# О стохастическом экстраградиентном методе для вариационных неравенств

Дмитрий Ковалев

Московский физико-технический институт

Физтех-школа прикладной математики и информатики

Кафедра интеллектуальных систем

Научный руководитель: д.ф-м.н. Гасников А.В.

# Вариационное неравенство

Найти вектор $x^* \in \mathbb{R}^d$, удовлетворяющий

$$g(x) - g\left(x^*\right) + \left\langle F\left(x^*\right), x - x^*\right\rangle \geq 0 \text{ для всех } x \in \mathbb{R}^d$$

- $g : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ – $\mu$-сильно выпуклая замкнутая функция
- $F : \mathbb{R}^d \to \mathbb{R}^d$ – монотонный $L$-липшицев оператор

# Стохастическое вариационное неравенство

Найти вектор $x^* \in \mathbb{R}^d$, удовлетворяющий

$$g(x) - g\left(x^*\right) + \left\langle F\left(x^*\right), x - x^* \right\rangle \geq 0 \text{ для всех } x \in \mathbb{R}^d$$

$$F(x) = \mathbb{E}_\xi[F(x;\xi)]$$

- $\xi$ – случайный вектор
- $F(x;\xi)$ – почти наверное монотонный $L$-липшицев оператор

# Пример: стохастическая минимизация

$$\boxed{\min_{x \in \mathbb{R}^d} \mathbb{E}_\xi[f(x;\xi)] + g(x)}$$

- $f(x;\xi) : \mathbb{R}^d \to \mathbb{R}$ – почти наверное выпуклая $L$-гладкая функция

$$F(x;\xi) = \nabla f(x;\xi)$$

# Пример: стохастическая седловая задача

$$\min_{x \in \mathbb{R}^{d_x}} \max_{y \in \mathbb{R}^{d_y}} \mathbb{E}_\xi [f(x, y; \xi)] + g_x(x) - g_y(y)$$

- $g_x(x) \colon \mathbb{R}^{d_x} \to \mathbb{R} \cup \{+\infty\}$ и $g_y(y) \colon \mathbb{R}^{d_x} \to \mathbb{R} \cup \{+\infty\}$ – $\mu$-сильно выпуклые функции
- $f(x, y; \xi) : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}$ – почти наверное выпуклая по $x$ и вогнутая по $y$ $L$-гладкая функция

$$F((x, y); \xi) = \begin{bmatrix} \nabla_x f(x, y; \xi) \\ -\nabla_y f(x, y; \xi) \end{bmatrix}, \quad g((x, y)) = g_x(x) + g_y(y)$$

# Экстраградиентный алгоритм

промежуточная точка

**Algorithm 1** Extragradient Method for Variational Inequalities.

1: **Parameters:** $x^0 \in \mathcal{K}$, stepsize $\eta > 0$
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: $\quad y^t = \mathrm{prox}_{\eta g}\left(x^t - \eta F(x^t)\right)$ градиентный шаг из $x^t$
4: $\quad x^{t+1} = \mathrm{prox}_{\eta g}\left(x^t - \eta F(y^t)\right)$
5: **end for**

градиентный шаг из $x^t$ с градиентом взятым в $y^t$

# Стохастический экстраградиентный алгоритм

**Algorithm 1** Extragradient Method for Variational Inequalities.

1: **Parameters:** $x^0 \in \mathcal{K}$, stepsize $\eta > 0$
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: $\quad y^t = \mathrm{prox}_{\eta g}\left(x^t - \eta F(x^t)\right) \longleftarrow F(x^t; \xi_1^t)$
4: $\quad x^{t+1} = \mathrm{prox}_{\eta g}\left(x^t - \eta F(y^t)\right) \longleftarrow F(y^t; \xi_2^t)$
5: **end for**

Векторы $\xi_1^t$ и $\xi_2^t$ равны или выбраны независимо?

# Экстраградиент с независимыми сэмплами

Anatoli Juditsky, Arkadi Nemirovski, Claire Tauvel
*Solving variational inequalities with stochastic mirror-prox algorithm.*
Stochastic Systems 1.1 (2011): 17-58.

