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

Методы оптимизации

Александр Безносиков

Московский физико-технический институт

19 октября 2023

Лагранжиан

Рассматриваем задачу условной оптимизации вида:

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

Лагранжиан

Рассматриваем задачу условной оптимизации вида:

$$\min_{x\in\mathbb{R}^d X} f_0(x)$$

$$\text{s.t. } f_i(x) \le 0, \ i = 1, \ldots, m$$

Здесь еще можно было немного обобщить постановку и добавить, что $x \in \mathcal{X} \cap \mathrm{dom} f_i$. Но мы предполагаем, что $\mathcal{X} \cap \underline{\mathrm{dom} f_i} = \mathbb{R}^d$.

Лагранжиан

Лагранжиан

Функция Лагранжа/Лагранжиан для этой задачи строится следующим образом:

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

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

Что делали на семинаре:

Что делали на семинаре:

- Рассматривали:
$$g(\lambda) = \inf_{x \in \mathbb{R}^d} L(x, \lambda).$$
Вопрос: как называется этот объект?

Что делали на семинаре:

- Рассматривали:

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

Вопрос: как называется этот объект? двойственная функция

Что делали на семинаре:

- Рассматривали:

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

Вопрос: как называется этот объект? двойственная функция

- Осознали, что для любой $\lambda \geq 0$

$$g(\lambda) \leq f(x^*).$$

λ^*

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

$$f_i(x) < 0, \quad i = 1, \ldots, m.$$

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

$$f_i(x) < 0, \;\; i = 1, \ldots, m.$$

- **Вопрос:** и что оно дает?

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

$$f_i(x) < 0, \;\; i = 1, \ldots, m.$$

- **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

$$\sup_{\lambda \geq 0} g(\lambda) = f(x^*).$$

Седловая точка

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

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

Теорема о седловой точке Куна-Таккера

f_0

$f_i \leq 0$

f_0, f_i - вып

Теорема о седловой точке Куна-Таккера

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

- для x^* существует $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа,
- для x^* – глобальное решение задачи оптимизации с ограничениями.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что (x^*, λ^*) – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

x^* – удовл. огран.? $\exists\, x^*$ не удовл. $\exists i$ $f_i(x^*) > 0$

$\lambda_j \geq 0$

$$L(x^*, \lambda) = f_0(x^*) + \sum_{j=1}^{m} \lambda_j f_i(x^*)$$

λ_i – варьируем

$f_i(x^*) > 0$ $\quad \lambda_i f_i(x^*) > 0$ $\quad \lambda_i \uparrow$

$$\sup_{\lambda \geq 0} L(x^*; \lambda) = +\infty$$

$$L(x^*, \lambda^*) \geq L(x^*, \lambda^*) \geq L(x^*; \lambda) \quad \forall \lambda \geq 0$$

$\lambda = \lambda^* \to \lambda_i^* = 2\lambda_i^*$

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$.

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$. ∎
- Заметим, что $f_0(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$.

x^* - удовл. огран.

$$\sup_{\lambda \geq 0} \boxed{f_0(x^*)} + \underbrace{\sum \lambda_j \underbrace{g_j(x^*)}_{\leq 0}}_{0}$$

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$.
- Заметим, что $f_0(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*) = f_0(x^*)$.

$$f(x^*) = \sup_{\lambda \geq 0} = L(x^*, \lambda^*) \geq L(x^*, \lambda)$$

λ^* дает супремум

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \ge 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \ge 0} L(x^*, \lambda)$? $\sup_{\lambda \ge 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$.
- Заметим, что $f(x^*) = \sup_{\lambda \ge 0} L(x^*, \lambda)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*) = f(x^*)$. Первое неравенство из определения седловой задачи дает:

$$\forall x \in X \quad f_0(x) \ge \underbrace{f_0(x) + \sum_{j=1}^m \underbrace{\lambda_j^*}_{\ge 0} \underbrace{f_j(x)}_{\le 0}}_{\le 0} \ge \cancel{L(x^*, \lambda^*)} = f_0(x^*).$$

огран., $f_i \le 0$

усл.

≤ 0

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$.
- Заметим, что $f(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*) = f(x^*)$. Первое неравенство из определения седловой задачи дает:

$$f_0(x) + \sum_{j=1}^m \lambda_j^* f_j(x) \geq L(x^*, y^*) = f_0(x^*).$$

А это и есть то, что мы хотели. **Вопрос:** почему?

