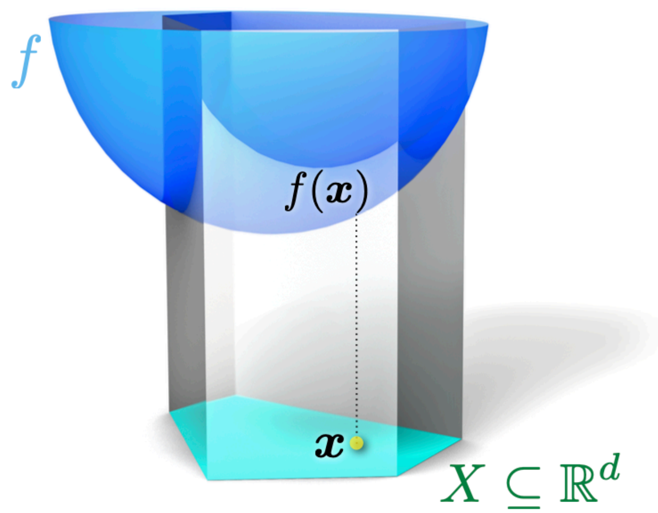
$$\min_{x \in X} f(x)$$



Условие оптимальности

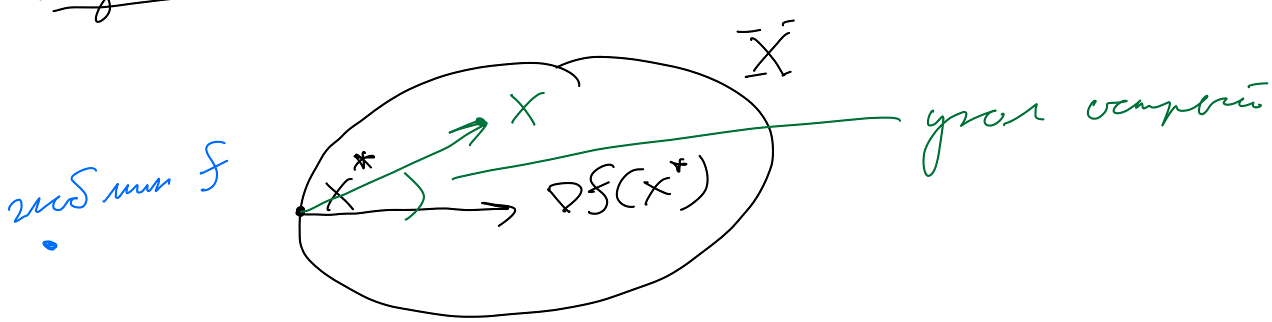
$\cancel{\nabla f(x^*) = 0}$

Условие оптимальности для задачи с ограничениями

Пусть дана выпуклая непрерывно дифференцируемая на $\mathbb{R}^d$ функция $f : \mathbb{R}^d \to \mathbb{R}$ и выпуклое множество $\mathcal{X}$. Тогда $x^* \in \mathcal{X}$ – глобальный минимум f на $\mathcal{X}$ тогда и только тогда, когда для всех $x \in \mathcal{X}$ выполнено

$$\langle \nabla f(x^*), x - x^* \rangle \geq 0.$$

Физический смысл:



Док-во:

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

- достаточность

выпуклость

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

опред. глоб мин. на $\underline{\overline{X}}$

- необходимость

x^* – глоб мин на $\underline{\overline{X}}$

от противного $\exists \hat{x} \in \underline{\overline{X}}$:

$$\langle \nabla f(x^*); \hat{x} - x^* \rangle < 0$$

$$\underbrace{x_\lambda}_{\in \underline{\overline{X}}} = \lambda \hat{x} + (1-\lambda) x^* \qquad \lambda \in [0;1]$$

выпуклость $\underline{\overline{X}}$

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

$$\varphi'(\lambda) = \langle \nabla f(x^* + \lambda(\hat{x} - x^*)); \hat{x} - x^* \rangle$$