- требует равномерную ограниченность шума
  на всей области определения для сходимости

- расходится на простой билинейной седловой задаче
  когда область определения неограничена

# Предлагаемый подход: экстраградиент с одинаковыми сэмплами

**Algorithm 2** Stochastic Extragradient Method for Variational Inequalities.

1: **Parameters:** $x^0 \in \mathcal{K}$, stepsize $\eta > 0$
2: **for** $t = 0, 1, 2, \ldots$ **do**
3: Sample $\xi^t$
4: $y^t = \mathrm{prox}_{\eta g}\left(x^t - \eta F(x^t; \xi^t)\right)$
5: $x^{t+1} = \mathrm{prox}_{\eta g}\left(x^t - \eta F(y^t; \xi^t)\right)$
6: **end for**

требует ограниченность шума только в оптимуме!

# Основная теорема

**Теорема.** Пусть $g\colon \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ – $\mu$-сильно выпуклая замкнутая функция, $F(\cdot;\xi)\colon \mathbb{R}^d \to \mathbb{R}^d$ – почти наверное монотонный $L$-липшицев оператор с ограниченной в оптимуме $x^*$ дисперсией $\mathbb{E}\|F(x^*;\xi) - \mathbb{E}[F(x^*;\xi)]\|^2 \le \sigma^2$. Тогда для любого $\eta \le {}^1\!/_{(2L)}$ выполнено

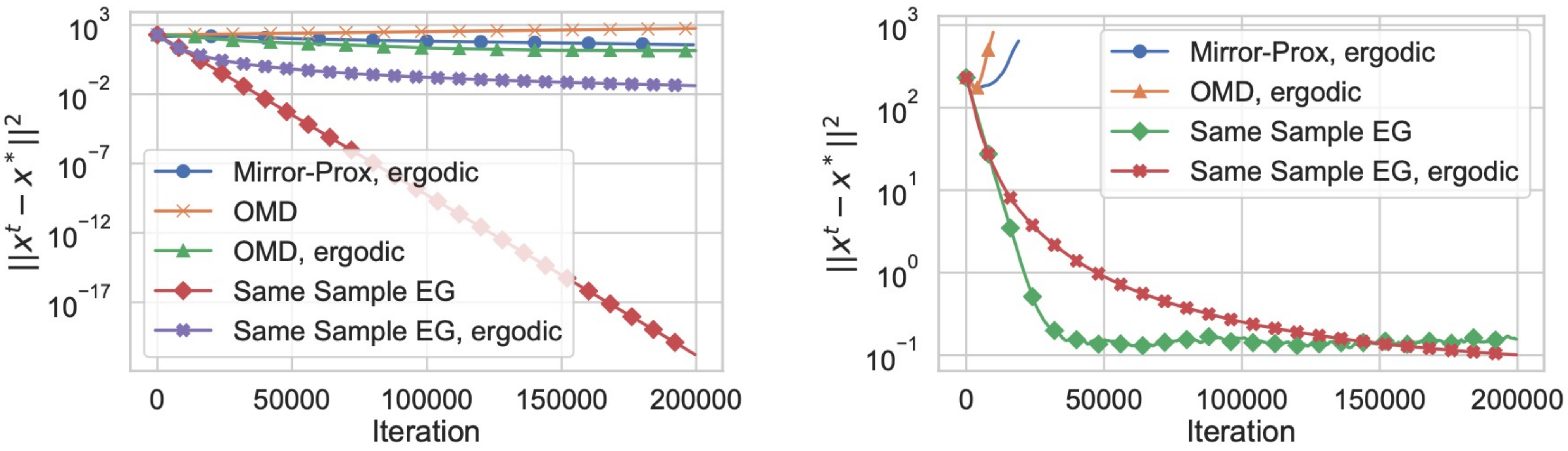
$$\mathbb{E}\|x^t - x^*\|^2 \le \left(1 - \frac{2\eta\mu}{3}\right)^t \|x^0 - x^*\|^2 + \frac{3\eta\sigma^2}{\mu}.$$

