# Децентрализованная оптимизация на меняющихся графах

Нят Чунг Нгуен

Московский физико-технический институт

16 декабря 2024 г.

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Распределенная оптимизация

Распределенная оптимизация - важное направление исследований в современной теории оптимизации. Ее приложения включают:

# Распределенная оптимизация

Распределенная оптимизация - важное направление исследований в современной теории оптимизации. Ее приложения включают:

- машинное обучение на больших данных

# Распределенная оптимизация

Распределенная оптимизация - важное направление исследований в современной теории оптимизации. Ее приложения включают:

- машинное обучение на больших данных
- распределенную обработку сигналов

# Распределенная оптимизация

Распределенная оптимизация - важное направление исследований в современной теории оптимизации. Ее приложения включают:

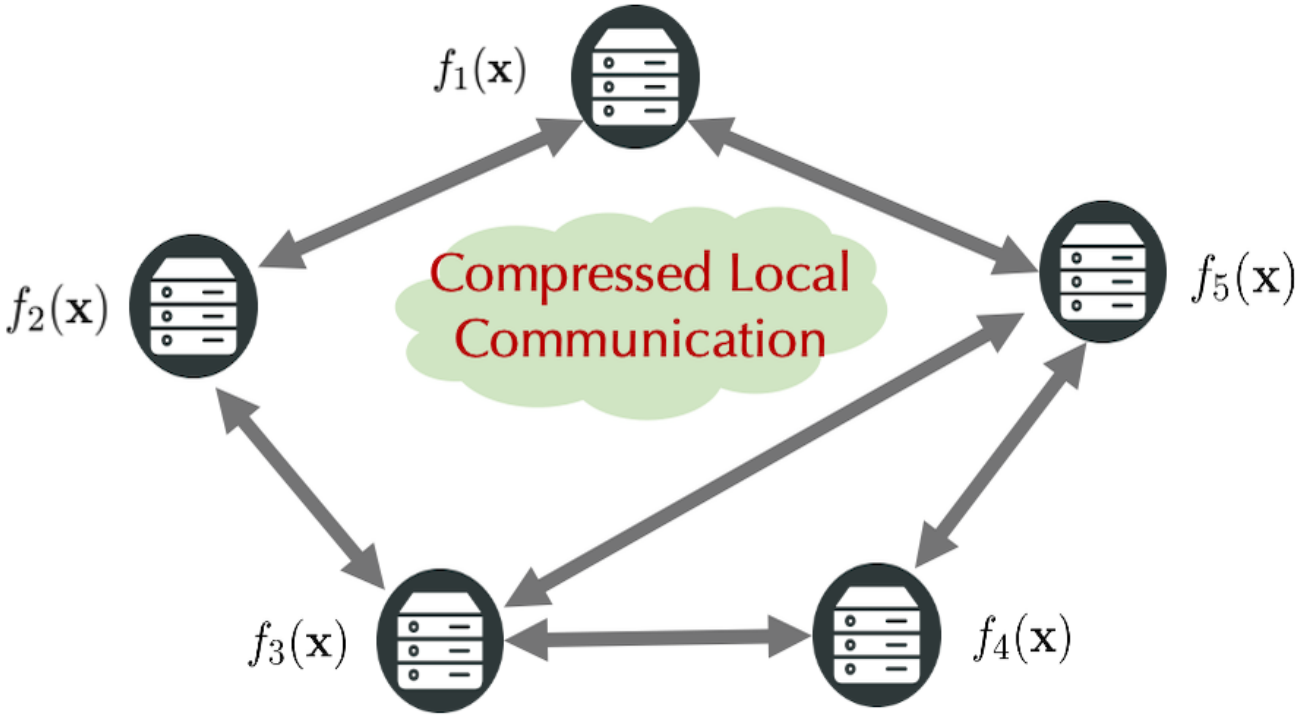
- машинное обучение на больших данных
- распределенную обработку сигналов
- распределенное управление

# Распределенная оптимизация

Распределенная оптимизация - важное направление исследований в современной теории оптимизации. Ее приложения включают:

- машинное обучение на больших данных
- распределенную обработку сигналов
- распределенное управление
- другие области

# Коммуникация




Основное внимание в работе уделяется оптимизации на меняющихся сетях.