Доказательство

$\Rightarrow$ Пусть для x^* существует $\lambda^* \geq 0$ такая, что x^*, λ^* – седловая точка функции Лагранжа, тогда x^* – глобальное решение задачи с ограничениями.

- Для начала удостоверимся, что x^* удовлетворяет ограничениям. Если нет, то $h_j(x^*) > 0$ для некоторого j. **Вопрос:** что можно сказать про $\sup_{\lambda \geq 0} L(x^*, \lambda)$? $\sup_{\lambda \geq 0} L(x^*, \lambda) = +\infty$. **Вопрос:** может ли такое быть? Нет, второе неравенство в определении седловой точки рушится для $\lambda = 2\lambda^*$.
- Заметим, что $f(x^*) = \sup_{\lambda \geq 0} L(x^*, \lambda)$. Второе неравенство в определении седловой задачи дает $L(x^*, \lambda^*) = f(x^*)$. Первое неравенство из определения седловой задачи дает:

$$f_0(x) + \sum_{j=1}^{m} \lambda_j^* f_j(x) \geq L(x^*, y^*) = f_0(x^*).$$

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

$$f(x^*) = \sup_{\lambda \geq 0} g(\lambda)$$

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – 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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной $\lambda^* \geq 0$, $f(x^*) = g(\lambda^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*)$.

sup

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной $\lambda^* \geq 0$, $f(x^*) = g(\lambda^*) = \inf_{x \in \mathbb{R}^d} L(x, \cancel{\lambda^*})$. Откуда $\cancel{f(x^*)} \leq L(x^*, \lambda^*) = f(\cancel{x^*}) + \underbrace{\sum_{j=1}^{m} \lambda_j^* f_j(x^*)}_{\leq 0}$. $= 0$

$\lambda \geq 0$

0

услов

$f_i \leq 0$

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

- **Вопрос:** что нам дает условие Слейтера, для выпуклой оптимизации? для решения двойственной $\lambda^* \geq 0$, $f(x^*) = g(\lambda^*) = \inf_{x \in \mathbb{R}^d} L(x, \lambda^*)$. Откуда $f(x^*) \leq L(x^*, \lambda^*) = f(x^*) + \sum_{j=1}^m \lambda_j^* f_j(x^*)$. **Вопрос:** что можем сказать про $\lambda_j^* f_j(x^*)$? все 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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

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

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

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

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

$L(x^*, \lambda^*)$
$= f(x^*) \leq L(x, \lambda^*)$

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 $\lambda^* \geq 0$ такое, что (x^*, λ^*) – седловая точка функции Лагранжа.

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

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

Вторая часть получается из того $f_j(x^*) \leq 0$, а значит $f(x^*) \geq f(x^*) + \underbrace{\sum_{j=1}^m \lambda_j^* f_j(x^*)}_{=0} \cancel{\geq} L(x^*, \lambda)$ для $\lambda_j \geq 0$.

$\forall \lambda \quad \leq 0$

Седловая задача и игра

Седловые задачи – это больше, чем просто функция Лагранжа.

Седловая задача и игра

Седловые задачи – это больше, чем просто функция Лагранжа.
Пусть есть некоторая функция:

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

Седловая задача и игра

Седловые задачи – это больше, чем просто функция 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 $x \in \mathcal{X}$ и втор $\lambda \in \Lambda$ (это может быть распределение ресурсов, выбор действие и т.д.).

Седловая задача и игра

Седловые задачи – это больше, чем просто функция 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 $x \in \mathcal{X}$ и $\lambda \in \Lambda$ (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 игра

Седловые задачи – это больше, чем просто функция 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 $x \in \mathcal{X}$ и $\lambda \in \Lambda$ (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 игра

Седловые задачи – это больше, чем просто функция 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 $x \in \mathcal{X}$ и $\lambda \in \Lambda$ (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 игра

- 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 игра

- 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 игра

- 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 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 игра

- 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 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 игра

- **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 игра

- **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 $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$

Седловая задача и игра

- **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 $\inf_x \sup_\lambda L(x, \lambda)$». 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 игра

- **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 $\inf_x \sup_\lambda L(x, \lambda)$». Противоположная ситуация во втором случае $\sup_\lambda \inf_x L(x, \lambda)$
- Используя эту интуицию можно понять, что в общем случае, что

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

- Формально:

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

Седловая задача и игра

- **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 $\inf_x \sup_\lambda L(x, \lambda)$». 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 игра

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

Седловая задача и игра

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.

min max $\Updownarrow$ max min

$\exists$ реш $\Rightarrow$ седл. задачу

$g(\lambda) = \inf_x L(x,\lambda)$

$\sup g(\lambda) = \sup \inf L(x,\lambda)$

Седловая задача и игра

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) y. Тогда L имеет седловые точки на $\mathcal{X} \times \Lambda$.

