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

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

$$Ax = b \qquad A \in \mathbb{R}^{n \times d} \qquad b \in \mathbb{R}^n$$

NB: $x \in X \cap \operatorname{dom} f_i$ (где $X \cap \operatorname{dom} f_i = \mathbb{R}^d$)

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.

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

NB: $g(\lambda, \nu) \le f_0(x^*) \quad \forall \lambda \succeq 0 \quad \nu \in \mathbb{R}^n$

Условие Слейтера

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

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

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

$$\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} g(\lambda, \nu) = f_0(x^*).$$

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

f_0 - в.
f_i - в.

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, $\Downarrow$ – важный
- x^* – глобальное решение задачи оптимизации с ограничениями.

1) $\Rightarrow$ (x^*, λ^*, ν^*) – с.т., тогда $\overset{?}{}$ x^* – решение исх. задачи

a) x^* – удовл. огранич.

от противного: $\exists i : f_i(x^*) > 0$

$$\sup_{\lambda \succeq 0, \nu \in \mathbb{R}^n} L(x^*, \lambda, \nu) = +\infty$$

$$\shortparallel$$

$$f_0(x^*) + \underset{\to +\infty}{\lambda_i} \underset{>0}{f_i(x^*)} + \sum_{\delta e_3 i} + \text{мн. ор.}$$

$$L(x^*, \lambda^*, \nu^*) \geq L(x^*, \lambda, \nu) \quad \forall \lambda \succeq 0 \quad \forall \nu \in \mathbb{R}^n$$

$$\exists \ \lambda' = \begin{pmatrix} \vdots \\ 2\lambda_i^* \\ \vdots \end{pmatrix}$$ противоречие

$\boxed{\text{значит } x^* \text{ – удовл. огранич.}}$

б). $f_0(x^*) \overset{?}{\leq} f_0(x) \quad \forall x$ удовл. огран

$$L(x^*, \lambda, \nu) = f_0(x^*) + \sum \underset{\geq 0}{\lambda_i} \underbrace{f_i(x^*)}_{\leq 0} + \underbrace{(\cancel{Ax^* - b})^T \nu}_{=0}$$

$$L(x^*, \lambda^*, \nu^*) \underset{\text{седловая точка}}{\geq} L(x^*, \lambda, \nu)$$

$$f_0(x^*) = \sup L(x^*, \lambda, \nu) = L(x^*, \lambda^*, \nu^*)$$

$$f_0(x^*) = L(x^*, \lambda^*, \nu^*)$$

$$f_0(x) + \sum \lambda_j^* f_j(x) + \nu_*^T (Ax - b) = L(x, \lambda^*, \nu^*)$$

из седл. точки: $L(x, \lambda^*, \nu^*) \geq L(x^*, \lambda^*, \nu^*)$
$= f_0(x^*)$

$$f_0(x) + \sum \lambda_j^* f_j(x) + \nu_*^T (Ax - b) = L(x, \lambda^*, \nu^*) \geq f_0(x^*)$$

Если x удовл. огранич., то $Ax - b = 0$

$$f_0(x) \geq f_0(x) + \underbrace{\sum \underbrace{\lambda_j^*}_{\geq 0} \underbrace{f_j(x)}_{\leq 0}}_{\leq 0} \geq f_0(x^*)$$

$$\boxed{f_0(x) \geq f_0(x^*)}$$ $\forall x$ удовл. огран.

2) $\Leftarrow$ будет в презентации

Забыли про лагранжиан:

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

I: $x \in X$ действие
II: $\lambda \in \Lambda$ действие

$\Rightarrow L(x, \lambda)$ – "выигрыш" II игрока

II выиграть больше

I проиграть меньше

Безразлично: $\exists\, \tilde\lambda$ для нек $\tilde x$: $\tilde\lambda = \sup L$

$\tilde x$ может быть никогда не выбрано

Более умно: хотим (x^*, λ^*):

$$L(x,\lambda^*) \ge L(x^*,\lambda^*) \ge L(x^*,\lambda)$$

игрок I не выгодно менять стр. x^*

игрок II невыгодно

Важен ли порядок выбора:

1) II выбрал $\tilde\lambda$

I игрок действует $\inf\limits_x L(x,\tilde\lambda)$

тогда как рассуждает игрок II при своем выборе:

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

одинаковы?

