Лагранжиан. Седловая задача. Метод экстраградиента. Прямо-двойственный метод.

Методы оптимизации в ML

Александр Безносиков

1 апреля 2024

Лагранжиан

Рассматриваем задачу условной оптимизации вида:

$$\begin{aligned} \min_{x \in \mathbb{R}^d} \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \leq 0, \ i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

Здесь матрица $A \in \mathbb{R}^{n \times d}$ и $b \in \mathbb{R}^n$.

Лагранжиан

Рассматриваем задачу условной оптимизации вида:

$$
\begin{aligned}
\min_{x \in \mathbb{R}^d} \;& f_0(x) \\
\text{s.t.} \;& f_i(x) \leq 0, \; i = 1, \ldots, m \\
& Ax = b
\end{aligned}
$$

Здесь матрица $A \in \mathbb{R}^{n \times d}$ и $b \in \mathbb{R}^n$. Здесь еще можно было немного обобщить постановку и добавить, что $x \in \mathcal{X} \cap \operatorname{dom} f_i$. Но мы предполагаем, что $\mathcal{X} \cap \operatorname{dom} f_i = \mathbb{R}^d$.

Лагранжиан

Лагранжиан

Функция Лагранжа/Лагранжиан для этой задачи строится следующим образом:

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \nu^T (Ax - b),$$

где $\lambda_i \geq 0$ для $i = 1, \ldots, m$, а $\nu \in \mathbb{R}^n$. λ_i можно записать в виде векторов λ соответствующей размерности.

Что делали на семинаре

Что делали на семинаре:

Что делали на семинаре

Что делали на семинаре:

- Рассматривали:

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^d} L(x, \lambda, \nu).$$

Вопрос: как называется этот объект?

Что делали на семинаре

Что делали на семинаре:

- Рассматривали:

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^d} L(x, \lambda, \nu).$$

Вопрос: как называется этот объект? двойственная функция

Что делали на семинаре

Что делали на семинаре:

- Рассматривали:
$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^d} L(x, \lambda, \nu).$$
Вопрос: как называется этот объект? двойственная функция
- Осознали, что для любой $\lambda \succeq 0$ и $\nu \in \mathbb{R}^n$
$$g(\lambda, \nu) \leq f(x^*).$$

Что делали на семинаре

Что делали на семинаре:

Что делали на семинаре

Что делали на семинаре:

- Узнали

Условие Слейтера

Будем говорить, что для задачи с ограничениями выполняется условие Слейтера, если существует $x \in \mathbb{R}^d$, такой что

$$f_i(x) < 0, \; i = 1, \ldots, m \quad \text{и} \quad Ax = b.$$

Что делали на семинаре

Что делали на семинаре:

- Узнали

Условие Слейтера

Будем говорить, что для задачи с ограничениями выполняется условие Слейтера, если существует $x \in \mathbb{R}^d$, такой что

$$f_i(x) < 0, \;\; i = 1, \ldots, m \quad \text{и} \quad Ax = b.$$

- **Вопрос:** и что оно дает?

Что делали на семинаре

Что делали на семинаре:

- Узнали

Условие Слейтера

Будем говорить, что для задачи с ограничениями выполняется условие Слейтера, если существует $x \in \mathbb{R}^d$, такой что

$$f_i(x) < 0, \; i = 1, \ldots, m \quad \text{и} \quad Ax = b.$$

- **Вопрос:** и что оно дает?

Теорема Слейтера

Если в задаче с ограничениями все функции являются выпуклыми и выполняется условие Слейтера, то тогда при построении двойственной задачи выполняется свойство сильной двойственности, а именно

$$\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} g(\lambda, \nu) = f(x^*).$$

Седловая точка

Седловая точка

Точка $(x^*, \lambda^*, \nu^*) \in \mathbb{R}^d \times \mathbb{R}^m_+ \times \mathbb{R}^n$ называется седловой для функции $L(x, \lambda, \nu)$, если для любых $(x, \lambda, \nu) \in \mathbb{R}^d \times \mathbb{R}^m_+ \times \mathbb{R}^n$ выполнено

$$L(x, \lambda^*, \nu^*) \geq L(x^*, \lambda^*, \nu^*) \geq L(x^*, \lambda, \nu).$$

Теорема о седловой точке Куна-Таккера

Теорема о седловой точке Куна-Таккера

Для задачи выпуклой оптимизации с выпуклыми ограничениями с выполненными условием Слейтера следующие утверждения эквиваленты:

- для x^* существует $\lambda^* \succeq 0$ и $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа,
- x^* – глобальное решение задачи оптимизации с ограничениями.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$? $= +\infty$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$? $= +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)? = +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$? $= +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)? = +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. Первое неравенство из определения седловой задачи дает:
$$f_0(x) + \sum_{j=1}^m \lambda_j^* f_j(x) + (\nu^*)^T(Ax - b) \geq L(x^*, \lambda^*, \nu^*) = f_0(x^*).$$

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)? = +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. Первое неравенство из определения седловой задачи дает:
$$f_0(x) + \sum_{j=1}^m \lambda_j^* f_j(x) + (\nu^*)^T(Ax - b) \geq L(x^*, \lambda^*, \nu^*) = f_0(x^*).$$
А это и есть то, что мы хотели. **Вопрос:** почему?

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \succeq 0$, $\nu^* \in \mathbb{R}^n$ такие, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $f_i(x^*) > 0$ для некоторого i (или $Ax^* \neq b$). **Вопрос:** что можно сказать про $\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$? $= +\infty$. **Вопрос:** может ли такое быть? Нет, 2ое неравенство в определении седловой точки рушится для $\lambda = (.\lambda^*_{i-1}, 2\lambda^*_i, \lambda^*_{i+1}.)^T$. Аналогично разбирается случай, когда $Ax^* \neq b$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. Первое неравенство из определения седловой задачи дает:
$$f_0(x) + \sum_{j=1}^m \lambda_j^* f_j(x) + (\nu^*)^T (Ax - b) \geq L(x^*, \lambda^*, \nu^*) = f_0(x^*).$$
А это и есть то, что мы хотели. **Вопрос:** почему? для допустимых x (удовлетворяет ограничениям), имеем, что левая часть $\leq f_0(x)$, так как λ^*_j неотрицательные.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации?

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$. Откуда $f_0(x^*) \leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*) + (\nu^*)^T (Ax^* - b)$.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$. Откуда $f_0(x^*) \leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*) + (\nu^*)^T(Ax^* - b)$. **Вопрос:** что можем сказать про $\lambda_j^* f_j(x^*)$ и $Ax^* - b$? равны 0.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$. Откуда $f_0(x^*) \leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*) + (\nu^*)^T (Ax^* - b)$. **Вопрос:** что можем сказать про $\lambda_j^* f_j(x^*)$ и $Ax^* - b$? равны 0. Поэтому $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$.