$f^* > -\infty$

$\|x^0 - x^*\|^2$

$\uparrow$

$-\infty$

Седловая задача и игра

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) y. 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) y. Тогда (гарантий существования тут нет)

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

Ищем решение седловой задачи

- Оптимизация функции Лагранжа — седловая задача.

Ищем решение седловой задачи

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

Ищем решение седловой задачи

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

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

Вопрос: как решали первое? Может быть поможет решать второе.

$$x^{k+1} = x^k - \gamma \nabla f(x^k)$$

$$x^{k+1} = x^k - \gamma \nabla_x f(x^k, \lambda^k)$$

$$\lambda^{k+1} = \lambda^k + \gamma \nabla_y f(x^k, \lambda^k)$$

Ищем решение седловой задачи

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

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

  Вопрос: как решали первое? Может быть поможет решать второе.

- Градиентный спуск-подъем:

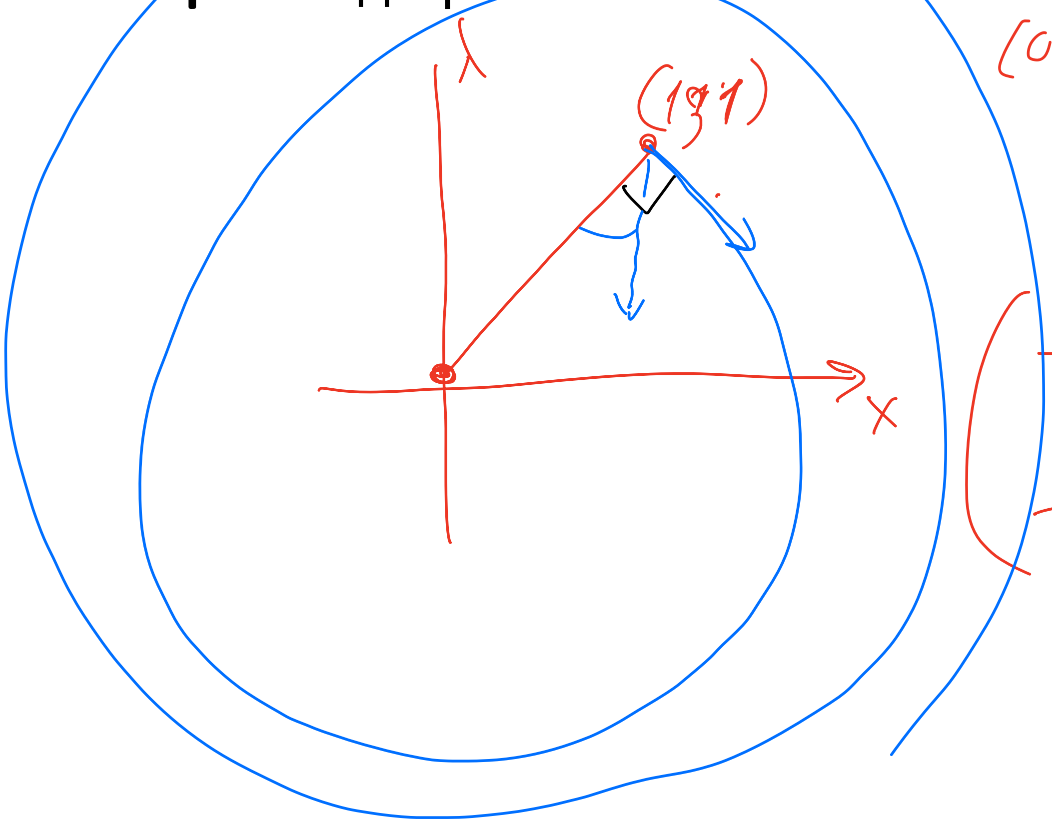
$$\begin{pmatrix} x^{k+1} \\ \cancel{x}^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \end{pmatrix} - \gamma \begin{pmatrix} \nabla_x f(x^k, \lambda^k) \\ -\nabla_\lambda f(x^k, \lambda^k) \end{pmatrix}$$

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?



