

# Координатные методы распределенной оптимизации в условиях гомогенности данных. Постановка задачи

Подготовил: Максимов Роман

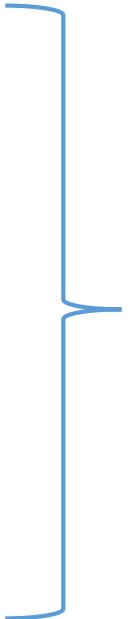
Участники проекта: Максимов Роман, Алимаскина Екатерина

Руководитель: Былинкин Дмитрий, МФТИ

11.02.2025

# Мотивация

- Данные могут быть не гомогенны
- Вычисление мастер-градиента зачастую намного дешевле
- Датасеты огромные и хранить и обрабатывать их приходится на множестве устройств
- Вычисление градиентов по разным блокам координат можно параллелизировать
- Хотим уменьшать коммуникационную сложность



Федеративное обучение

# План

- Берем координатный спуск

---

## Algorithm 1 Parallel Coordinate Descent Method 1 (PCDM1)

---

```
1: Choose initial point  $x_0 \in \mathbf{R}^N$ 
2: for  $k = 0, 1, 2, \dots$  do
3:   Randomly generate a set of blocks  $S_k \subseteq \{1, 2, \dots, n\}$ 
4:    $x_{k+1} \leftarrow x_k + (h(x_k))_{[S_k]}$ 
5: end for
```

---

Литература: <https://arxiv.org/pdf/1212.0873.pdf>

Square Loss	$\frac{1}{2}(A_j^T x - y_j)^2$
Logistic Loss	$\log(1 + e^{-y_j A_j^T x})$
Hinge Square Loss	$\frac{1}{2} \max\{0, 1 - y_j A_j^T x\}^2$

Примеры частично  
сепарабельных функций

- Применяем к экстраградиенту:

---

**Algorithm 1** Accelerated Extragradient
 

---

```

1: Input:  $x^0 = x_f^0 \in \mathbb{R}^d$ 
2: Parameters:  $\tau \in (0, 1), \eta, \theta, \alpha > 0, K \in \{1, 2, \dots\}$ 
3: for  $k = 0, 1, 2, \dots, K - 1$  do
4:    $x_g^k = \tau x^k + (1 - \tau)x_f^k$ 
5:    $x_f^{k+1} \approx \arg \min_{x \in \mathbb{R}^d} [A_\theta^k(x) := p(x_g^k) + \langle \nabla p(x_g^k), x - x_g^k \rangle + \frac{1}{2\theta} \|x - x_g^k\|^2 + q(x)]$ 
6:    $x^{k+1} = x^k + \eta \alpha (x_f^{k+1} - x^k) - \eta \nabla r(x_f^{k+1})$ 
7: end for
8: Output:  $x^K$ 
  
```

---

$$\begin{cases} p \rightarrow (f - f_1)(x, y) \\ q \rightarrow f_1(x, y) \end{cases}$$

$$\nabla p \rightarrow \zeta = \begin{cases} \frac{1}{p} \nabla_x (f - f_1), & p \\ \frac{1}{1-p} \nabla_y (f - f_1), & 1-p \end{cases}$$

Сэмплируем координату при вычислении градиента

Литература: <https://arxiv.org/pdf/2205.15136>

# Текущая задача:

**Lemma 1.** Consider Algorithm 1. Let  $\theta$  be defined as in Theorem 1:  $\theta = \frac{1}{2L_p}$ . Then, under Assumptions 1-3, the following inequality holds for all  $\bar{x} \in \mathbb{R}^d$

$$\begin{aligned} 2\langle \bar{x} - x_g^k, \nabla r(x_f^{k+1}) \rangle &\leq 2 \left[ r(\bar{x}) - r(x_f^{k+1}) \right] - \mu \|x_f^{k+1} - \bar{x}\|^2 - \theta \|\nabla r(x_f^{k+1})\|^2 \\ &\quad + 3\theta \left( \|\nabla A_\theta^k(x_f^{k+1})\|^2 - \frac{L_p^2}{3} \|x_g^k - \arg \min_{x \in \mathbb{R}^d} A_\theta^k(x)\|^2 \right). \end{aligned} \quad (19)$$

**Lemma 2.** Consider Algorithm 1 for Problem 1 under Assumption 1-3 with the following tuning:

$$\tau = \min \left\{ 1, \frac{\sqrt{\mu}}{2\sqrt{L_p}} \right\}, \quad \theta = \frac{1}{2L_p}, \quad \eta = \min \left\{ \frac{1}{2\mu}, \frac{1}{2\sqrt{\mu L_p}} \right\}, \quad \alpha = \mu, \quad (20)$$

and let  $x_f^{k+1}$  in line 5 satisfy

$$\|\nabla A_\theta^k(x_f^{k+1})\|^2 \leq \frac{L_p^2}{3} \|x_g^k - \arg \min_{x \in \mathbb{R}^d} A_\theta^k(x)\|^2. \quad (21)$$

Then, the following inequality holds:

$$\frac{1}{\eta} \|x^{k+1} - x^*\|^2 + \frac{2}{\tau} \left[ r(x_f^{k+1}) - r(x^*) \right] \leq (1 - \rho) \left[ \frac{1}{\eta} \|x^k - x^*\|^2 + \frac{2}{\tau} [r(x_f^k) - r(x^*)] \right], \quad (22)$$

where

$$\rho := \frac{1}{2} \min \left\{ 1, \sqrt{\frac{\mu}{L_p}} \right\}. \quad (23)$$

## Дальнейшие задачи:

- Варьирование исходя из симилярити по разным координатам (блокам координат) для достижения лучших оценок
- Вычислительные эксперименты