• Оптимизация на „простых" множествах

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

Условие оптимальности

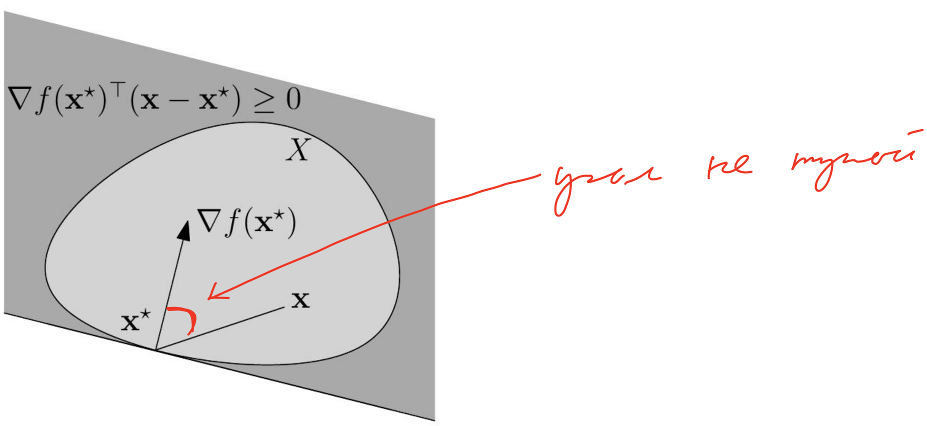
- f – непр. дифф. на $\mathbb{R}^d$
- f – выпуклая
- X – выпуклое

тогда $x^* \in X$ – глоб. минимум $f(x)$ на X

$$\Longleftrightarrow$$

$$\langle \nabla f(x^*); x - x^* \rangle \geq 0 \quad \forall x \in X$$

Физический смысл:



- x^* (безусловный)

Док-во:

- достаточность. $\Leftarrow$

$$\langle \nabla f(x^*); x - x^* \rangle \geq 0 \quad \forall x \in X$$

выпуклость f:

$$f(x) \geq f(x^*) + \underbrace{\langle \nabla f(x^*); x - x^* \rangle}_{\geq 0} \geq f(x^*) \quad \forall x \in X$$

x^* – глобальный минимум на X

- необходимость $\Rightarrow$

x^* – глобальный минимум на X

от противного: $\exists \tilde{x} \in X : \langle \nabla f(x^*), \tilde{x} - x^* \rangle < 0$

$$x_\lambda = \lambda \tilde{x} + (1-\lambda) x^* \qquad \lambda \in [0;1]$$

$\in X$

$$\varphi(\lambda) = f(x_\lambda) = f(\lambda \tilde{x} + (1-\lambda) x^*)$$

$$\frac{d\varphi}{d\lambda} = \frac{d}{d\lambda}\Big(f(\lambda(\tilde{x}-x^*)+x^*)\Big) = \langle \nabla f(\lambda(\tilde{x}-x^*)+x^*); \tilde{x}-x^* \rangle$$

$$\left.\frac{d\varphi}{d\lambda}\right|_{\lambda=0} = \langle \nabla f(x^*), \tilde{x}-x^* \rangle \underbrace{<0}_{\text{по предположению}}$$

φ убывает в окр. 0, а значит $\exists \hat{\lambda} > 0$:

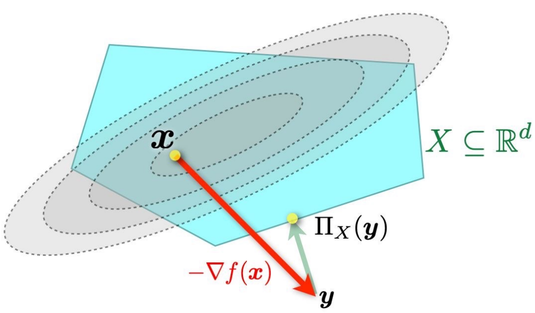
$$f(\underbrace{x^* + \hat{\lambda}(\tilde{x}-x^*)}_{\neq x^*}) = \varphi(\hat{\lambda}) < \varphi(0) = f(x^*)$$

пришли к противоречию, что x^* – гл. мин ■

- Метод град. спуска с проекцией

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

$x^k \in X$, не значит $x^{k+1} \in X$



$$\boxed{\begin{aligned} x^{k+1} &= \Pi\left(x^k - \gamma \nabla f(x^k)\right) \\ \Pi(y) &= \operatorname*{argmin}_{x \in X} \|x-y\|_2^2 \end{aligned}}$$