$(0;0)$

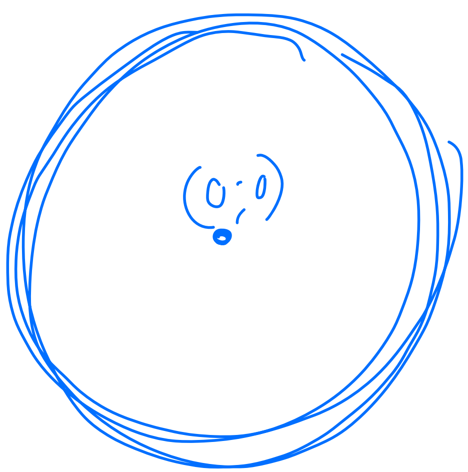
$x = 0$

$\lambda = 0$

$$\begin{pmatrix} +\nabla_x f(x^0, \lambda^0) \\ -\nabla_\lambda f(x^0, \lambda^0) \end{pmatrix} = \begin{pmatrix} \lambda \\ -x \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

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 $\binom{\nabla_x f(x^k, \lambda^k)}{-\nabla_\lambda f(x^k, \lambda^k)}$ 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 $\binom{\nabla_x f(x^k, \lambda^k)}{-\nabla_\lambda f(x^k, \lambda^k)}$ 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 $\binom{\nabla_x f(x^k, \lambda^k)}{-\nabla_\lambda f(x^k, \lambda^k)}$ всегда ортогонален направлению на решение $\binom{x^k - x^*}{\lambda^k - \lambda^*}$. **Вопрос:** что это значит? Метод не стремится к решению.
- Интуиция не является сторогой, но может подсказать, что нужно попробовать что-то чуть-чуть другое.

Метод экстраградиента

ExtraGradient

Алгоритм 1 ~~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: $x^{k+1/2} = x^k - \eta \nabla_x f(x^k, y^k)$

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

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

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

6: **end for**

Выход: x^K

Метод экстраградиента

Алгоритм 2 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: $\quad x^{k+1/2} = x^k - \eta \nabla_x f(x^k, y^k)$
3: $\quad \lambda^{k+1/2} = \lambda^k + \eta \nabla_\lambda f(x^k, \lambda^k)$
4: $\quad x^{k+1} = x^k - \eta \nabla_x f(x^{k+1/2}, \lambda^{k+1/2})$
5: $\quad \lambda^{k+1} = \lambda^k + \eta \nabla_\lambda f(x^{k+1/2}, \lambda^{k+1/2})$
6: **end for**

Выход: x^K

$\begin{pmatrix} \lambda \\ -x \end{pmatrix} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}$ $\qquad \begin{pmatrix} \eta \\ -\eta \end{pmatrix}$

$\eta < 1$

$\begin{pmatrix} x^1 \\ \lambda^1 \end{pmatrix} = \begin{pmatrix} 1-\eta+\eta^2 \\ 1-\eta+\eta^2 \end{pmatrix} \qquad \nabla_{1/2} = \begin{pmatrix} 1-\eta \\ \eta-1 \end{pmatrix} \qquad \begin{pmatrix} x^{1/2} \\ \lambda^{1/2} \end{pmatrix} = \begin{pmatrix} 1-\eta \\ 1-\eta \end{pmatrix}$

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.

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

$$z^{k+1/2} = z^k - \eta F(z^k)$$

$$z^{k+1} = z^k - \eta F(z^{k+1/2})$$

$$\|z^{k+1} - z\|_2^2 = \|z^k - \eta F(z^{k+1/2}) - z\|^2 =$$

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

$$-2\eta \langle F(z^{k+1/2}); z^{k+1} - z\rangle \qquad \|z^{k+1} - z^k\|_2^2$$

$$-2\eta \langle F(z^{k+1/2}); z^k - z^{k+1}\rangle$$

$$\| $$

$$2 \langle z^{k+1} - z^k; z^k - z^{k+1}\rangle$$

$$\| $$

$$-2\|z^k - z^{k+1}\|_2^2$$

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

$$\|z^{k+1/2} - u\|_2^2 = \|z^k - \eta F(z^k) - u\|_2^2$$

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

$$\|z^{k+1} - z\|_2^2 = \|z^k - z\|_2^2 - \underline{2\eta \langle F(z^{k+1/2}); z^{k+1} - z\rangle} - \cancel{\|z^{k+1} - z^k\|_2^2}$$