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$. Откуда $f_0(x^*) \leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*) + (\nu^*)^T (Ax^* - b)$. **Вопрос:** что можем сказать про $\lambda_j^* f_j(x^*)$ и $Ax^* - b$? равны 0. Поэтому $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. Откуда первую часть определения седловой задачи:
$$L(x^*, \lambda^*, \nu^*) = f_0(x^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*).$$

Доказательство

$\Leftarrow$ Пусть x^* – глобальное решение задачи с ограничениями, а также пусть все f_0 и $\{f_i\}$ выпуклые и выполнено условие Слейтера, то существует $\lambda^* \succeq 0$ и $\nu \in \mathbb{R}^n$ такое, что (x^*, λ^*, ν^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной (λ^*, ν^*), $f_0(x^*) = g(\lambda^*, \nu^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*)$. Откуда $f_0(x^*) \leq L(x^*, \lambda^*, \nu^*) = f_0(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*) + (\nu^*)^T(Ax^* - b)$. **Вопрос:** что можем сказать про $\lambda_j^* f_j(x^*)$ и $Ax^* - b$? равны 0. Поэтому $L(x^*, \lambda^*, \nu^*) = f_0(x^*)$. Откуда первую часть определения седловой задачи:
$$L(x^*, \lambda^*, \nu^*) = f_0(x^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*, \nu^*).$$
Вторая часть получается из того $f_j(x^*) \leq 0$ и $Ax^* - b = 0$, а значит $f_0(x^*) \geq f_0(x^*) + \sum_{j=1}^m \lambda_j f_j(x^*) + \nu^T(Ax^* - b) = L(x^*, \lambda, \nu)$ для $\lambda_j \geq 0$.

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.

Абстрогируемся от функции Лагранжа. Пусть есть некоторая функция:

$$L(x, \lambda) : \mathcal{X} \times \Lambda \to \mathbb{R}.$$

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.

Абстрогируемся от функции Лагранжа. Пусть есть некоторая функция:

$$L(x, \lambda) : \mathcal{X} \times \Lambda \to \mathbb{R}.$$

Рассмотрим следующую игровую интерпретацию

- Пусть есть два игрока: первый игрок может выбирать $x \in \mathcal{X}$, а второй – $\lambda \in \Lambda$ (это может быть распределение ресурсов, выбор действие и т.д.).

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.

Абстрогируемся от функции Лагранжа. Пусть есть некоторая функция:

$$L(x, \lambda) : \mathcal{X} \times \Lambda \to \mathbb{R}.$$

Рассмотрим следующую игровую интерпретацию

- Пусть есть два игрока: первый игрок может выбирать $x \in \mathcal{X}$, а второй – $\lambda \in \Lambda$ (это может быть распределение ресурсов, выбор действие и т.д.).
- Функция $L(x, \lambda)$ некоторое значение прибыли в зависимости от выбранных $x \in \mathcal{X}$ и $\lambda \in \Lambda$. Первый игрок платит второму игроку сумму $L(x, \lambda)$.

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.

Абстрогируемся от функции Лагранжа. Пусть есть некоторая функция:

$$L(x, \lambda) : \mathcal{X} \times \Lambda \to \mathbb{R}.$$

Рассмотрим следующую игровую интерпретацию

- Пусть есть два игрока: первый игрок может выбирать $x \in \mathcal{X}$, а второй – $\lambda \in \Lambda$ (это может быть распределение ресурсов, выбор действие и т.д.).
- Функция $L(x, \lambda)$ некоторое значение прибыли в зависимости от выбранных $x \in \mathcal{X}$ и $\lambda \in \Lambda$. Первый игрок платит второму игроку сумму $L(x, \lambda)$.
- **Вопрос:** чего хочет первый, а чего хочет второй?

Седловая задача и игры

Седловые задачи – это больше, чем просто функция Лагранжа. Это отдельный и более общий класс задач, который имеет свои приложения.
Абстрогируемся от функции Лагранжа. Пусть есть некоторая функция:

$$L(x, \lambda) : \mathcal{X} \times \Lambda \to \mathbb{R}.$$

Рассмотрим следующую игровую интерпретацию

- Пусть есть два игрока: первый игрок может выбирать $x \in \mathcal{X}$, а второй – $\lambda \in \Lambda$ (это может быть распределение ресурсов, выбор действие и т.д.).
- Функция $L(x, \lambda)$ некоторое значение прибыли в зависимости от выбранных $x \in \mathcal{X}$ и $\lambda \in \Lambda$. Первый игрок платит второму игроку сумму $L(x, \lambda)$.
- **Вопрос:** чего хочет первый, а чего хочет второй? Первый хочет платить меньше, а второй хочет получить больше.

Седловая задача и игры

- Нужно найти равновесие между игроками, потому что экстремальная стратегия не значит лучшая. Например, если игрок два знает, что при некотором $\tilde{x}$ игрока один можно получить огромный выигрыш при правильном выборе $\lambda(\tilde{x})$. В то же время может так оказаться, что при других $x \neq \tilde{x}$ выбор $\lambda(\tilde{x})$ дает нулевой выигрыш второму игроку, тогда игрок один может просто никогда не выбирать $\tilde{x}$. При этом может быть $\tilde{\lambda}$, которая при любом x будет давать небольшой фиксированный выигрыш.

Седловая задача и игры

- Нужно найти равновесие между игроками, потому что экстремальная стратегия не значит лучшая. Например, если игрок два знает, что при некотором $\tilde{x}$ игрока один можно получить огромный выигрыш при правильном выборе $\lambda(\tilde{x})$. В то же время может так оказаться, что при других $x \neq \tilde{x}$ выбор $\lambda(\tilde{x})$ дает нулевой выигрыш второму игроку, тогда игрок один может просто никогда не выбирать $\tilde{x}$. При этом может быть $\tilde{\lambda}$, которая при любом x будет давать небольшой фиксированный выигрыш.
- С точки зрения седловой задачи:
$$L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda),$$
получается следующее:

Седловая задача и игры

- Нужно найти равновесие между игроками, потому что экстремальная стратегия не значит лучшая. Например, если игрок два знает, что при некотором $\tilde{x}$ игрока один можно получить огромный выигрыш при правильном выборе $\lambda(\tilde{x})$. В то же время может так оказаться, что при других $x \neq \tilde{x}$ выбор $\lambda(\tilde{x})$ дает нулевой выигрыш второму игроку, тогда игрок один может просто никогда не выбирать $\tilde{x}$. При этом может быть $\tilde{\lambda}$, которая при любом x будет давать небольшой фиксированный выигрыш.
- С точки зрения седловой задачи:
$$L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda),$$
получается следующее: пусть $(\tilde{x}, \tilde{\lambda})$ – седло, тогда любые изменения x игрока один будут приводить к тому, что он будет платить больше (обратно для игрока два).

Седловая задача и игры

- Нужно найти равновесие между игроками, потому что экстремальная стратегия не значит лучшая. Например, если игрок два знает, что при некотором $\tilde{x}$ игрока один можно получить огромный выигрыш при правильном выборе $\lambda(\tilde{x})$. В то же время может так оказаться, что при других $x \neq \tilde{x}$ выбор $\lambda(\tilde{x})$ дает нулевой выигрыш второму игроку, тогда игрок один может просто никогда не выбирать $\tilde{x}$. При этом может быть $\tilde{\lambda}$, которая при любом x будет давать небольшой фиксированный выигрыш.
- С точки зрения седловой задачи:
$$L(x, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*, \lambda),$$
получается следующее: пусть $(\tilde{x}, \tilde{\lambda})$ – седло, тогда любые изменения x игрока один будут приводить к тому, что он будет платить больше (обратно для игрока два). В обратную сторону, если, например, $\tilde{x}$ не часть решения седловой задачи, то игрок один сможет изменить x при фиксированной $\tilde{\lambda}$ и платить меньше.

Седловая задача и игры

- **Вопрос:** одинаковый ли будет результат игры, если
  1) сначала будет выбирать первый игрок, а потом второй
  2) сначала будет выбирать второй игрок, а потом первый?