$\leftarrow$ евклидова проекция

• Св-ва проекции:

0) Проекция однозначна (сильная выпуклость задачи $\|x-y\|_2^2$)

1) X – выпуклое, $x \in X$, $y \in \mathbb{R}^d$, тогда

$$\langle x - \Pi(y);\ y - \Pi(y) \rangle \le 0$$

Док-во: $$\Pi(y) = \underset{z \in \underbrace{X}_{\text{выпуклое}}}{\operatorname{argmin}} \Big[d(z) := \underbrace{\|z-y\|_2^2}_{\text{выпуклая}} \Big]$$

Условие оптимальности для $d(z)$, X и решения $\Pi(y)$

$$\langle \nabla d(z^*);\ z - z^* \rangle \ge 0 \quad \forall z \in X$$

$z^* = \Pi(y)$, $z = x$

$$\langle \nabla d(\Pi(y));\ x - \Pi(y) \rangle \ge 0$$

$$\nabla d(z) = 2(z-y)$$

$$2 \langle \Pi(y) - y;\ x - \Pi(y) \rangle \ge 0 \quad \blacksquare$$

2) Нерасширяемость оператора проекции

X – выпуклое мн-во, $x_1, x_2 \in \mathbb{R}^d$, тогда

$$\|\Pi(x_1) - \Pi(x_2)\|_2 \le \|x_1 - x_2\|_2$$

Док-во: св-во 1) $y = x_1$ $x = \Pi(x_2)$

$$\langle \underbrace{\Pi(x_2) - \Pi(x_1)}; \ x_1 - \Pi(x_1) \rangle \le 0$$

аналогично $y = x_2$ $x = \Pi(x_1)$

$$\langle \underbrace{\Pi(x_1) - \Pi(x_2)}; \ x_2 - \Pi(x_2) \rangle \le 0$$

складываем

$$\langle \Pi(x_2) - \Pi(x_1);\ x_1 - \Pi(x_1) - x_2 + \Pi(x_2) \rangle \le 0$$

$$\underbrace{\langle \Pi(x_2)-\Pi(x_1);\ \Pi(x_2)-\Pi(x_1)\rangle}_{\|\Pi(x_2)-\Pi(x_1)\|_2^2} \le \underbrace{\langle x_2-x_1;\ \Pi(x_2)-\Pi(x_1)\rangle}_{\text{КБШ}}$$

$$\|\Pi(x_2)-\Pi(x_1)\|_2^{\cancel{2}} \le \|x_2-x_1\|_2 \cdot \cancel{\|\Pi(x_2)-\Pi(x_1)\|_2}$$

$$\|\Pi(x_2)-\Pi(x_1)\|_2 \le \|x_2-x_1\|_2 \quad \blacksquare$$

3) Стационарная точка град. спуска с проекцией

$$x^* = \Pi\left(x^*-\gamma\nabla f(x^*)\right)$$

Док-во:

$$\Pi\left(x^*-\gamma\nabla f(x^*)\right) = \arg\min_{x\in X}\left[\|x-x^*+\gamma\nabla f(x^*)\|_2^2\right]$$

$$= \arg\min_{x\in X}\Big[\underbrace{\underbrace{\|x-x^*\|_2^2}_{\ge 0} + 2\underbrace{\gamma}_{>0}\underbrace{\langle\nabla f(x^*); x-x^*\rangle}_{\ge 0 \text{ по усл.}}}_{\ge 0} + \cancel{\gamma^2\|\nabla f(x^*)\|_2^2}\Big]$$

потому что argmin

0 под argmin достигается в $x = x^*$ $\blacksquare$

- Док-во сходимости:

$$\|x^{k+1}-x^*\|_2^2 = \|\Pi(x^k-\gamma\nabla f(x^k)) - x^*\|_2^2$$

3-е св-во проекции

$$= \|\Pi(x^k-\gamma\nabla f(x^k)) - \Pi(x^*-\gamma\nabla f(x^*))\|_2^2$$

2-е св-во проекции

$$\le \|x^k-x^* - \gamma(\nabla f(x^k)-\nabla f(x^*))\|_2^2$$

$$= \|x^k - x^*\|_2^2 - 2\gamma \underbrace{\langle \nabla f(x^k) - \nabla f(x^*); x^k - x^* \rangle}_{\mu \text{ сильная выпуклость}} + \gamma^2 \underbrace{\|\nabla f(x^k) - \nabla f(x^*)\|_2^2}_{L\text{- гладкость}}$$