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Постановка задачи

Мы рассматриваем оптимизационную задачу типа

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

# Постановка задачи

Мы рассматриваем оптимизационную задачу типа

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ - выпуклое множество

# Постановка задачи

Мы рассматриваем оптимизационную задачу типа

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ - выпуклое множество
- Каждая функция $f_i(x)$ хранится локально на некотором компьютере

# Постановка задачи

Мы рассматриваем оптимизационную задачу типа

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ - выпуклое множество
- Каждая функция $f_i(x)$ хранится локально на некотором компьютере
- Компьютеры соединены друг с другом через децентрализованную коммуникационную сеть. Они могут выполнять локальные вычисления и обмениваться информацией со своими непосредственными соседями в сети

# Постановка задачи

Мы рассматриваем оптимизационную задачу типа

$$\min_{x \in \mathcal{X}} f(x) := \frac{1}{n} \sum_{i=1}^{n} f_i(x)$$

- $\mathcal{X} \subseteq \mathbb{R}^d$ - выпуклое множество
- Каждая функция $f_i(x)$ хранится локально на некотором компьютере
- Компьютеры соединены друг с другом через децентрализованную коммуникационную сеть. Они могут выполнять локальные вычисления и обмениваться информацией со своими непосредственными соседями в сети
- Сеть может меняться со временем. Этот тип сетей называется *меняющимися графами*

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Обозначения и предположения

## Определение (Выпуклость, Сильно Выпуклось, и Гладкость)

Рассмотрим функцию $h : \mathbb{R}^d \to \mathbb{R}$.

1. $h$ называется **выпуклой** если для любых $x_1, x_2 \in \mathbb{R}^d$, выполняется:

$$h(x_2) \geq h(x_1) + \langle \nabla h(x_1), x_2 - x_1 \rangle.$$

2. $h$ называется $\mu$-**сильно выпукло** если для любых $x_1, x_2 \in \mathbb{R}^d$, выполняется:

$$h(x_2) \geq h(x_1) + \langle \nabla h(x_1), x_2 - x_1 \rangle + \frac{\mu}{2}\|x_2 - x_1\|^2.$$

3. $h$ называется $L$-**гладкой** если для любых $x_1, x_2 \in \mathbb{R}^d$, выполняется:

$$h(x_2) \leq h(x_1) + \langle \nabla h(x_1), x_2 - x_1 \rangle + \frac{L}{2}\|x_2 - x_1\|^2.$$

# Обозначения и предположения

**Предположение**

Для каждого $i = 1, \ldots, n$ функция $f_i$ является выпуклой.

**Предположение**

Для каждого $i = 1, \ldots, n$ функция $f_i$ является $\mu_i$-сильно выпуклой.

**Предположение**

Для каждого $i = 1, \ldots, n$ функция $f_i$ является $L_i$-гладкой.

**Предположение**

Для каждого $i = 1, \ldots, n$ норма субградиентов функции $f_i$ ограничена величиной $M_i$, то есть

$$\|\partial f_i(x)\| \leq M_i.$$

# Обозначения и предположения

## Предположение

Предположим, что допустимая область $\mathcal{X}$ является выпуклой и ограниченной. Обозначим через $R$ радиус шара, содержащего $\mathcal{X}$, то есть

$$\forall x \in \mathcal{X} : \ \|x\| \leq R.$$

Определим следующие локальные константы для задачи:

$$L = \max_{i=1,\dots,m} L_i, \quad M = \max_{i=1,\dots,m} M_i, \quad \mu = \min_{i=1,\dots,m} \mu_i.$$

# Обозначения и предположения

**Децентрализованная коммуникация**

# Обозначения и предположения

**Децентрализованная коммуникация**

- На каждом шаге обмена данными мы используем граф для представления соединения между вычислительными машинами

# Обозначения и предположения

**Децентрализованная коммуникация**

- На каждом шаге обмена данными мы используем граф для представления соединения между вычислительными машинами
- Обозначим коммуникационную сеть, меняющуюся со времени, последовательностью графов $\{\mathcal{G}^k\}_{k=0}^{\infty} = \{(\mathcal{V}, \mathcal{E}^k)\}_{k=0}^{\infty}$, где $\mathcal{V} = \{1, \ldots, n\}$ - множество узлов (каждый узел соответствует одному компьютеру), а $\mathcal{E}^k$ - множество доступных соединений на $k$-м коммуникационном раунде