# Основная теорема

**Следствие.** Чтобы достичь точности $\|x^t - x^*\|^2 \leq \epsilon$, требуется следующее число итераций:

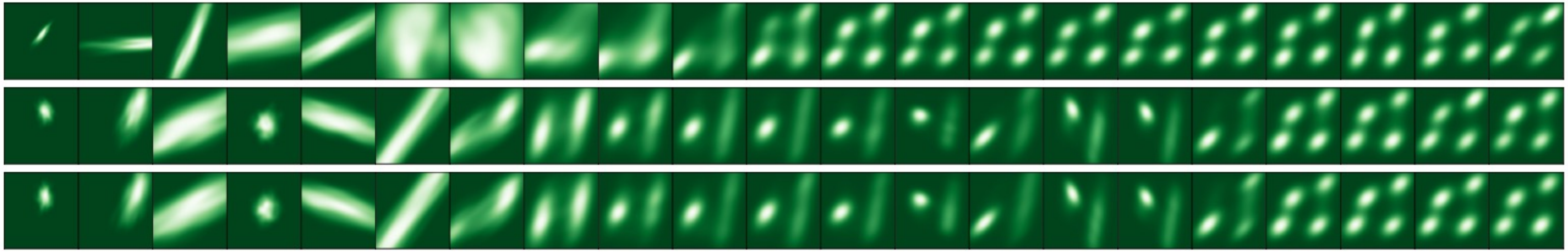
$$t = \mathcal{O}\left(\max\left\{\frac{L}{\mu}, \frac{\sigma^2}{\epsilon\mu^2}\right\}\log\frac{\|x^0 - x^*\|^2}{\epsilon}\right).$$

# Эксперимент: билинейная седловая задача




Сходимость алгоритмов Mirror-Prox (Juditsky et al., 2011), Optimistic Mirror Descent (Gidel et al., 2019) и предложенного алгоритма Same Sample EG. Слева: стохастическая седловая задача $\min_x \max_y \sum_{i=1}^n x^\top \mathbf{B}_i y$. Так как в оптимуме шум равен 0, предложенный алгоритм сходится линейно, в отличие от медленной сходимости алгоритмов (Juditsky et al., 2011) и (Gidel et al., 2019). Справа: стохастическая седловая задача с линейными членами. Так как шум в оптимуме не равен 0, (Juditsky et al., 2011) и (Gidel et al., 2019) расходятся в отличие от предложенного метода.

# Эксперимент: генерация смеси гауссиан



Верхний ряд: предложенный алгоритм. Средний ряд: градиентный спуск-подъем. Нижний ряд: (Juditsky et al., 2011).

# Эксперимент: GAN, CelebA

Результаты обучения self attention GAN с помощью алгоритмов Adam (верхний ряд) и ExtraAdam (нижний ряд) за два прохода по данным. Представлены результаты для трех лучших шагов $10^{-2}, 2 \cdot 10^{-3}, 4 \cdot 10^{-3}$ в порядке от левого к правому.

# Выносится на защиту

- Предложен новый вариант стохастического экстраградиентного метода для вариационных неравенств
- Доказано теоретическое превосходство предложенного метода над известными результатами
- Проведены численные эксперименты, показывающие практическую эффективность предложенного подхода

# Избранные публикации за время обучения

- Dmitry Kovalev, Elnur Gasanov, Peter Richtarik, Alexander Gasnikov
  *Lower Bounds and Optimal Algorithms for Smooth and Strongly Convex Decentralized Optimization Over Time-Varying Networks*
  arXiv preprint arXiv:2106.04469, 2021

- Dmitry Kovalev, Egor Shulgin, Peter Richtarik, Alexander Rogozin, Alexander Gasnikov
  *ADOM: Accelerated Decentralized Optimization Method for Time-Varying Networks*
  Proceedings of the 38th International Conference on Machine Learning, 2021

- Dmitry Kovalev, Adil Salim, Peter Richtarik
  *Optimal and Practical Algorithms for Smooth and Strongly Convex Decentralized Optimization*
  34th Conference on Neural Information Processing Systems, 2020

- Eduard Gorbunov, Dmitry Kovalev, Dmitry Makarenko, Peter Richtarik
  *Linearly Converging Error Compensated SGD*
  34th Conference on Neural Information Processing Systems, 2020

# Избранные публикации за время обучения

- Filip Hanzely, Dmitry Kovalev, Peter Richtarik
  *Variance Reduced Coordinate Descent with Acceleration:*
  *New Method With a Surprising Application to Finite-Sum Problems*
  Proceedings of the 37th International Conference on Machine Learning, 2020

- Zhize Li, Dmitry Kovalev, Xun Qian, Peter Richtarik
  *Acceleration for Compressed Gradient Descent in Distributed and Federated Optimization*
  Proceedings of the 37th International Conference on Machine Learning, 2020

- Grigory Malinovskiy, Dmitry Kovalev, Elnur Gasanov, Laurent Condat, Peter Richtarik
  *From Local SGD to Local Fixed-Point Methods for Federated Learning*
  Proceedings of the 37th International Conference on Machine Learning, 2020

- Konstantin Mishchenko, Dmitry Kovalev, Egor Shulgin, Peter Richtarik, Yura Malitsky
  *Revisiting Stochastic Extragradient*
  Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics, 2020