$$+ \|z^{k+1/2} - z^{k+1}\|_2^2$$

$$+ \cancel{\|z^k - z^{k+1}\|^2} - \underline{2\eta \langle F(z^k), z^{k+1/2} - z^{k+1}\rangle} - \|z^{k+1/2} - z^k\|_2^2$$

$$- \Big(\langle F(z^{k+1/2}); z^{k+1/2} - z\rangle + \langle F(z^{k+1/2}); \underline{z^{k+1} - z^{k+1/2}}\rangle + \langle F(z^k); z^{k+1/2} - \underbrace{u}_{z^{k+1}}\rangle \Big)$$

$$= - \Big(\langle F(z^{k+1/2}); z^{k+1/2} - z\rangle + \langle F(z^{k+1/2}) - F(z^k); z^{k+1} - z^{k+1/2}\rangle \Big)$$

CBS

$$\|z^{k+1} - z\|_2^2 + \cancel{\|z^{k+1/2} - z^{k+1}\|_2^2} \le \|z^k - z\|_2^2 - \|z^{k+1/2} - z^k\|_2^2 - 2\eta \langle F(z^{k+1/2}); z^{k+1/2} - z\rangle + \eta^2 \|F(z^{k+1/2}) - F(z^k)\|^2 + \cancel{\|z^{k+1} - z^{k+1/2}\|^2}$$

$\nabla_x f(x,y)$ – L-Lipsch. $\nabla_y f(x,y)$ – L-Lip.

$$\|F(z_1) - F(z_2)\|_2^2 \le L^2 \cdot \|z_1 - z_2\|_2^2$$

$$f(x,\cancel{y}^*) - f(x^*,\cancel{y}) = 0$$

$$f(x,\lambda^*) \qquad f(x^*,\lambda)$$

$\min_{\mathbb{R}} \max_{\mathbb{R}} (x-1)(y+1)$

$(1,-1)$

$$\max_{x\in C} f\left(\frac{1}{k}\sum x^{k+1/2}, y\right) - \min f\left(x, \frac{1}{k}\sum y^{k+1/2}\right) \le \frac{\|z^0 - z\|_2^2}{2k}$$

$$\|z^{k+1} - z\|_2^2 \le \|z^k - z\|_2^2 \cancel{- \|z^{k+1/2} - z^k\|_2^2}$$
$$- 2\eta \langle F(z^{k+1/2}), z^{k+1/2} - z\rangle$$
$$\cancel{+ \eta^2 L^2 \|z^{k+1/2} - z^k\|_2^2}$$

$$\eta \le \frac{1}{L}$$

$$\langle F(z^{k+1/2}), z^{k+1/2} - z\rangle \le \frac{\|z^k - z\|_2^2 - \|z^{k+1} - z\|_2^2}{2\eta}$$

∇f gro on.

$$f(x^{k+1/2}, y) - f(x, y^{k+1/2}) \le \frac{\|z^k - z\|_2^2 - \|z^{k+1} - z\|_2^2}{2\eta}$$

$$\frac{1}{k}\sum f(x^{k+1/2}, y) - f(x, y^{k+1/2}) \le \frac{\|z^0 - z\|_2^2}{2\eta k} \cancel{- \|z^k - z\|_2^2}$$

Прямо-двойственный алгоритм

Рассмотрим специализированную 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 специализированную 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:

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

Прямо-двойственный алгоритм

Рассмотрим специализированную 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

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

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: x^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: x^K

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

Доказательство

- Выпуклость и гладкость f:

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

Доказательство

- Выпуклость и гладкость g:

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

Доказательство

- С предыдущих слайдов:

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

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

- Суммируем:

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

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)

$$\langle z^k - z^{k+1}, z^{k+1} - z \rangle_P.$$

- 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:

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

Доказательство

- Суммируем:

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

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

Доказательство

- Суммируем:

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

- **Вопрос:** что потребуем? $P > 0$ и $P - \max(L_g, L_f)I > 0$, 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 любого $x \in \mathbb{R}^d$

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