# Обозначения и предположения

**Децентрализованная коммуникация**

- На каждом шаге обмена данными мы используем граф для представления соединения между вычислительными машинами
- Обозначим коммуникационную сеть, меняющуюся со времени, последовательностью графов $\{\mathcal{G}^k\}_{k=0}^{\infty} = \{(\mathcal{V}, \mathcal{E}^k)\}_{k=0}^{\infty}$, где $\mathcal{V} = \{1, \ldots, n\}$ - множество узлов (каждый узел соответствует одному компьютеру), а $\mathcal{E}^k$ - множество доступных соединений на $k$-м коммуникационном раунде
- Мы обозначаем $\mathcal{N}_i^k = \{j \in \mathcal{V} | (i,j) \in \mathcal{E}^k\}$ - множество соседей $i$-го узла на $k$-м раунде, и в это время он может взаимодействовать только с узлами в $\mathcal{N}_i^k$.

# Обозначения и предположения

## Матрицы коммуникации

# Обозначения и предположения

**Матрицы коммуникации**

Каждому графу в коммуникационной сети соответствует матрица $W^k \in \mathbb{R}^{n \times n}$, которая удовлетворяет следующим свойствам.

# Обозначения и предположения

**Матрицы коммуникации**

Каждому графу в коммуникационной сети соответствует матрица $W^k \in \mathbb{R}^{n \times n}$, которая удовлетворяет следующим свойствам.

## Предположение

1. $W^k$ является положительно полуопределенной,
2. $W_{ij}^k = 0$ если $i \neq j$ и $(i, j) \notin \mathcal{E}^k$,
3. $\ker W^k = \text{span}(\overline{1})$, где $\overline{1} = (1, \ldots, 1) \in \mathbb{R}^n$.

# Обозначения и предположения

**Матрицы коммуникации**

Каждому графу в коммуникационной сети соответствует матрица $W^k \in \mathbb{R}^{n \times n}$, которая удовлетворяет следующим свойствам.

## Предположение

1. $W^k$ является положительно полуопределенной,
2. $W^k_{ij} = 0$ если $i \neq j$ и $(i,j) \notin \mathcal{E}^k$,
3. $\ker W^k = \operatorname{span}(\overline{1})$, где $\overline{1} = (1, \ldots, 1) \in \mathbb{R}^n$.

## Пример

Лапласиан графа

$$W^k_{ij} \triangleq \begin{cases} -1, & \text{если } (i,j) \in \mathcal{E}^k, \\ \deg(i), & \text{если } i = j, \\ 0, & \text{иначе.} \end{cases}$$

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Нижние оценки и оптимальные алгоритмы для гладкой и сильно выпуклой оптимизации

| | Статические сети | Произв. меняющ. | Медленно меняющ. |
|---|---|---|---|
| Коммуникация | $\sqrt{\chi}\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ | $\chi\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ | $\chi\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ |
| Выч. градиента | $\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ | $\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ | $\sqrt{\frac{L}{\mu}}\log\frac{1}{\varepsilon}$ |
| Статья | Scaman et al. (2017), Kovalev et al. (2020) | Kovalev et al. (2021) | Metelev et al. (2023), Kovalev et al. (2021) |

Таблица: Скорости сходимости для гладкой и сильно выпуклой оптимизации

# Нижние оценки и оптимальные алгоритмы для негладкой оптимизации

| Настройка | Тип сложности | Статические сети | Изменяющиеся во времени сети |
|---|---|---|---|
| Сильно выпуклая | Децентрализованная коммуникация | $\sqrt{\chi}\frac{M}{\sqrt{\mu\varepsilon}}$ | $\chi\frac{M}{\sqrt{\mu\varepsilon}}$ |
| | Вычисление субградиента | $\frac{M^2}{\mu\varepsilon}$ | $\frac{M^2}{\mu\varepsilon}$ |
| Выпуклая | Децентрализованная коммуникация | $\sqrt{\chi}\frac{MR}{\varepsilon}$ | $\chi\frac{MR}{\varepsilon}$ |
| | Вычисление субградиента | $\frac{M^2R^2}{\varepsilon^2}$ | $\frac{M^2R^2}{\varepsilon^2}$ |
| Статья | | Scaman et al. (2018), Lan et al. (2020) | Kovalev et al. (2024) |