Седловая задача и игры

- **Вопрос:** одинаковый ли будет результат игры, если
  1) сначала будет выбирать первый игрок, а потом второй
  2) сначала будет выбирать второй игрок, а потом первый?
- Нет, как рассуждает первый игрок в первом случае: «если я выберу некоторый x, то второй игрок тут же будет максимизировать свой выигрыш: $\sup_\lambda L(x, \lambda)$, значит мне надо выбрать x так, чтобы минимизировать потери: $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$

Седловая задача и игры

- **Вопрос:** одинаковый ли будет результат игры, если
  1) сначала будет выбирать первый игрок, а потом второй
  2) сначала будет выбирать второй игрок, а потом первый?
- Нет, как рассуждает первый игрок в первом случае: «если я выберу некоторый x, то второй игрок тут же будет максимизировать свой выигрыш: $\sup_\lambda L(x, \lambda)$, значит мне надо выбрать x так, чтобы минимизировать потери: $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$
- Используя эту интуицию можно понять, что в общем случае, что

$$\sup_\lambda \inf_x L(x, \lambda) \leq \inf_x \sup_\lambda L(x, \lambda)$$

Седловая задача и игры

- **Вопрос:** одинаковый ли будет результат игры, если
  1) сначала будет выбирать первый игрок, а потом второй
  2) сначала будет выбирать второй игрок, а потом первый?
- Нет, как рассуждает первый игрок в первом случае: «если я выберу некоторый x, то второй игрок тут же будет максимизировать свой выигрыш: $\sup_\lambda L(x, \lambda)$, значит мне надо выбрать x так, чтобы минимизировать потери: $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$
- Используя эту интуицию можно понять, что в общем случае, что

$$\sup_\lambda \inf_x L(x, \lambda) \leq \inf_x \sup_\lambda L(x, \lambda)$$

- Формально:

$$\inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \lambda) \leq L(x, \lambda) \; \forall x \in \mathcal{X} \;\; \Rightarrow \;\; \sup_{\lambda \in \Lambda} \inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \lambda) \leq \sup_{\lambda \in \Lambda} L(x, \lambda) \; \forall x \in \mathcal{X}$$

Седловая задача и игры

- **Вопрос:** одинаковый ли будет результат игры, если
  1) сначала будет выбирать первый игрок, а потом второй
  2) сначала будет выбирать второй игрок, а потом первый?
- Нет, как рассуждает первый игрок в первом случае: «если я выберу некоторый x, то второй игрок тут же будет максимизировать свой выигрыш: $\sup_\lambda L(x, \lambda)$, значит мне надо выбрать x так, чтобы минимизировать потери: $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$
- Используя эту интуицию можно понять, что в общем случае, что

$$\sup_\lambda \inf_x L(x, \lambda) \leq \inf_x \sup_\lambda L(x, \lambda)$$

- Формально:

$$\inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \lambda) \leq L(x, \lambda) \; \forall x \in \mathcal{X} \;\; \Rightarrow \;\; \sup_{\lambda \in \Lambda} \inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \lambda) \leq \sup_{\lambda \in \Lambda} L(x, \lambda) \; \forall x \in \mathcal{X}$$

Откуда $\sup_{\lambda \in \Lambda} \inf_{\tilde{x} \in \mathcal{X}} L(\tilde{x}, \lambda) \leq \inf_{x \in \mathcal{X}} \sup_{\lambda \in \Lambda} L(x, \lambda)$

Седловая задача и игры

Как две игры: $\sup_\lambda \inf_x L(x, \lambda)$ $\inf_x \sup_\lambda L(x, \lambda)$ связаны с седловой точкой?

Седловая задача и игры

Как две игры: $\sup_\lambda \inf_x L(x, \lambda)$ $\inf_x \sup_\lambda L(x, \lambda)$ связаны с седловой точкой?

Теорема о седловой точке

Множество седловых точек функции $L : \mathcal{X} \times \Lambda \to \mathbb{R}$ непустое тогда и только тогда, когда обе задачи $\sup_\lambda \inf_x L(x, \lambda)$ и $\inf_x \sup_\lambda L(x, \lambda)$ имеют решение и эти решения совпадают.

Седловая задача и игры

Теорема Сиона-Какутани

Пусть $\mathcal{X}$, Λ выпуклые компактные множества, пусть также $L : \mathcal{X} \times \Lambda \to \mathbb{R}$ непрерывна, выпукла по x (для любого фиксированного λ) и вогнута по λ (для любого фиксированного x). Тогда L имеет седловые точки на $\mathcal{X} \times \Lambda$.

Седловая задача и игры

Теорема Сиона-Какутани

Пусть $\mathcal{X}$, Λ выпуклые компактные множества, пусть также $L : \mathcal{X} \times \Lambda \to \mathbb{R}$ непрерывна, выпукла по x (для любого фиксированного λ) и вогнута по λ (для любого фиксированного x). Тогда L имеет седловые точки на $\mathcal{X} \times \Lambda$.

Теорема Сиона-Какутани

Пусть $\mathcal{X}$, Λ выпуклые множества, и $\mathcal{X}$ или Λ дополнительно компактно, пусть также $L : \mathcal{X} \times \Lambda \to \mathbb{R}$ непрерывна, выпукла по x (для любого фиксированного λ) и вогнута по λ (для любого фиксированного x). Тогда (гарантий существования тут нет)

$$\sup_{\lambda \in \Lambda} \inf_{x \in \mathcal{X}} L(x, \lambda) = \inf_{x \in \mathcal{X}} \sup_{\lambda \in \Lambda} L(x, \lambda)$$

Седловая задача

- Оптимизация функции Лагранжа — седловая задача.

Седловая задача

- Оптимизация функции Лагранжа — седловая задача.
- Седловые задачи возникают как отдельный большой класс задач.

Седловая задача

- Оптимизация функции Лагранжа — седловая задача.
- Седловые задачи возникают как отдельный большой класс задач.
- Будем рассматривать следующую задачу:

$$\min_{x \in \mathbb{R}^d} \max_{\lambda \in \mathbb{R}^d} L(x, \lambda),$$

где L непрерывно дифференцируема по обеим группам переменных, выпукла-вогнута: выпукла по x (для любого фиксированного λ) и вогнута по λ (для любого фиксированного x), а также градиенты по обеим группам переменных являются $L/\sqrt{2}$-Липшицевыми:

$$\|\nabla_x L(x_1, \lambda_1) - \nabla_x L(x_2, \lambda_2)\|_2^2 \leq \frac{L^2}{2}(\|x_1 - x_2\|_2^2 + \|\lambda_1 - \lambda_2\|_2^2)$$

$$\|\nabla_\lambda L(x_1, \lambda_1) - \nabla_\lambda L(x_2, \lambda_2)\|_2^2 \leq \frac{L^2}{2}(\|x_1 - x_2\|_2^2 + \|\lambda_1 - \lambda_2\|_2^2)$$

С

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.
- Рассмотрим задачу $\min_{x \in R} \max_{\lambda \in \mathbb{R}} x\lambda$. Стартовая точка $(1, 1)$. **Вопрос:** где решение?

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.
- Рассмотрим задачу $\min_{x \in R} \max_{\lambda \in \mathbb{R}} x\lambda$. Стартовая точка $(1, 1)$. **Вопрос:** где решение? Точка $(0, 0)$.