$$\le \|x^k - x^*\|_2^2 + 2\gamma \langle \nabla f(x^*); x^k - x^* \rangle - 2\gamma\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right) + 2L\gamma^2\left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^* \rangle\right)$$

$$= (1-\gamma\mu)\|x^k - x^*\|_2^2 + 2\gamma(\gamma L - 1)\underbrace{\left(f(x^k) - f(x^*) - \langle \nabla f(x^*); x^k - x^* \rangle\right)}_{\ge 0 \text{ по выпуклости}}$$

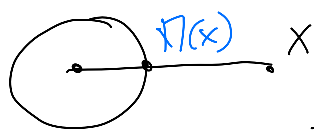
$$\overset{\gamma \le \frac{1}{L}}{\le} (1-\gamma\mu)\|x^k - x^*\|_2^2 \quad \blacksquare$$

Та же сходимость, что и для GD.

- Проекция:

1) L_2-шар

$X = \{x \in \mathbb{R}^d \mid \|x\|_2^2 \le 1\}$ $\quad \Pi(x) = \min\left\{1; \frac{1}{\|x\|_2}\right\} x$



2) $X = \{x \in \mathbb{R}^d \mid a_i \le x_i \le b_i\}$ $\quad [\Pi(x)]_i = \begin{cases} a_i & x_i < a_i \\ x_i & a_i \le x_i \le b_i \\ b_i & b_i < x_i \end{cases}$

3) $X = \{x \in \mathbb{R}^d \mid Ax = b\}$ $\quad \Pi(x) = x - A^T(AA^T)^{-1}(Ax - b)$

4) алгоритмы, находящие точные проекции

- Линейная задача (альтернатива проекции):

$$\min_{s \in \overline{X}} \langle s; g \rangle \qquad g \text{ – некоторый вектор}$$

ограничено

1) L_1 - шар

$$\overline{X} = \{ x \in \mathbb{R}^d \mid \|x\|_1 \leq 1 \}$$

$$i = \arg\max_j |g_j|$$

$$s^* = -\operatorname{sign}(g_j) e_i$$ ← базисный вектор

2) симплекс

$$\overline{X} = \{ x \in \mathbb{R}^d \mid \sum_{i=1}^{d} x_i = 1 \quad x_i \geq 0 \}$$

$$s^* = e_i \qquad i = \arg\min_j g_j$$

3) L_∞ - шар

$$\overline{X} = \{ x \in \mathbb{R}^d \mid \|x\|_\infty \leq 1 \}$$

$$s^* = -\sum_{i=1}^{d} \operatorname{sign}(g_i) e_i$$

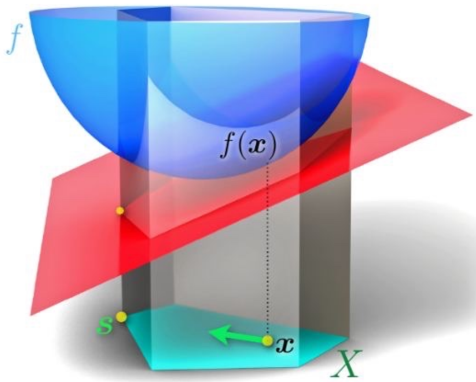
- Метод Франк-Вульфа:

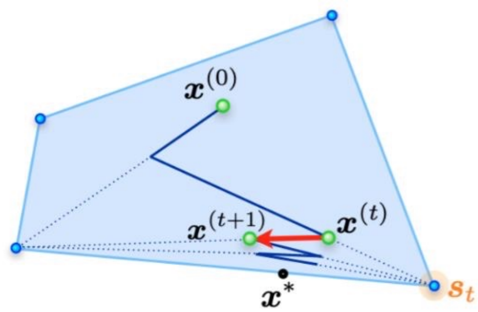
метод условного градиента

$$\boxed{\begin{aligned} s^k &= \arg\min_{s \in \overline{X}} \langle s; \nabla f(x^k) \rangle \\ x^{k+1} &= (1-\gamma_k) x^k + \gamma_k s^k \qquad \gamma_k = \frac{2}{k+2} \end{aligned}}$$