2) обратное:

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

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

$$\inf_{\tilde x} L(\tilde x,\lambda) \le L(x,\lambda) \Rightarrow \sup_\lambda \inf_{\tilde x} L(\tilde x,\lambda) \le \sup_\lambda L(x,\lambda)$$

$$\Downarrow$$

$$\sup\inf L(x,\lambda) \le \inf\sup L(x,\lambda)$$

sup inf и inf sup связаны с наличием с.т.

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.

(x^*, λ^*) — искомая $\Leftrightarrow$ $\inf \sup L(x,\lambda)$, $\sup \inf L(x,\lambda)$

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)

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

$\inf\limits_{x} \sup\limits_{\lambda,\nu}$ или $\sup\limits_{\lambda,\nu} \inf\limits_{x} L(x,\lambda,\nu)$, когда найдем с.т. $L(x,\lambda,\nu)$

тогда найдем решение x^* исходной задачи

$\min\limits_{x} \max\limits_{\lambda,\nu}$ или $\max\limits_{\lambda,\nu} \min\limits_{x}$

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

Простая идея: град. спуск $\Rightarrow$ град. спуска-подъема

$$x^{k+1} = x^k - \gamma \nabla_x L(x^k, \lambda^k) \quad \leftarrow \text{спуск}$$

$$\lambda^{k+1} = \lambda^k + \gamma \nabla_\lambda L(x^k, \lambda^k) \quad \leftarrow \text{подъем}$$

x^{k+1} чуть лучше по памяти

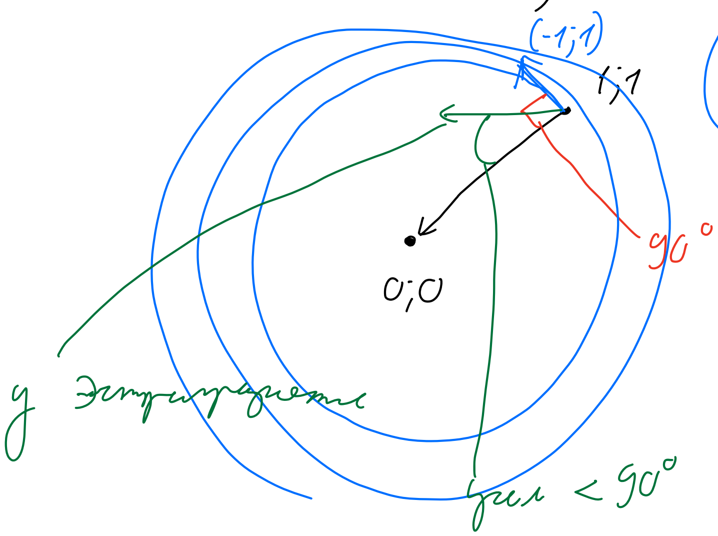
Пример: $\min\limits_{x \in \mathbb{R}} \max\limits_{\lambda \in \mathbb{R}} x\lambda$

$\nabla_x = \lambda$

$\nabla_\lambda = x$

$(x^*, \lambda^*) = (0;0) \leftarrow \nabla = 0$

$(x^0; \lambda^0) = (1; 1)$



$$\begin{pmatrix} -\nabla_x \\ \nabla_\lambda \end{pmatrix} = \begin{pmatrix} -\lambda \\ x \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$$

Г. Корпелевич „взгляд в буд.“

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

---

1) похоже на моментум

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

неявная схема

прокс. метод

экстраполяция прокса

$x^{k+1} \to x^{k+1/2}$

$x^{k+1} = \arg\min \left\{ \gamma f(x) + \frac{1}{2}\|x - x^k\|_2^2 \right\}$

$= \underline{\mathrm{prox}_{\gamma f}(x^k)}$

$\gamma \nabla f(x) + x - x^k \Big|_{x^{k+1}} = 0$

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

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

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

- Почему не $u_\lambda = \lambda^*$ $\quad u_x = x^*$

$$\min_x \max_\lambda L(x,\lambda) = x\lambda \qquad x^* = 0, \quad \lambda^* = 0$$

$$L(x^*, \lambda) = 0 \qquad L(x, \lambda^*) = 0$$

$$L(x^k, \lambda^*) - L(x^*, \lambda^k) = 0$$

- Используем

$$\max_\lambda L(x^k, \lambda) - \min_x L(x, \lambda^k)$$