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.
- Рассмотрим задачу $\min_{x \in R} \max_{\lambda \in \mathbb{R}} x\lambda$. Стартовая точка $(1, 1)$. **Вопрос:** где решение? Точка $(0, 0)$.
- Вектор $\binom{\nabla_x L(x^k, \lambda^k)}{-\nabla_\lambda L(x^k, \lambda^k)}$ всегда ортогонален направлению на решение $\binom{x^k - x^*}{\lambda^k - \lambda^*}$. **Вопрос:** что это значит?

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.
- Рассмотрим задачу $\min_{x \in R} \max_{\lambda \in \mathbb{R}} x\lambda$. Стартовая точка $(1, 1)$. **Вопрос:** где решение? Точка $(0, 0)$.
- Вектор $\binom{\nabla_x L(x^k, \lambda^k)}{-\nabla_\lambda L(x^k, \lambda^k)}$ всегда ортогонален направлению на решение $\binom{x^k - x^*}{\lambda^k - \lambda^*}$. **Вопрос:** что это значит? Метод не стремится к решению.

Ищем решение седловой задачи

- Идея спуска-подъема неплохая и часто рабочая, но интуиция подсказывает, что у него есть не самые приятные аспекты.
- Рассмотрим задачу $\min_{x \in R} \max_{\lambda \in \mathbb{R}} x\lambda$. Стартовая точка $(1, 1)$. **Вопрос:** где решение? Точка $(0, 0)$.
- Вектор $\begin{pmatrix} \nabla_x L(x^k, \lambda^k) \\ -\nabla_\lambda L(x^k, \lambda^k) \end{pmatrix}$ всегда ортогонален направлению на решение $\begin{pmatrix} x^k - x^* \\ \lambda^k - \lambda^* \end{pmatrix}$. **Вопрос:** что это значит? Метод не стремится к решению.
- Интуиция не является сторогой, но может подсказать, что нужно попробовать что-то чуть-чуть другое.

Экстраградиентный метод

Алгоритм 1 Экстраградиентный метод

Вход: размер шага $\gamma > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $x^{k+1/2} = x^k - \gamma \nabla_x L(x^k, \lambda^k)$

3: $\lambda^{k+1/2} = \lambda^k + \gamma \nabla_\lambda L(x^k, \lambda^k)$

4: $x^{k+1} = x^k - \gamma \nabla_x L(x^{k+1/2}, \lambda^{k+1/2})$

5: $\lambda^{k+1} = \lambda^k + \gamma \nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2})$

6: **end for**

Выход: $\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1/2}$

Экстраградиентный метод

Алгоритм 2 Экстраградиентный метод

Вход: размер шага $\gamma > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**
2: $\quad x^{k+1/2} = x^k - \gamma \nabla_x L(x^k, \lambda^k)$
3: $\quad \lambda^{k+1/2} = \lambda^k + \gamma \nabla_\lambda L(x^k, \lambda^k)$
4: $\quad x^{k+1} = x^k - \gamma \nabla_x L(x^{k+1/2}, \lambda^{k+1/2})$
5: $\quad \lambda^{k+1} = \lambda^k + \gamma \nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2})$
6: **end for**

Выход: $\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1/2}$

Легко проверить, что для этого метода на задаче $\min_{x\in R} \max_{\lambda\in\mathbb{R}} x\lambda$, направления итогового градиентного шага в скалярном приведении с направлением на решение дает число больше 0, а значит острый угол.

Доказательство

Для удобства введем следующие обозначения

- Вектор переменных z и оператор:

$$z = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(z) = F(x, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -\nabla_\lambda L(x, \lambda) \end{pmatrix}$$

Доказательство

Для удобства введем следующие обозначения

- Вектор переменных z и оператор:

$$z = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(z) = F(x, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -\nabla_\lambda L(x, \lambda) \end{pmatrix}$$

- Если $\nabla_x L$ и $\nabla_\lambda L$ L-Липшицевы, то L-Липшицев и оператор F. Как проявляется выпуклость по x и вогнутость по λ увидим позже.

Доказательство

Для удобства введем следующие обозначения

- Вектор переменных z и оператор:

$$z = \begin{pmatrix} x \\ \lambda \end{pmatrix}, \quad F(z) = F(x, \lambda) = \begin{pmatrix} \nabla_x L(x, \lambda) \\ -\nabla_\lambda L(x, \lambda) \end{pmatrix}$$

- Если $\nabla_x L$ и $\nabla_\lambda L$ L-Липшицевы, то L-Липшицев и оператор F. Как проявляется выпуклость по x и вогнутость по λ увидим позже.
- **Вопрос:** как экстраградиентный метод будет выглядеть в новых обозначениях?

$$\begin{aligned} z^{k+1/2} &= z^k - \gamma F(z^k) \\ z^{k+1} &= z^k - \gamma F(z^{k+1/2}) \end{aligned}$$

Доказательство

Для начала докажем следующую лемму:

Лемма

Пусть $z, y \in \mathbb{R}^d$, и $z^+ = z - y$, тогда для любого $u \in \mathbb{R}^d$:

$$\|z^+ - u\|_2^2 = \|z - u\|_2^2 - 2\langle y, z^+ - u\rangle - \|z^+ - z\|_2^2.$$

Доказательство: тут достаточно обычных алгебраических преобразований:

$$\begin{aligned}
\|z^+ - u\|_2^2 &= \|z^+ - z + z - u\|_2^2 \\
&= \|z - u\|_2^2 + 2\langle z^+ - z, z - u\rangle + \|z^+ - z\|_2^2 \\
&= \|z - u\|_2^2 + 2\langle z^+ - z, z^+ - u\rangle - \|z^+ - z\|_2^2 \\
&= \|z - u\|_2^2 - 2\langle y, z^+ - u\rangle - \|z^+ - z\|_2^2.
\end{aligned}$$

Доказательство

Применим доказанную лемму два раза для итерации экстраградиентного метода:

- Для $z = z^k$, $y = \gamma F(z^{k+1/2})$ и $z^+ = z^{k+1}$:

$$\|z^{k+1} - u\|_2^2 = \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1} - u\rangle - \|z^{k+1} - z^k\|_2^2$$

Доказательство

Применим доказанную лемму два раза для итерации экстраградиентного метода:

- Для $z = z^k$, $y = \gamma F(z^{k+1/2})$ и $z^+ = z^{k+1}$:

$$\|z^{k+1} - u\|_2^2 = \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1} - u\rangle - \|z^{k+1} - z^k\|_2^2$$

- Для $z = z^k$, $y = \gamma F(z^k)$ и $z^+ = z^{k+1/2}$:

$$\|z^{k+1/2} - \tilde{u}\|_2^2 = \|z^k - \tilde{u}\|_2^2 - 2\gamma\langle F(z^k), z^{k+1/2} - \tilde{u}\rangle - \|z^{k+1/2} - z^k\|_2^2$$

Доказательство

- С предыдущего слайда:

$$\|z^{k+1} - u\|_2^2 = \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1} - u\rangle - \|z^{k+1} - z^k\|_2^2$$
$$\|z^{k+1/2} - \tilde{u}\|_2^2 = \|z^k - \tilde{u}\|_2^2 - 2\gamma\langle F(z^k), z^{k+1/2} - \tilde{u}\rangle - \|z^{k+1/2} - z^k\|_2^2$$