$$\varphi'(0) = \langle \nabla f(x^*); \hat{x} - x^* \rangle < 0$$

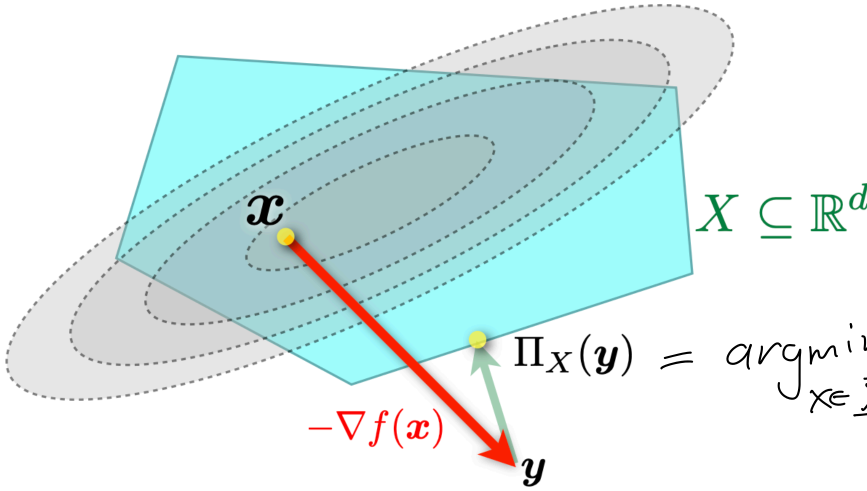
φ в окр нуля убывает

$\exists \hat{\lambda} > 0$:

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

противоречие ■

Проекция



$= \arg\min_{x \in X} \|y - x\|_2^2$

Алгоритм 1 Градиентный спуск с проекцией

Вход: размеры шагов $\{\gamma_k\}_{k=0} > 0$, стартовая точка $x^0 \in \mathbb{R}^d$, количество итераций K

for $k = 0, 1, \ldots, K-1$ **do**
  Вычислить $\nabla f(x^k)$
  $x^{k+1} = \Pi_{\mathcal{X}}[x^k - \gamma_k \nabla f(x^k)]$
end for

Выход: x^K

Свойство оператора проекции

Для выпуклого замкнутого множества $\mathcal{X}$ и любой точки оператор проекции существует и принимает единственно значение.

Док-во: сильно выпуклая ф. ($\| \ \|_2^2$) мин. на вып мн-ве: решение сущ. и ед.

Свойство оператора проекции

Пусть $\mathcal{X} \subseteq \mathbb{R}^d$ выпуклое замкнутое множество, $x \in \mathcal{X}, y \in \mathbb{R}^d$. Тогда

$$\langle x - \Pi_{\mathcal{X}}(y), y - \Pi_{\mathcal{X}}(y) \rangle \leq 0.$$

Док-во:

$$\Pi_X(y) = \arg\min_{x \in X} \|x - y\|_2^2$$

x^* для $f(x) = \|x - y\|_2^2$ на X

$\nabla f(x) = 2(x - y)$

Условие оптим.:

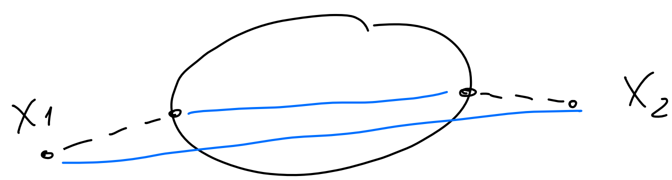
$$\langle \cancel{2}(\underbrace{x^*}_{\Pi_X(y)} - y);\ x - x^* \rangle \geq 0 \qquad x \in X$$

$$\langle \Pi_X(y) - y;\ x - \Pi_X(y) \rangle \geq 0 \qquad \blacksquare$$

Свойство нерасширяемости оператора проекции

Пусть $\mathcal{X} \subseteq \mathbb{R}^d$ выпуклое замкнутое множество, $x_1, x_2 \in \mathbb{R}^d$. Тогда

$$\|\Pi_{\mathcal{X}}(x_1) - \Pi_{\mathcal{X}}(x_2)\|_2 \leq \|x_1 - x_2\|_2.$$



Док-во: $\quad x \in X \quad y \in \mathbb{R}^d$

$$\langle \Pi_X(y) - y;\ x - \Pi_X(y) \rangle \geq 0$$

$y = x_1$

$$\langle \Pi_X(x_1) - x_1;\ \underbrace{x}_{\Pi_X(x_2)} - \Pi_X(x_1) \rangle \geq 0$$

$$\langle \Pi_X(x_1) - x_1;\ \Pi_X(x_2) - \Pi_X(x_1) \rangle \geq 0$$

аналогично

$$+ \quad \langle \Pi_X(x_2) - x_2;\ \Pi_X(x_1) - \Pi_X(x_2) \rangle \geq 0$$

$$\langle \underline{\Pi(x_1)} - x_1 - \underline{\Pi(x_2)} + x_2; \underline{\Pi(x_2) - \Pi(x_1)} \rangle \geq 0$$

$$-\|\Pi(x_1) - \Pi(x_2)\|_2^2 + \langle x_2 - x_1; \Pi(x_2) - \Pi(x_1) \rangle \geq 0$$