Таблица: Скорости сходимости для негладкой оптимизации

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Задачи с аффинными ограничениями

Задача децентрализованной оптимизации с аффинными ограничениями:

$$\min_{x \in \mathcal{K}} f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x) \quad \text{s.t.} \quad Ax = b,\ A \in \mathbb{R}^{m \times d},\ \mathcal{K} \subseteq \mathbb{R}^d.$$

# Задачи с аффинными ограничениями

Децентрализованную оптимизацию с глобальными аффинными ограничениями:

$$\min_{x_1 \in \mathbb{R}^{d_1}, \ldots, x_n \in \mathbb{R}^{d_n}} f(x) = \sum_{i=1}^{n} f_i(x) \quad \text{s.t.} \quad \sum_{i=1}^{n} (A_i x_i - b_i) = 0,$$

где $A_i \in \mathbb{R}^{m \times d_i}$ и $b_i \in \mathbb{R}^m$. Мы предполагаем, что функции $f_i$, матрицы $A_i$ и векторы $b_i$ хранятся локально узлами вычислительной сети.

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Постановка задачи

Мы рассматриваем задачу минимаксной оптимизации типа

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \; f(x, y) := \frac{1}{n} \sum_{i=1}^{n} f_i(x, y)$$

# Постановка задачи

Мы рассматриваем задачу минимаксной оптимизации типа

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y) := \frac{1}{n} \sum_{i=1}^{n} f_i(x, y)$$

- $\mathcal{X}, \mathcal{Y}$ - замкнутые выпуклые множества

# Седловые задачи

Мы вводим множество $\mathcal{Z} = \mathcal{X} \times \mathcal{Y}$, и оператор $F$:

$$F_i(z) = F_i(x, y) = \begin{pmatrix} \nabla_x f_i(x, y) \\ -\nabla_y f_i(x, y) \end{pmatrix}, \quad F(z) = \frac{1}{n} \sum_{i=1}^{n} F_i(z)$$

## Предположение

Пусть функции $f(x, y)$ и $f_i(x, y)$ удовлетворяют следующим условиям:

1. Функция $f(x, y)$ является $L$-гладкой, т.е. для всех $z_1, z_2 \in \mathcal{Z}$ выполнено
$$\|F(z_1) - F(z_2)\| \leq L\|z_1 - z_2\|.$$
2. Для любого $i$, $f_i(x, y)$ является $L_{\max}$-гладкой, т.е. для всех $z_1, z_2 \in \mathcal{Z}$ выполнено
$$\|F_i(z_1) - F_i(z_2)\| \leq L_{\max}\|z_1 - z_2\|.$$
3. Функция $f(x, y)$ является сильно выпуклой-сильно вогнутой с константой $\mu$, т.е. для всех $z_1, z_2 \in \mathcal{Z}$ выполнено
$$\langle F(z_1) - F(z_2), z_1 - z_2\rangle \geq \mu\|z_1 - z_2\|^2.$$

# Седловые задачи

| | Статические сети | Произв. меняющ. | Медленно меняющ. |
|---|---|---|---|
| Коммуникация | $\sqrt{\chi}\frac{L}{\mu}\log\frac{1}{\varepsilon}$ | $\chi\frac{L}{\mu}\log\frac{1}{\varepsilon}$ | $\chi\frac{L}{\mu}\log\frac{1}{\varepsilon}$ |
| Выч. градиента | $\frac{L}{\mu}\log\frac{1}{\varepsilon}$ | $\frac{L}{\mu}\log\frac{1}{\varepsilon}$ | $\frac{L}{\mu}\log\frac{1}{\varepsilon}$ |
| Статья | Beznosikov et al. (2021b), Kovalev et al. (2022) | Beznosikov et al. (2021a), Kovalev et al. (2022) | Nguyen et al. (2023), Kovalev et al. (2022) |

Таблица: Скорости сходимости для гладких и сильно выпукло-вогнутых задач седловой точки

# Негладкие седловые задачи

Негладкие минимаксные задачи могут быть решены методом субградиента или зеркальным проксимальным методом

Однако эти алгоритмы не являются оптимальными с точки зрения коммуникации (нижние границы не были установлены), и поэтому негладкие децентрализованные задачи седловой точки остаются открытым направлением исследований как для статических, так и для изменяющихся во времени графов.