- Подставим вместо $\tilde{u} = z^{k+1}$ и сложим

$$\begin{aligned}\|z^{k+1} - u\|_2^2 + \|z^{k+1/2} - z^{k+1}\|_2^2 &= \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1} - u\rangle \\ &\quad - 2\gamma\langle F(z^k), z^{k+1/2} - z^{k+1}\rangle - \|z^{k+1/2} - z^k\|_2^2\end{aligned}$$

Доказательство

- Немного поработаем с выражением с прошлого слайда:

$$
\begin{aligned}
\|z^{k+1} - u\|_2^2 + \|z^{k+1/2} - z^{k+1}\|_2^2 \\
= & \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle \\
& - 2\gamma\langle F(z^k) - F(z^{k+1/2}), z^{k+1/2} - z^{k+1}\rangle - \|z^{k+1/2} - z^k\|_2^2
\end{aligned}
$$

Доказательство

- Немного поработаем с выражением с прошлого слайда:

$$\begin{aligned}\|z^{k+1} - u\|_2^2 + \|z^{k+1/2} - z^{k+1}\|_2^2 \\ =& \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle \\ &- 2\gamma\langle F(z^k) - F(z^{k+1/2}), z^{k+1/2} - z^{k+1}\rangle - \|z^{k+1/2} - z^k\|_2^2\end{aligned}$$

- КБШ:

$$\begin{aligned}\|z^{k+1} - u\|_2^2 + \|z^{k+1/2} - z^{k+1}\|_2^2 \\ \leq& \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle - \|z^{k+1/2} - z^k\|_2^2 \\ &+ \gamma^2\|F(z^k) - F(z^{k+1/2})\|_2^2 + \|z^{k+1/2} - z^{k+1}\|_2^2\end{aligned}$$

или

$$\begin{aligned}\|z^{k+1} - u\|_2^2 \leq& \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle \\ &+ \gamma^2\|F(z^k) - F(z^{k+1/2})\|_2^2 - \|z^{k+1/2} - z^k\|_2^2\end{aligned}$$

Доказательство

- L-Липшицевость F:

$$\|z^{k+1} - u\|_2^2 \leq \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle + \gamma^2 L^2\|z^k - z^{k+1/2}\|_2^2 - \|z^{k+1/2} - z^k\|_2^2$$

Доказательство

- L-Липшицевость F:

$$\|z^{k+1} - u\|_2^2 \leq \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle + \gamma^2 L^2 \|z^k - z^{k+1/2}\|_2^2 - \|z^{k+1/2} - z^k\|_2^2$$

- $\gamma \leq \frac{1}{L}$:

$$\|z^{k+1} - u\|_2^2 \leq \|z^k - u\|_2^2 - 2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle$$

или

$$2\gamma\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle \leq \|z^k - u\|_2^2 - \|z^{k+1} - u\|_2^2$$

Доказательство

- Работаем с

$$
\begin{aligned}
\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle
&= \left\langle \begin{pmatrix} \nabla_x L(x^{k+1/2}, \lambda^{k+1/2}) \\ -\nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2}) \end{pmatrix}, \begin{pmatrix} x^{k+1/2} \\ \lambda^{k+1/2} \end{pmatrix} - \begin{pmatrix} u_x \\ u_\lambda \end{pmatrix} \right\rangle \\
&= \langle \nabla_x L(x^{k+1/2}, \lambda^{k+1/2}), x^{k+1/2} - u_x\rangle \\
&\quad + \langle -\nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2}), \lambda^{k+1/2} - u_\lambda\rangle
\end{aligned}
$$

Доказательство

- Работаем с

$$\begin{aligned}\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle &= \left\langle \begin{pmatrix} \nabla_x L(x^{k+1/2}, \lambda^{k+1/2}) \\ -\nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2}) \end{pmatrix}, \begin{pmatrix} x^{k+1/2} \\ \lambda^{k+1/2} \end{pmatrix} - \begin{pmatrix} u_x \\ u_\lambda \end{pmatrix} \right\rangle \\ &= \langle \nabla_x L(x^{k+1/2}, \lambda^{k+1/2}), x^{k+1/2} - u_x\rangle \\ &\quad + \langle -\nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2}), \lambda^{k+1/2} - u_\lambda\rangle\end{aligned}$$

- Выпуклость по x и вогнутость по λ:

$$\begin{aligned}\langle F(z^{k+1/2}), z^{k+1/2} - u\rangle &= \langle \nabla_x L(x^{k+1/2}, \lambda^{k+1/2}), x^{k+1/2} - u_x\rangle \\ &\quad + \langle -\nabla_\lambda L(x^{k+1/2}, \lambda^{k+1/2}), \lambda^{k+1/2} - u_\lambda\rangle \\ &\geq L(x^{k+1/2}, \lambda^{k+1/2}) - L(u_x, \lambda^{k+1/2}) \\ &\quad - L(x^{k+1/2}, \lambda^{k+1/2}) + L(x^{k+1/2}, u_\lambda)\end{aligned}$$

Доказательство

- Итого получаем:

$$2\gamma\left(L(x^{k+1/2}, u_\lambda) - L(u_x, \lambda^{k+1/2})\right) \leq \|z^k - u\|_2^2 - \|z^{k+1} - u\|_2^2$$

Доказательство

- Итого получаем:

$$2\gamma\left(L(x^{k+1/2},u_\lambda) - L(u_x,\lambda^{k+1/2})\right) \leq \|z^k - u\|_2^2 - \|z^{k+1} - u\|_2^2$$

- Суммируем по всем k от 0 до $K-1$ и делим на $2\gamma K$:

$$\frac{1}{K}\sum_{k=0}^{K-1}\left(L(x^{k+1/2},u_\lambda) - L(u_x,\lambda^{k+1/2})\right) \leq \frac{\|z^0 - u\|_2^2 - \|z^K - u\|_2^2}{2\gamma K} \leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

Доказательство

- Итого получаем:

$$2\gamma\left(L(x^{k+1/2}, u_\lambda) - L(u_x, \lambda^{k+1/2})\right) \leq \|z^k - u\|_2^2 - \|z^{k+1} - u\|_2^2$$

- Суммируем по всем k от 0 до $K-1$ и делим на $2\gamma K$:

$$\frac{1}{K}\sum_{k=0}^{K-1}\left(L(x^{k+1/2}, u_\lambda) - L(u_x, \lambda^{k+1/2})\right) \leq \frac{\|z^0 - u\|_2^2 - \|z^K - u\|_2^2}{2\gamma K}$$

$$\leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

- Неравенство Йесена для выпуклой и вогнутой функции дает:

$$\left(L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda\right) - L\left(u_x, \frac{1}{K}\sum_{k=0}^{K-1}\lambda^{k+1/2}\right)\right) \leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

Сходимость экстраградиентного метода

Теорема о сходимости экстраградиентного метода

Пусть дана непрерывно дифференцируемая по обеим группам переменным выпуклая-вогнутая L-гладкая функция $L : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$, тогда для экстраградиентного метода справедлива следующая оценка сходимости для любого $u \in \mathbb{R}^d \times \mathbb{R}^n$ и для любого $\gamma \leq \frac{1}{L}$:

$$\left(L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda\right) - L\left(u_x, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1/2}\right)\right) \leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

Сходимость экстраградиентного метода

Теорема о сходимости экстраградиентного метода

Пусть дана непрерывно дифференцируемая по обеим группам переменным выпуклая-вогнутая L-гладкая функция $L : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$, тогда для экстраградиентного метода справедлива следующая оценка сходимости для любого $u \in \mathbb{R}^d \times \mathbb{R}^n$ и для любого $\gamma \leq \frac{1}{L}$:

$$\left(L\left(\frac{1}{K} \sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda \right) - L\left(u_x, \frac{1}{K} \sum_{k=0}^{K-1} \lambda^{k+1/2} \right) \right) \leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