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

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

Свойство оператора проекции

Для x^* – решения условной задачи минимизации выпуклой непрерывно дифференцируемой функции f на выпуклом замкнутом множестве $\mathcal{X}$ справедливо

$$x^* = \Pi_{\mathcal{X}}(x^* - \gamma \nabla f(x^*))$$

Док-во:

$$\Pi_{\mathcal{X}}(x^* - \gamma \nabla f(x^*)) = \underset{x \in \mathcal{X}}{\arg\min} \underline{\|x - x^* + \gamma \nabla f(x^*)\|_2^2}$$

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

$$+ 2\gamma \langle \nabla f(x^*); x - x^* \rangle$$

$$\underbrace{\|x^* - x\|_2^2}_{\geq 0} + 2\gamma \underbrace{\langle \nabla f(x^*); x - x^* \rangle}_{\geq 0 \text{ усл. опт.}}$$

$x = x^*$ даёт 0

решение

Док-во сходимости град. спуска с проекцией

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

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

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

нерастяг. проекц

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

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

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

просто так не добавить

L - гладкость

$$\|\nabla f(x^k) - \nabla f(x^*)\|_2^2 \le 2L\left(f(x^k) - f(x^*) - \underbrace{\langle \nabla f(x^*); x^k - x^* \rangle}_{\neq 0}\right)$$

μ - сильная выпуклость

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

$$\leq \|x^k - x^*\|_2^2 - 2\gamma\langle -\nabla f(x^*); x^k - x^*\rangle$$

$$- 2\gamma\left(\frac{\mu}{2}\|x^k - x^*\|_2^2 + f(x^k) - f(x^*)\right)$$

$$+ \gamma^2 \cdot 2L\left(f(x^k) - f(x^*) - \langle\nabla f(x^*); x^k - x^*\rangle\right)$$

$$= (1 - \gamma\mu)\|x^k - x^*\|_2^2$$

$$+ (\gamma^2 \cdot 2L - 2\gamma)\cdot$$

$$\underbrace{\left(f(x^k) - f(x^*) - \langle\nabla f(x^*); x^k - x^*\rangle\right)}_{\text{по выпуклости } \geq 0}$$

$$2\gamma(\gamma L - 1) \leq 0 \Rightarrow \gamma \leq \frac{1}{L}$$

То же самое, что и для град. спуска

Важна „простота" множества X

Аналитические решения:

- ℓ_2-шар радиуса 1 с центром в 0:

$$\mathcal{X} = \left\{x \in \mathbb{R}^d \mid \|x\|_2^2 = \sum_{i=1}^{d} x_i^2 \leq 1\right\} \quad \Pi_{\mathcal{X}}(x) = \min\left\{1, \frac{1}{\|x\|_2}\right\} x$$

- Куб

$$\mathcal{X} = \left\{x \in \mathbb{R}^d \mid a_i \leq x_i \leq b_i\right\} \quad [\Pi_{\mathcal{X}}(x)]_i = \begin{cases} a_i, & \text{если } x_i \leq a_i \\ x_i, & \text{если } a_i < x_i \leq b_i \\ b_i, & \text{иначе} \end{cases}$$

- Линейные ограничения типа равенств

$$\mathcal{X} = \left\{x \in \mathbb{R}^d \mid Ax = b\right\} \quad \Pi_{\mathcal{X}}(x) = x - A^T(AA^T)^{-1}(Ax - b)$$

- Для некоторых множеств существуют эффективные алгоритмы проекций.

Квадратичная функция – сложно

Проще квадратичной – линейную

$$\min_{s \in \mathcal{X}} \langle s; g \rangle$$

ограничено

- ℓ_1-шар с радиусом 1 и центром в 0:

$$\mathcal{X} = \left\{ x \in \mathbb{R}^d \mid \|x\|_1 = \sum_{i=1}^{d} |x_i| \le 1 \right\} \quad s^* = -\mathrm{sign}(g_i) \mathrm{e}_i, \; i = \operatorname*{argmax}_j |g_j|$$

- Вероятностный симплекс:

$$\mathcal{X} = \left\{ x \in \mathbb{R}^d \mid x_i \ge 0, \; \sum_{i=1}^{d} x_i = 1 \right\} \quad s^* = \mathrm{e}_i, \text{ где } i = \operatorname*{argmin}_j g_j$$

- ℓ_∞-шар радиуса 1 с центром в 0:

$$\mathcal{X} = \left\{ x \in \mathbb{R}^d \mid \|x\|_\infty = \max_{i=1,\dots,d} |x_i| \le 1 \right\} \quad s^* = -\sum_{i=1}^{d} \mathrm{sign}(g_i) \mathrm{e}_i$$

