

Универсальные методы для стохастических вариационных неравенств

А. А. Климза, Ф. С. Стонякин, А. В. Гасников

Московский физико-технический институт (национальный
исследовательский университет)

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

Задачу стохастической выпуклой оптимизации уже разбирали в статье [3], в которой предлагается универсальный метод для решения монотонных стохастических вариационных неравенств на базе проксимального зеркального метода. В новой статье [4] авторы предлагают свой универсальный градиентный спуск для задач стохастической выпуклой оптимизации. Преимущества универсального градиентного спуска в том, что он сам настраивается на гладкость задачи и не требует параметров на входе. Мы предлагаем применение этого метода для стохастических вариационных неравенств. Мы доказываем теоретическую сходимость нашего метода и получаем оценку скорости сходимости:

$$O\left(\inf_{\nu \in [0,1]} \left(\frac{(3-\nu)L_\nu}{\varepsilon}\right)^{\frac{2}{1+\nu}} D^2\right).$$

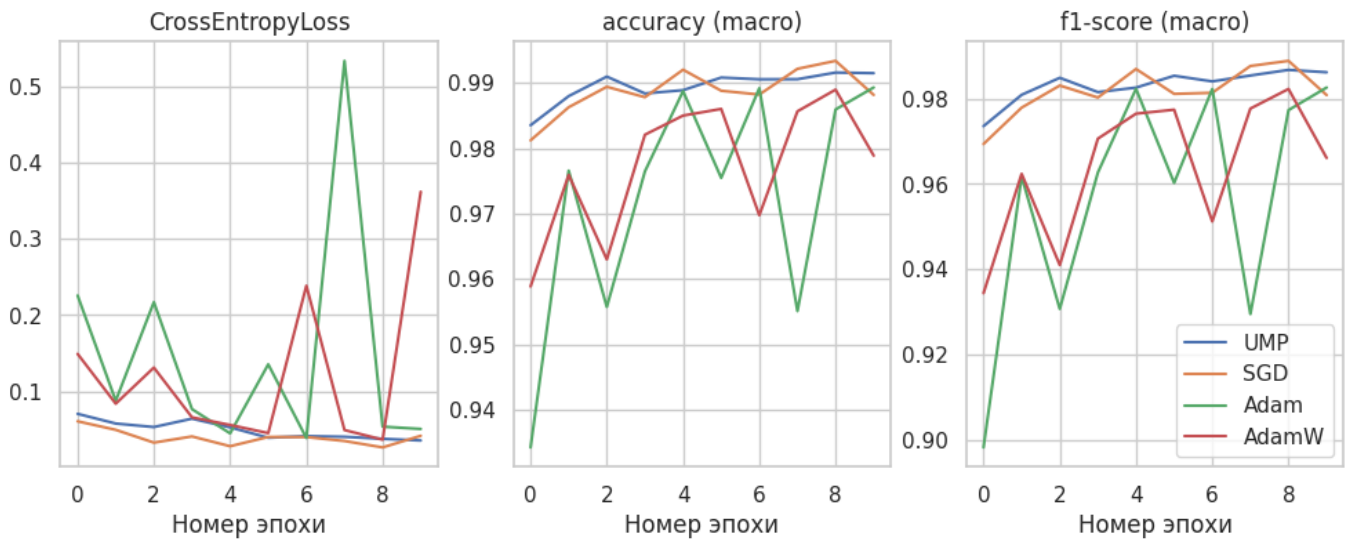
Algorithm 1 Универсальный проксимальный зеркальный метод (UMP)

- 1: Set $z_0 = \arg \min_{u \in Q} d(u)$, $L_0 = \|g(z_0)\|$.
 - 2: **for** $k = 0, 1, \dots$ **do**
 - 3: $w_k = \arg \min_{x \in Q} (\langle g(z_k), x \rangle + L_k \frac{1}{2} \|z_k - x\|^2)$.
 - 4: $z_{k+1} = \arg \min_{x \in Q} (\langle g(w_k), x \rangle + L_k \frac{1}{2} \|z_k - x\|^2)$.
 - 5: $L_{k+1} = L_k + \max\left(0, \frac{-2\langle g(w_k), z_{k+1} - w_k \rangle - L_k \|w_k - z_{k+1}\|^2}{D^2 + \|w_k - z_{k+1}\|^2}\right)$.
 - 6: **end for**
-