Идея:

$$\underset{s \in \mathbb{X}}{\operatorname{argmin}} \langle s; \nabla f(x^k) \rangle = \underset{s \in \mathbb{X}}{\operatorname{argmin}} \Big[\underbrace{f(x^k) + \langle s - x^k; \nabla f(x^k) \rangle}_{\text{минимизация линейной аппроксимации}} \Big]$$





$$x^{k+1} = \frac{k}{k+1} x^k + \frac{1}{k+1} s^k \longleftarrow \text{вычисление среднего}$$

$$= \left(1 - \frac{1}{k+1}\right) x^k + \frac{1}{k+1} s^k$$

Метод Ф-В = „усреднение угадок линейной аппроксим."

Док-во сходимости: L-гладкость

$$f(x^{k+1}) = f(x^k) + \langle \nabla f(x^k); x^{k+1} - x^k \rangle + \frac{L}{2} \|x^k - x^{k+1}\|_2^2$$

$$x^{k+1} = x^k + \gamma_k (s^k - x^k)$$

$$= f(x^k) + \gamma_k \langle \nabla f(x^k); s^k - x^k \rangle + \frac{L \gamma_k^2}{2} \|s^k - x^k\|_2^2$$

$\mathbb{X}$ - ограничено, $D = \operatorname{diam} \mathbb{X} = \max\limits_{x, y \in \mathbb{X}} \|x - y\|_2$

$$= f(x^k) + \gamma_k \langle \nabla f(x^k); s^k - x^k \rangle + \frac{L D^2}{2} \gamma_k^2$$

$$\langle s^k; \nabla f(x^k) \rangle = \min_{s \in \mathbb{X}} \langle s; \nabla f(x^k) \rangle \le \langle x^*; \nabla f(x^k) \rangle$$

$$\le f(x^k) + \gamma_k \langle \nabla f(x^k); x^* - x^k \rangle + \frac{L D^2}{2} \gamma_k^2$$

выпуклость

$$\le f(x^k) + \gamma_k \left(f(x^*) - f(x^k)\right) + \frac{LD^2}{2}\gamma_k^2$$

Вычтем $f(x^*)$

$$f(x^{k+1}) - f(x^*) \le (1-\gamma_k)\left(f(x^k) - f(x^*)\right) + \frac{LD^2}{2}\gamma_k^2$$

По индукции: если $\gamma_k = \frac{2}{k+2}$, тогда

$$f(x^k) - \underbrace{f(x^*)}_{f^*} \le \frac{2\max\{LD^2;\ f(x^0) - f(x^*)\}}{k+2}$$

- БИ верна
- ПИ

$$f(x^{k+1}) - f(x^*) \le (1-\gamma_k)\left(f(x^k) - f(x^*)\right) + \frac{LD^2}{2}\gamma_k^2$$

$$\gamma_k = \frac{2}{k+2}$$

$$= \left(1 - \frac{2}{k+2}\right)\left(f(x^k) - f(x^*)\right) + \frac{LD^2}{2} \cdot \frac{4}{(k+2)^2}$$

предположение

$$\le \frac{k}{k+2} \cdot \frac{2\max\{LD^2;\ f(x^0) - f(x^*)\}}{k+2} + \frac{LD^2}{2} \cdot \frac{4}{(k+2)^2}$$

$$LD^2 \le \max\{LD^2;\ f(x^0) - f(x^*)\}$$

$$= \frac{1}{(k+2)^2}(k+1) \cdot 2\max\{LD^2;\ f(x^0) - f(x^*)\}$$

$$\frac{k+1}{(k+2)^2} \le \frac{1}{k+3} = \frac{1}{k+1+2}$$

$$\le 2\max\{LD^2;\ f(x^0) - f(x^*)\} \cdot \frac{1}{k+1+2} \qquad \blacksquare$$

- Сходимость метода Франк-Вульфа

f – выпуклая, L-гладкая

X – ограничено диаметром D, выпуклое

$$\boxed{f(x^k) - f(x^*) \le \frac{2\max\{LD^2;\ f(x^0) - f(x^*)\}}{k+2}}$$

Итоги по Ф-В:

⊕ сублинейная сходимость для вып. задач. (как у GD)

⊕ простота реализации

⊕ работа с линейной миним.

⊖ „простые" множества

⊖ для сильно вып. сходимость та же