

# Универсальные методы для стохастических вариационных неравенств

Климза Антон

Руководитель: А. В. Гасников

## Аннотация

В данной статье рассматривается задача оптимизации стохастических вариационных неравенств. Мы предлагаем использование универсального градиентного спуска и показываем его универсальность. Получены оценки необходимого числа итераций для достижения заданного качества решения вариационного неравенства. Также, мы сравниваем полученный алгоритм с другими популярными методами на задаче оптимизации для генеративно-состязательных сетей.

## 1 Введение

Вариационные неравенства нередко возникают в самых разных проблемах оптимизации и имеют многочисленные приложения [1] в математической экономике, теории игр и машинном обучении для задач негладких оптимизаций [2], генеративно-состязательных сетей [3] и обучения с подкреплением [4, 5]. Наиболее известным аналогом градиентного метода для вариационных неравенств является экстраградиентный метод Г.М. Корпелевич [6]. Одним из современных вариантов экстраградиентного метода является проксимальный зеркальный метод А.С. Немировского [7].

Задачу стохастической выпуклой оптимизации уже разбирали в статье [9], в которой предлагается универсальный метод для решения монотонных стохастических вариационных неравенств на базе проксимального зеркального метода. По сути, используется стандартный проксимальный зеркальный метод, в котором  $L$  предлагается выбирать специальным образом, схожим со способом, используемым в Adagrad. Однако этот метод не является полностью адаптивным, поскольку, так же как и в Adagrad, в стратегии выбора шага существенно используется информация о размере решения. Полностью адаптивный метод решения гладких стохастических монотонных вариационных неравенств был построен (с небольшими оговорками) в работе [10].

В новой статье [8] авторы предлагают свой универсальный градиентный спуск для задач стохастической выпуклой оптимизации. Мы предлагаем применение этого метода для стохастических вариационных неравенств, в частности для седловых задач. Такие постановки, например, возникают в задачах состязательного обучения. Преимущества универсального градиентного спуска в том, что он сам настраивается на гладкость задачи и не требует параметров на входе.

## 2 Постановка задачи

Для некоторого оператора  $g : Q \rightarrow \mathbb{R}^n$ , заданного на выпуклом компакте  $Q \in \mathbb{R}^n$ , будем рассматривать сильные вариационные неравенства вида:

$$\langle g(x^*), x^* - x \rangle \leq 0$$

где удовлетворяет условию Липшица. Отметим, что в этом неравенстве требуется найти решение вариационного неравенства  $x^* \in Q$ , для которого

$$\max_{x \in Q} \langle g(x^*), x^* - x \rangle \leq 0$$

В случае монотонного поля наш подход позволяет рассматривать также слабые вариационные неравенства

$$\langle g(x), x^* - x \rangle \leq 0$$

в котором требуется найти  $x^* \in Q$ , такое, что неравенство верно при всех  $x \in Q$ .

## 3 Вычислительный эксперимент

Для сравнения работы универсального градиентного спуска с другими известными методами оптимизации обучим генеративно-сопоставительную сеть с разными оптимизаторами и построим графики метрик качества предсказаний.

Возьмём датасет образцов рукописного написания цифр [MNIST](#). Он содержит 60000 тренировочных и 10000 тестовых картинок размера 28x28, каждая подписана соответствующей ей цифрой.

В качестве модели возьмём простую модель, предложенную в [данной статье](#). В качестве оптимизаторов рассмотрим SGD (Стохастический градиентный спуск), Adam, AdamW (Adam с сокращением веса) и USGM (Универсальный стохастический градиентный спуск). После обучения модели построим графики метрик BCELoss (Бинарный кросс-энтропийный лосс) и WGAN (Wasserstein Loss) [11] зависящих от номера эпохи обучения.

## 4 Список литературы

- [1] Facchinei F., Pang J.S. Finite-Dimensional Variational Inequality and Complementarity Problems. New York: Springer, 2003. V. 1, 2. 693 p
- [2] Y. Nesterov, Smooth minimization of non-smooth functions. Math. Program. 103, 127–152 (2005)
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. WardeFarley, S. Ozair, A. Courville and Y. Bengio, Generative adversarial networks. Commun. ACM 63, 139–144 (2020)
- [4] Y. Jin and A. Sidford, Efficiently solving MDPs with stochastic mirror descent. In Proceedings of the 37th International Conference on Machine Learning (ICML), Proc. Mach. Learn. Res. 119, 4890–4900 (2020)
- [5] S. Omidshafiei, J. Pazis, C. Amato, J. P. How and J. Vian, Deep decentralized multi-task multi-agent reinforcement learning under partial observability. In Proceedings of the 34th International Conference on Machine Learning (ICML), Proc. Mach. Learn. Res. 70, 2681–2690 (2017)

- [6] Корпелевич Г.М. Экстраградиентный метод для отыскания седловых точек и других задач Экономика и матем. методы. Т. 12. № 4. С. 747–756.
- [7] Nemirovski A. 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. 2004. V. 15. P. 229–251.
- [8] Anton Rodomanov Ali Kavis Yongtao Wu Kimon Antonakopoulos Volkan Cevher Universal Gradient Methods for Stochastic Convex Optimization. 2024.
- [9] Bach F., Levy K. Y. A universal algorithm for variational inequalities adaptive to smoothness and noise // arXiv:1902.01637.
- [10] Iusem A. N. et al. Variance-based extragradient methods with line search for stochastic variational inequalities // SIAM Journal on Optimization. 2019. V. 29, № 1. С. 175–206.
- [11] Martin Arjovsky, Soumith Chintala, Léon Bottou, Wasserstein GAN // arXiv:1701.07875, 2017