Подставим $\gamma = \frac{1}{L}$:

$$\left(L\left(\frac{1}{K} \sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda \right) - L\left(u_x, \frac{1}{K} \sum_{k=0}^{K-1} \lambda^{k+1/2} \right) \right) \leq \frac{L\|z^0 - u\|_2^2}{2K}$$

Сходимость экстраградиентного метода

Теорема о сходимости экстраградиентного метода

Пусть дана непрерывно дифференцируемая по обеим группам переменным выпуклая-вогнутая L-гладкая функция $L : \mathbb{R}^d \times \mathbb{R}^n \to \mathbb{R}$, тогда для экстраградиентного метода справедлива следующая оценка сходимости для любого $u \in \mathbb{R}^d \times \mathbb{R}^n$ и для любого $\gamma \leq \frac{1}{L}$:

$$\left(L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda\right) - L\left(u_x, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1/2}\right)\right) \leq \frac{\|z^0 - u\|_2^2}{2\gamma K}$$

Подставим $\gamma = \frac{1}{L}$:

$$\left(L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1/2}, u_\lambda\right) - L\left(u_x, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1/2}\right)\right) \leq \frac{L\|z^0 - u\|_2^2}{2K}$$

Остается вопрос: что подставлять в качестве $u = (u_x, u_\lambda)$?

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи.

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи. Рассмотрим самую простую седловую задачу $\min_x \max_\lambda (x-1)\cdot(\lambda+1)$. Решение этой задачи $x = 1, \lambda = -1, g(1,-1) = 0$. Тогда $L(x, \lambda^*) - L(x^*, \lambda) = 0$.

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи. Рассмотрим самую простую седловую задачу $\min_x \max_\lambda (x-1)\cdot(\lambda+1)$. Решение этой задачи $x=1, \lambda=-1, g(1,-1)=0$. Тогда $L(x,\lambda^*) - L(x^*,\lambda) = 0$. Такой критерий не подходит для выпукло-вогнутых седел, но подходит для сильно выпукло–сильно вогнутых. Но в сильно выпуклом–сильно вогнутом случае можно доказать линейную сходимость по аргументу, что более сильный результат.

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи. Рассмотрим самую простую седловую задачу $\min_x \max_\lambda (x-1)\cdot(\lambda+1)$. Решение этой задачи $x = 1, \lambda = -1, g(1,-1) = 0$. Тогда $L(x, \lambda^*) - L(x^*, \lambda) = 0$. Такой критерий не подходит для выпукло-вогнутых седел, но подходит для сильно выпукло–сильно вогнутых. Но в сильно выпуклом–сильно вогнутом случае можно доказать линейную сходимость по аргументу, что более сильный результат.
- Нужный вариант: $\max_\lambda L(x^k, \lambda) - \min_x L(x, \lambda^k)$.

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи. Рассмотрим самую простую седловую задачу $\min_x \max_\lambda (x-1)\cdot(\lambda+1)$. Решение этой задачи $x = 1, \lambda = -1, g(1,-1) = 0$. Тогда $L(x,\lambda^*) - L(x^*, \lambda) = 0$. Такой критерий не подходит для выпукло-вогнутых седел, но подходит для сильно выпукло–сильно вогнутых. Но в сильно выпуклом–сильно вогнутом случае можно доказать линейную сходимость по аргументу, что более сильный результат.
- Нужный вариант: $\max_\lambda L(x^k, \lambda) - \min_x L(x, \lambda^k)$.
- Если $\min_{x\in\mathcal{X}} \max_{\lambda\in\Lambda} (x-1)\cdot(\lambda+1)$ $\mathcal{X} = \mathbb{R}, \Lambda = \mathbb{R}$, то **вопрос:** что можно сказать про $\max_\lambda g(x^k, \lambda) - \min_x g(x, y^k)$?

Как измерять сходимость?

- Ожидаемый вариант: $L(x^k, \lambda^*) - L(x^*, \lambda^k)$, где x^* и λ^* – решение седловой задачи. Рассмотрим самую простую седловую задачу $\min_x \max_\lambda (x-1)\cdot(\lambda+1)$. Решение этой задачи $x = 1, \lambda = -1, g(1,-1) = 0$. Тогда $L(x, \lambda^*) - L(x^*, \lambda) = 0$. Такой критерий не подходит для выпукло-вогнутых седел, но подходит для сильно выпукло–сильно вогнутых. Но в сильно выпуклом–сильно вогнутом случае можно доказать линейную сходимость по аргументу, что более сильный результат.
- Нужный вариант: $\max_\lambda L(x^k, \lambda) - \min_x L(x, \lambda^k)$.
- Если $\min_{x\in\mathcal{X}} \max_{\lambda\in\Lambda} (x-1)\cdot(\lambda+1)$ $\mathcal{X} = \mathbb{R}, \Lambda = \mathbb{R}$, то **вопрос:** что можно сказать про $\max_\lambda g(x^k, \lambda) - \min_x g(x, y^k)$? $= +\infty$.

Как измерять сходимость?

- Поэтому критерий вида: $\max_{\lambda \in \Lambda} L(x^k, \lambda) - \min_{x \in \mathcal{X}} L(x, \lambda^k)$, формально можно рассматривать только на компактах.

Как измерять сходимость?

- Поэтому критерий вида: $\max_{\lambda \in \Lambda} L(x^k, \lambda) - \min_{x \in \mathcal{X}} L(x, \lambda^k)$, формально можно рассматривать только на компактах.
- С первого взгляда кажется, что это значит, что выпукло-вогнутые седловые задачи можно решать только на компактах (об этом говорила теорема Сиона-Какутани).

Как измерять сходимость?

- Поэтому критерий вида: $\max_{\lambda \in \Lambda} L(x^k, \lambda) - \min_{x \in \mathcal{X}} L(x, \lambda^k)$, формально можно рассматривать только на компактах.
- С первого взгляда кажется, что это значит, что выпукло-вогнутые седловые задачи можно решать только на компактах (об этом говорила теорема Сиона-Какутани).
- Такую же проблему встречали и в выпуклой минимизации (на $\mathbb{R}^d$ выпуклая задача может не иметь решения).

Как измерять сходимость?

- Поэтому критерий вида: $\max_{\lambda \in \Lambda} L(x^k, \lambda) - \min_{x \in \mathcal{X}} L(x, \lambda^k)$, формально можно рассматривать только на компактах.
- С первого взгляда кажется, что это значит, что выпукло-вогнутые седловые задачи можно решать только на компактах (об этом говорила теорема Сиона-Какутани).
- Такую же проблему встречали и в выпуклой минимизации (на $\mathbb{R}^d$ выпуклая задача может не иметь решения).
- В выпуклой минимизации предполагали, что решение существует.

Как измерять сходимость?

- Поэтому критерий вида: $\max_{\lambda \in \Lambda} L(x^k, \lambda) - \min_{x \in \mathcal{X}} L(x, \lambda^k)$, формально можно рассматривать только на компактах.
- С первого взгляда кажется, что это значит, что выпукло-вогнутые седловые задачи можно решать только на компактах (об этом говорила теорема Сиона-Какутани).
- Такую же проблему встречали и в выпуклой минимизации (на $\mathbb{R}^d$ выпуклая задача может не иметь решения).
- В выпуклой минимизации предполагали, что решение существует.
- Здесь можно сделать то же самое: предположим что решение существует и лежит в некотором компактном множестве $\mathcal{X}_* \times \Lambda_*$, тогда в критерии сходимости можно брать $\Lambda = \Lambda_*$ и $\mathcal{X} = \mathcal{X}_*$.