Расширение задач седловой точки для включения аффинных и связанных ограничений представляет собой перспективное направление для дальнейших исследований.

# План


1. Актуальность
2. Постановка задачи
3. Обозначения и предположения
4. Нижние оценки и оптимальные алгоритмы
5. Задачи с аффинными ограничениями
6. Седловые задачи
7. Заключение

# Заключение

- Децентрализованная оптимизация находит применение в различных областях, включая распределенное управление и обработка сигналов, машинное обучение на больших данных.

# Заключение

- Децентрализованная оптимизация находит применение в различных областях, включая распределенное управление и обработка сигналов, машинное обучение на больших данных.
- Одной из ключевых задач в этой области является разработка методов, способных учитывать изменения в структуре сети. Наблюдается отсутствие некоторых исследований, посвященных негладкому случаю, что оставляет пробел в этом направлении.

# Заключение

- Децентрализованная оптимизация находит применение в различных областях, включая распределенное управление и обработка сигналов, машинное обучение на больших данных.
- Одной из ключевых задач в этой области является разработка методов, способных учитывать изменения в структуре сети. Наблюдается отсутствие некоторых исследований, посвященных негладкому случаю, что оставляет пробел в этом направлении.
- В данной работе рассматриваются существующие результаты и методы, а также выделяются несколько открытых вопросов, которые могут послужить ориентиром для дальнейших исследований.

# Список литературы I


Aleksandr Beznosikov, Alexander Rogozin, Dmitry Kovalev, and Alexander Gasnikov. Near-optimal decentralized algorithms for saddle point problems over time-varying networks. In *International Conference on Optimization and Applications*, pages 246–257. Springer, 2021a.

Aleksandr Beznosikov, Valentin Samokhin, and Alexander Gasnikov. Distributed saddle-point problems: Lower bounds, optimal algorithms and federated gans. *arXiv preprint arXiv:2010.13112*, 2021b.

Dmitry Kovalev, Adil Salim, and Peter Richtárik. Optimal and practical algorithms for smooth and strongly convex decentralized optimization. *Advances in Neural Information Processing Systems*, 33:18342–18352, 2020.

Dmitry Kovalev, Elnur Gasanov, Alexander Gasnikov, and Peter Richtarik. Lower bounds and optimal algorithms for smooth and strongly convex decentralized optimization over time-varying networks. *Advances in Neural Information Processing Systems*, 34, 2021.

# Список литературы II

Dmitry Kovalev, Aleksandr Beznosikov, Abdurakhmon Sadiev, Michael Persiianov, Peter Richtárik, and Alexander Gasnikov. Optimal algorithms for decentralized stochastic variational inequalities. *Advances in Neural Information Processing Systems*, 35:31073–31088, 2022.

Dmitry Kovalev, Ekaterina Borodich, Alexander Gasnikov, and Dmitrii Feoktistov. Lower bounds and optimal algorithms for non-smooth convex decentralized optimization over time-varying networks. *arXiv preprint arXiv:2405.18031*, 2024.

Guanghui Lan, Soomin Lee, and Yi Zhou. Communication-efficient algorithms for decentralized and stochastic optimization. *Mathematical Programming*, 180(1):237–284, 2020.

Dmitry Metelev, Aleksandr Beznosikov, Alexander Rogozin, Alexander Gasnikov, and Anton Proskurnikov. Decentralized optimization over slowly time-varying graphs: Algorithms and lower bounds. *arXiv preprint arXiv:2307.12562*, 2023.

# Список литературы III

Nhat Trung Nguyen, Alexander Rogozin, Dmitry Metelev, and Alexander Gasnikov. Min-max optimization over slowly time-varying graphs. In *Doklady Mathematics*, volume 108, pages S300–S309. Springer, 2023.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Yin Tat Lee, and Laurent Massoulié. Optimal algorithms for smooth and strongly convex distributed optimization in networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 3027–3036. JMLR. org, 2017.

Kevin Scaman, Francis Bach, Sébastien Bubeck, Laurent Massoulié, and Yin Tat Lee. Optimal algorithms for non-smooth distributed optimization in networks. *Advances in Neural Information Processing Systems*, 31, 2018.