007.

Alexander Rakhlin and Karthik Sridharan. Online learning with predictable sequences. 2013.

Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, Xi Chen, and Xi Chen. Improved techniques for training gans. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 2234–2242. Curran Associates, Inc., 2016. URL `http://papers.nips.cc/paper/6125-improved-techniques-for-training-gans.pdf`.

Paul Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms, 2017.

Han Zhang, Ian Goodfellow, Dimitris Metaxas, and Augustus Odena. Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*, 2018.