Экстраградиентный метод

- Имеет сходимость $1/K$ для выпукло-вогнутых гладких седловых задач.
- Можно добавить проекции и решать седловую задачу на множествах $\mathcal{X} \neq \mathbb{R}^d$ и $\Lambda \neq R^n$.
- Можно получить линейную сходимость для сильно выпуклых–сильно вогнутых задач.
- В случае, если целевая функция седловой задачи – это функция Лагранжа, то мы по факту переносим ограничения исходной задачи минимизации в целевую функцию седловой задачи. При этом теперь у седловой задачи будут простые ограничения: оставшиеся для x, которые не занесли в Лагранжиан, а также простые ограничения на $\lambda_i \geq 0$ (можно также добавить искусственно ограничения сверху на λ_i и на ν: $\lambda_i \leq A$ и $\nu_j \in [-B, B]$).

Прямо-двойственный алгоритм

Рассмотрим следующую задачу:

$$\min_{x \in \mathbb{R}^d} \max_{\lambda \in \mathbb{R}^n} L(x, \lambda) = f(x) - \lambda^T A x - g(\lambda),$$

где функция f – L_f-гладкая и выпуклая, а функция g – L_g-гладкая и выпуклая.

Прямо-двойственный алгоритм

Рассмотрим следующую задачу:

$$\min_{x \in \mathbb{R}^d} \max_{\lambda \in \mathbb{R}^n} L(x, \lambda) = f(x) - \lambda^T A x - g(\lambda),$$

где функция f – L_f-гладкая и выпуклая, а функция g – L_g-гладкая и выпуклая.

Примеры

- Минимизация с ограничениями вида равенств:

$$\min_{x \in \mathbb{R}^d} f(x)$$
$$\text{s.t. } Ax = b$$

Прямо-двойственный алгоритм

Рассмотрим следующую задачу:

$$\min_{x\in\mathbb{R}^d}\max_{\lambda\in\mathbb{R}^n} L(x,\lambda) = f(x) - \lambda^T Ax - g(\lambda),$$

где функция f – L_f-гладкая и выпуклая, а функция g – L_g-гладкая и выпуклая.

Примеры

- Минимизация с ограничениями вида равенств:

$$\min_{x\in\mathbb{R}^d} f(x)$$
$$\text{s.t. } Ax = b$$

Лагранж:

$$\min_{x\in\mathbb{R}^d}\max_{\lambda\in\mathbb{R}^n} L(x,\lambda) = f(x) - \lambda^T Ax.$$

Прямо-двойственный алгоритм

Примеры

- Линейная модель с регуляризатором:

$$\min_{x \in \mathbb{R}^d} f(x) + \ell(Ax)$$

Прямо-двойственный алгоритм

Примеры

- Линейная модель с регуляризатором:

$$\min_{x \in \mathbb{R}^d} f(x) + \ell(Ax)$$

- Заметим, что для самосопряженной функции $\ell(Ax) = \ell^{**}(Ax) = \max_\lambda \{(Ax)^T \lambda - \ell^*(\lambda)\}$, тогда

$$\min_{x \in \mathbb{R}^d} \max_{\lambda \in \mathbb{R}^n} f(x) - \lambda^T Ax - \ell^*(\lambda).$$

Прямо-двойственный алгоритм

Алгоритм 3 Прямо-двойственный алгоритм

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\quad x^{k+1} = x^k - \eta \left(\nabla f(x^k) - A^T \lambda^k \right)$

3: $\quad \lambda^{k+1} = \lambda^k - \eta \left(\nabla g(\lambda^k) + A(2x^{k+1} - x^k) \right)$

4: **end for**

Выход: $\frac{1}{K} \sum_{k=1}^{K} x^k$, $\frac{1}{K} \sum_{k=1}^{K} \lambda^k$

Прямо-двойственный алгоритм

Алгоритм 4 Прямо-двойственный алгоритм

Вход: стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

1: **for** $k = 0, 1, \ldots, K-1$ **do**

2: $\quad x^{k+1} = x^k - \eta\left(\nabla f(x^k) - A^T \lambda^k\right)$

3: $\quad \lambda^{k+1} = \lambda^k - \eta\left(\nabla g(\lambda^k) + A(2x^{k+1} - x^k)\right)$

4: **end for**

Выход: $\frac{1}{K}\sum_{k=1}^{K} x^k$, $\frac{1}{K}\sum_{k=1}^{K} \lambda^k$

Если вместо $(2x^{k+1} - x^k)$ подставить просто x^k получится просто спуск-подъем. В $(2x^{k+1} - x^k)$ зашита "экстраградиентность".

Доказательство

- Выпуклость и гладкость f:

$$
\begin{aligned}
L(x^{k+1}, \lambda^{k+1}) - L(x, \lambda^{k+1})& \\
=& f(x^{k+1}) - f(x) - (\lambda^{k+1})^T A(x^{k+1} - x) \\
=& f(x^{k+1}) - f(x^k) + f(x^k) - f(x) - (\lambda^{k+1})^T A(x^{k+1} - x) \\
\leq& \langle \nabla f(x^k), x^{k+1} - x^k \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2 \\
&+ \langle \nabla f(x^k), x^k - x \rangle - (\lambda^{k+1})^T A(x^{k+1} - x) \\
=& \langle \nabla f(x^k) - A^T \lambda^k, x^{k+1} - x \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2 \\
&- (\lambda^{k+1} - \lambda^k)^T A(x^{k+1} - x) \\
=& \eta^{-1} \langle x^k - x^{k+1}, x^{k+1} - x \rangle + \frac{L_f}{2} \|x^{k+1} - x^k\|_2^2 \\
&- (\lambda^{k+1} - \lambda^k)^T A(x^{k+1} - x)
\end{aligned}
$$

Доказательство

- Выпуклость и гладкость g:

$$
\begin{aligned}
L(x^{k+1}, \lambda) - L(x^{k+1}, \lambda^{k+1}) \\
&= g(\lambda^{k+1}) - g(\lambda) - (x^{k+1})^T A^T(\lambda - \lambda^{k+1}) \\
&= g(\lambda^{k+1}) - g(\lambda^k) + g(\lambda^k) - g(\lambda) - (x^{k+1})^T A^T(\lambda - \lambda^{k+1}) \\
&\leq \langle \nabla g(\lambda^k), \lambda^{k+1} - \lambda^k \rangle + \frac{L_g}{2} \|\lambda^{k+1} - \lambda^k\|_2^2 \\
&\quad + \langle \nabla g(\lambda^k), \lambda^k - x \rangle - (x^{k+1})^T A^T(\lambda - \lambda^{k+1}) \\
&= \langle \nabla g(\lambda^k) + A(2x^{k+1} - x^k), \lambda^{k+1} - \lambda \rangle + \frac{L_g}{2} \|\lambda^{k+1} - \lambda^k\|_2^2 \\
&\quad + (x^k - x^{k+1})^T A^T(\lambda^{k+1} - \lambda) \\
&= \eta^{-1} \langle \lambda^k - \lambda^{k+1}, \lambda^{k+1} - \lambda \rangle + \frac{L_g}{2} \|\lambda^{k+1} - \lambda^k\|_2^2 \\
&\quad + (x^k - x^{k+1})^T A^T(\lambda^{k+1} - \lambda)
\end{aligned}
$$