- линейная не накладна, альтернатива проекции

Алгоритм 2 Метод Франк-Вульфа

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

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

2: Вычислить $\nabla f(x^k)$

3: Найти $s^k = \operatorname*{argmin}_{s \in \mathcal{X}} \langle s, \nabla f(x^k) \rangle$

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

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

6: **end for**

Выход: x^K

Физика:

- $s^k = \operatorname*{argmin}_{s \in \mathcal{X}} \langle s; \nabla f(x^k) \rangle$

$$= \operatorname*{argmin}_{s \in \mathcal{X}} \underbrace{f(x^k) + \langle \nabla f(x^k); s - x^k \rangle}_{\text{линейная аппрокс. ф.}}$$

s^k – на границе мн-ва

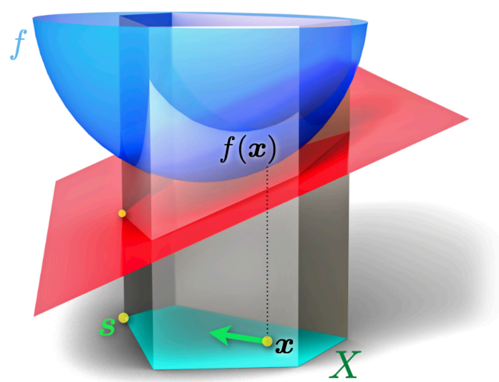
s^k – крайняя точка

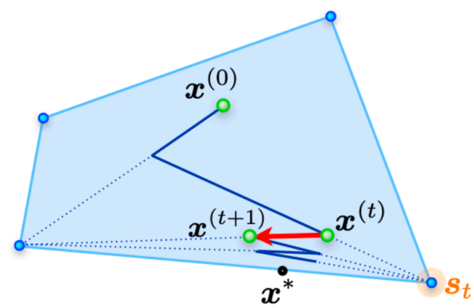
- $(1-\gamma_k)x_k + \gamma_k s_k \qquad \gamma_k = \frac{2}{k+2}$

динамическое среднее:

$$x^{k+1} = \frac{k}{k+1}x^k + \frac{1}{k+1}s^k$$

среднее точек с границы





Док-во сходимости:

- f – L-гладкая, выпуклая

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

L-гладкость (физ. смысл)

$$\le f(x^k) + \gamma_k\langle \nabla f(x^k); s^k - x^k\rangle + \frac{\gamma_k^2 L}{2}\underbrace{\|s^k - x^k\|_2^2}_{\mathrm{diam}^2(X) = D^2}$$

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

$$\langle \nabla f(x^k); s^k \rangle = \min \langle \nabla f(x^k); s \rangle \leq \langle \nabla f(x^k); x^* \rangle$$

$$\leq f(x^k) + \gamma_k \langle \nabla f(x^k); x^* - x^k \rangle + \gamma_k^2 \cdot \frac{LD^2}{2}$$

выпуклость

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

$$- f(x^*)$$

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

По индукции докажем

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

- База: $f(x^0) - f(x^*) \leq \dfrac{2 \cdot \left(f(x^0) - f(x^*)\right)}{0+2}$

• Предпол.: $f(x^{(k-1)}) - f(x^*) \le \dfrac{2\max\{\dots\}}{k-1+2}$

• Переход:

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

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

предполож.

$$\le \left(\frac{k-1}{k+1}\right)\cdot \frac{2\max\{\dots\}}{k+1} + \frac{2LD^2}{(k+1)^2} \qquad 2LD^2 \le \max\{\ \}$$

$$= 2\max\{\ \}\cdot\left(\frac{(k-1)}{(k+1)^2} + \frac{1}{(k+1)^2}\right)$$

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

$k^2+2k \qquad k^2+2k+1$

$$\le \frac{2\max\{\ \}}{k+2}$$

Теорема о сходимости метода Франк-Вульфа

Пусть дана непрерывно дифференцируемая выпуклая L-гладкая функция $f : \mathbb{R}^d \to \mathbb{R}$, тогда для метода Франк-Вульфа справедлива следующая оценка сходимости

$$f(x^K) - f(x^*) \leq \frac{\max\{2L\,\mathrm{diam}(\mathcal{X})^2, f(x^0) - f(x^*)\}}{K+2}$$

где $\mathrm{diam}(\mathcal{X}) := \max_{x,y\in\mathcal{X}} \|x-y\|_2$ – диаметр множества $\mathcal{X}$.

сублинейно! $\frac{1}{K}$ как град. спуск для гладких вып. задач.