

# Extragradient Meets VR: Bridging Two Giants to Boost Variational Inequalities

Чирков Георгий, Кабиков Юрий

Руководители: Медяков Даниил и Молодцов Глеб

Московский физико-технический институт

25 марта 2025 г.

# Постановка задачи и идея

## Вариационные неравенства

Ищем  $z^* \in \mathcal{Z}$  такое что:

$$\forall z \in \mathcal{Z} \hookrightarrow \langle F(z^*), z - z^* \rangle \geq 0$$

где  $F$  -  $\mu$  сильно монотонный оператор

## Стохастическая постановка

$$F(z) = \frac{1}{n} \sum_{i=1}^n F_i(z)$$

- Большие датасеты  $\Rightarrow$  нужны стохастические методы
- Обычный стохастический градиент имеет не уменьшающуюся дисперсию

# Предпосылки: Почему VI - это больше чем оптимизация

## Частные случаи:

- Минимизация ( $F = \nabla f$ )
- Седловые задачи:

$$F(x, y) = \begin{pmatrix} \nabla_x f(x, y) \\ -\nabla_y f(x, y) \end{pmatrix}$$

- Фиксированные точки ( $F(z) = z - T(z)$ )

## Приложения:

- GANs (равновесия Нэша)
- Adversarial training
- Экономические модели
- Теория игр
- Оптимальное управление

## Ключевая проблема

Традиционные методы оптимизации не работают для общего случая VI:

- Градиентный спуск расходится
- SGD "осциллирует" около решения
- Нужны специальные методы (Extragradient)

# Проблема дисперсии в Stochastic Extragradient

## Классический алгоритм

$$\begin{aligned} z^{t+\frac{1}{2}} &= z^t - \gamma F_{i_t}(z^t) \\ z^{t+1} &= z^t - \gamma F_{j_t}(z^{t+\frac{1}{2}}) \end{aligned}$$

## Проблема дисперсии

$$\mathbb{E}[\|F_\xi(z^t) - F(z^t)\|^2] \not\rightarrow 0 \quad \text{при} \quad z^t \rightarrow z^*$$

Последствия:

- Сходимость только к окрестности решения
- Необходимость уменьшения шага  $\gamma$
- Медленная сходимость при гетерогенных данных

# Редукция дисперсии через SAGA

## Идея SAGA

Стохастическим образом храним динамический "градиент":

$$\begin{cases} y_{i_t}^{t+1} = F_{i_t}(z^{t+\frac{1}{2}}), \text{ и для } j \neq i_t: y_j^{t+1} = y_j^t \\ G^t = y_j^{t+1} - y_j^t + \frac{1}{n} \sum_{j=1}^n y_j^t \end{cases}$$

Редукция дисперсии - уменьшение разности аппроксиматора и реального градиента:

$$\mathbb{E}[\|G^t - F(z^{t+\frac{1}{2}})\|^2] \rightarrow 0 \quad \text{при} \quad t \rightarrow \infty$$

Преимущества:

- Вообще не вычисляем полный градиент
- Уменьшение дисперсии без хранения референсных точек
- Но строим матрицу динамического градиента

## Что мы сделали

- Тщательно изучили оригинальную статью по совмещению алгоритмов Extragradient и PAGE для решения аналогичных задач вариационного неравенства
- Получили важный результат редукции дисперсии в нашем алгоритме ExtraSAGA
- Проработали несколько альтернативных идей с промежуточным шагом экстраградиента (использование динамического градиента, стохастического градиента в точке)

# Планы на будущее

- Получение финальных оценок сходимости нашего метода
- Проведение качественных экспериментов нашего метода (в частности на GAN)

# Заключение

- Нами предложен новый алгоритм ExtraSAGA - комбинация Extragradient и SAGA.
- Расширение на область задач вариационных неравенств позволяет решать более широкий класс задач, возникающих на практике.

Спасибо за внимание!