Доказательство

- С предыдущих слайдов:

$$L(x^{k+1},\lambda^{k+1}) - L(x,\lambda^{k+1}) \leq \eta^{-1}\langle x^k - x^{k+1}, x^{k+1} - x\rangle + \frac{L_f}{2}\|x^{k+1} - x^k\|_2^2 - (\lambda^{k+1} - \lambda^k)^T A (x^{k+1} - x)$$

$$L(x^{k+1},\lambda) - L(x^{k+1},\lambda^{k+1}) \leq \eta^{-1}\langle \lambda^k - \lambda^{k+1}, \lambda^{k+1} - \lambda\rangle + \frac{L_g}{2}\|\lambda^{k+1} - \lambda^k\|_2^2 + (x^k - x^{k+1})^T A^T (\lambda^{k+1} - \lambda)$$

- Суммируем:

$$L(x^{k+1},\lambda) - L(x,\lambda^{k+1}) \leq \begin{pmatrix} (x^k - x^{k+1})^T \\ (\lambda^k - \lambda^{k+1})^T \end{pmatrix}^T \begin{pmatrix} \frac{1}{\eta} & A \\ A & \frac{1}{\eta} \end{pmatrix} \begin{pmatrix} x^{k+1} - x \\ \lambda^{k+1} - \lambda \end{pmatrix} + \frac{L_g}{2}\|\lambda^{k+1} - \lambda^k\|_2^2 + \frac{L_f}{2}\|x^{k+1} - x^k\|_2^2$$

Доказательство

- Индуцированное скалярное произведение: $\langle x, Py \rangle = \langle x, y \rangle_P$ и норма $\|x\|_P^2 = \langle x, x \rangle_P$. У нас сейчас скалярное произведение вида (z – вектор из x и λ)

$$\langle z^k - z^{k+1}, z^{k+1} - z \rangle_P, \quad \text{где} \quad P = \begin{pmatrix} \frac{1}{\eta} & A \\ A & \frac{1}{\eta} \end{pmatrix}$$

- Ровно, как для обычного скалярного произведения:

$$\begin{aligned} L(x^{k+1}, \lambda) - L(x, \lambda^{k+1}) \leq & \langle z^k - z^{k+1}, z^{k+1} - z \rangle_P \\ & + \frac{\max(L_g, L_f)}{2} \|z^{k+1} - z^k\|_2^2 \\ = & \frac{1}{2}\|z^k - z\|_P^2 - \frac{1}{2}\|z^{k+1} - z\|_P^2 - \frac{1}{2}\|z^{k+1} - z^k\|_P^2 \\ & + \frac{\max(L_g, L_f)}{2} \|z^{k+1} - z^k\|_2^2. \end{aligned}$$

Доказательство

- Суммируем:

$$
\begin{aligned}
\sum_{k=0}^{K-1} &\left(L(x^{k+1}, \lambda) - L(x, \lambda^{k+1})\right) \\
&\leq \frac{1}{2}\|z^0 - z\|_P^2 - \frac{1}{2}\|z^K - z\|_P^2 \\
&\quad + \sum_{k=0}^{K-1}\left(\frac{\max(L_g, L_f)}{2}\|z^{k+1} - z^k\|_2^2 - \frac{1}{2}\|z^{k+1} - z^k\|_P^2\right).
\end{aligned}
$$

Доказательство

- Суммируем:

$$
\begin{aligned}
&\sum_{k=0}^{K-1} \left(L(x^{k+1}, \lambda) - L(x, \lambda^{k+1}) \right) \\
&\quad \leq \frac{1}{2}\|z^0 - z\|_P^2 - \frac{1}{2}\|z^K - z\|_P^2 \\
&\qquad + \sum_{k=0}^{K-1} \left(\frac{\max(L_g, L_f)}{2}\|z^{k+1} - z^k\|_2^2 - \frac{1}{2}\|z^{k+1} - z^k\|_P^2 \right).
\end{aligned}
$$

- **Вопрос:** что потребуем?

Доказательство

- Суммируем:

$$\sum_{k=0}^{K-1}\left(L(x^{k+1},\lambda) - L(x,\lambda^{k+1})\right)$$
$$\leq \frac{1}{2}\|z^0 - z\|_P^2 - \frac{1}{2}\|z^K - z\|_P^2$$
$$+ \sum_{k=0}^{K-1}\left(\frac{\max(L_g, L_f)}{2}\|z^{k+1} - z^k\|_2^2 - \frac{1}{2}\|z^{k+1} - z^k\|_P^2\right).$$

- **Вопрос:** что потребуем? $P \succ 0$ и $P - \max(L_g, L_f)I \succ 0$, чтобы "убить" последнюю строку и оставить $\|z^0 - z\|^2$.

Доказательство

- Суммируем:

$$\sum_{k=0}^{K-1}\left(L(x^{k+1},\lambda) - L(x,\lambda^{k+1})\right)$$
$$\leq \frac{1}{2}\|z^0 - z\|_P^2 - \frac{1}{2}\|z^K - z\|_P^2$$
$$+ \sum_{k=0}^{K-1}\left(\frac{\max(L_g, L_f)}{2}\|z^{k+1} - z^k\|_2^2 - \frac{1}{2}\|z^{k+1} - z^k\|_P^2\right).$$

- **Вопрос:** что потребуем? $P \succ 0$ и $P - \max(L_g, L_f)I \succ 0$, чтобы "убить" последнюю строку и оставить $\|z^0 - z\|^2$. Легко проверить, что это достигается с помощью $\eta \leq \frac{1}{\max(L_g,L_f)+\|A\|_2}$.

Доказательство

- Делим на K и получаем

$$\frac{1}{K}\sum_{k=0}^{K-1}\left(L(x^{k+1},\lambda) - L(x,\lambda^{k+1})\right) \leq \frac{1}{2K}\|z^0 - z\|_P^2.$$

Доказательство

- Делим на K и получаем

$$\frac{1}{K}\sum_{k=0}^{K-1}\left(L(x^{k+1},\lambda) - L(x,\lambda^{k+1})\right) \leq \frac{1}{2K}\|z^0 - z\|_P^2.$$

- Неравенство Йенсена для выпуклой и вогнутой функции дает

$$L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1},\lambda\right) - L\left(x,\frac{1}{K}\sum_{k=0}^{K-1}\lambda^{k+1}\right) \leq \frac{1}{2K}\|z^0 - z\|_P^2.$$

Сходимость прямо-двойственного метода

Сходимость прямо-двойственного метода

Если в билинейной седловой задаче функция f является выпуклой и L_f-гладкой, а функция g является вогнутой и L_g-гладкой, то прямо-двойственный метод имеет следующую оценку сходимости для любых $x \in \mathbb{R}^d$ и $u \in \mathbb{R}^n$

$$L\left(\frac{1}{K}\sum_{k=0}^{K-1} x^{k+1}, \lambda\right) - L\left(x, \frac{1}{K}\sum_{k=0}^{K-1} \lambda^{k+1}\right) \leq \frac{1}{2K}\|z^0 - z\|_P^2.$$

Здесь абсолютно эквивалетная ситуация с критерием сходимости, что была в обсуждении сходимости экстраградиентного метода.