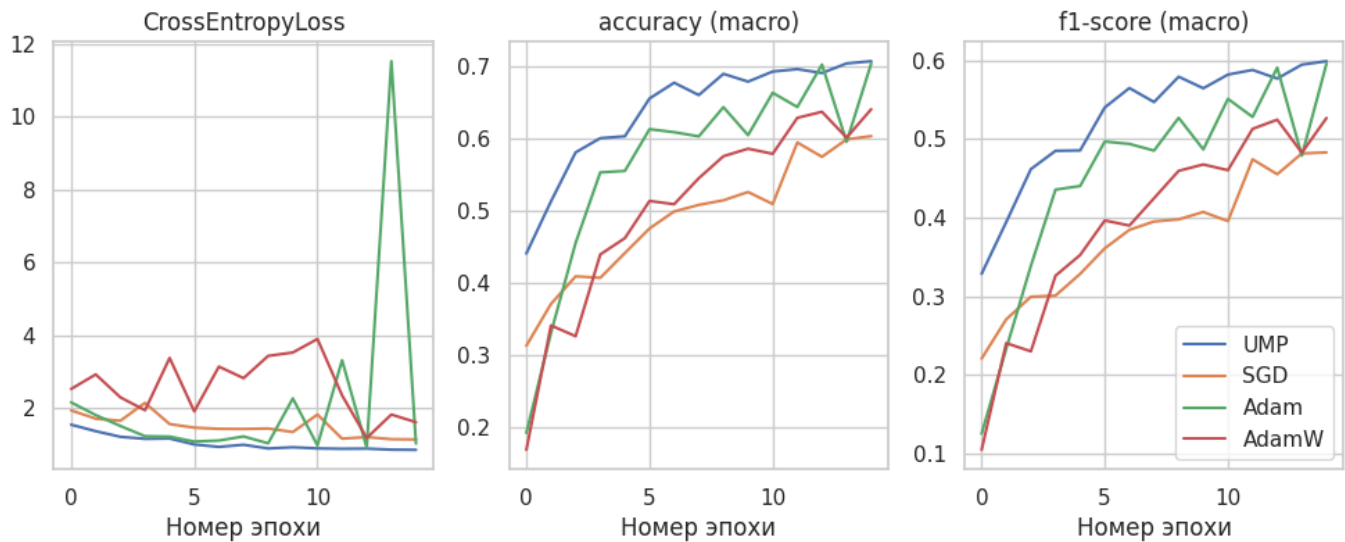
Для сравнения работы универсального проксимального зеркального метода с другими оптимизаторами мы обучаем свёрточную нейронную сеть [resnet50](#) реализованную в библиотеке `pytorch` для классификации изображений датасетов [MNIST](#) и [CIFAR10](#) также предоставленных библиотекой `pytorch`.

В качестве оптимизаторов рассматриваем SGD, Adam, AdamW и наш универсальный проксимальный зеркальный метод UMP.

Получаем, что на датасете MNIST наш метод получил такое же качество, как и SGD, и на датасете CIFAR10 наш метод показал наилучшие результаты. Данный эксперимент показывает, что наш универсальный проксимальный зеркальный метод может быть успешно использован для задач классической минимизации.



Сравнение метрик оптимизаторов на валидации при обучении resnet50 датасете MNIST



Сравнение метрик оптимизаторов на валидации при обучении resnet50 датасете CIFAR10

1 Список литературы

- [1] Корпелевич Г.М. Экстраградиентный метод для отыскания седловых точек и других задач Экономика и матем. методы. Т. 12. № 4. С. 747–756.
- [2] 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.
- [3] Bach F., Levy K. Y. A universal algorithm for variational inequalities adaptive to smoothness and noise. arXiv:1902.01637.
- [4] Anton Rodomanov Ali Kavis Yongtao Wu Kimon Antonakopoulos Volkan Cevher Universal Gradient Methods for Stochastic Convex Optimization. 2024.