Градиентный спуск

$$\min_{x\in\mathbb{R}^n} f(x)$$

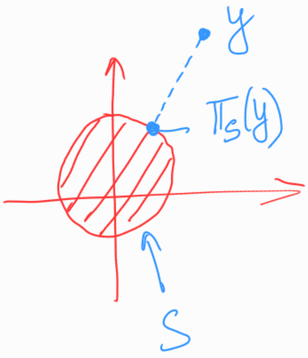
$$\boxed{x_{k+1} = x_k - \alpha_k \cdot \nabla f(x_k)}$$

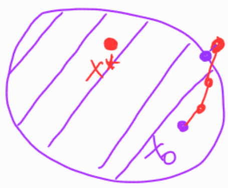
1) Как выбирать α_k?

2) Как выбирать x_0?

А что если

$$\min_{x\in S} f(x) \quad ?$$





Метод проекции градиента

$$x_{k+1} = \underset{x\in S}{\text{PROJ}}\left(x_k - \alpha_k \nabla f(x_k)\right)$$

$$\boxed{x_{k+1} = \Pi_S\left(x_k - \alpha_k \nabla f(x_k)\right)}$$

$$\|\Pi_S(x) - \Pi_S(y)\| \le \|x-y\|$$

$$\|\Pi_S(x_k) - \Pi_S(x^*)\| \le \|x_k - x^*\|$$

Условия оптимальности

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

1) Безусловная оптимизация $S = \mathbb{R}^n$

необх. усл-ие локального экстремума: $\boxed{\nabla f(\tilde{x}) = 0}$

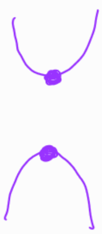
достаточное условие локального экстремума

1) $\nabla f(\tilde{x}) = 0$

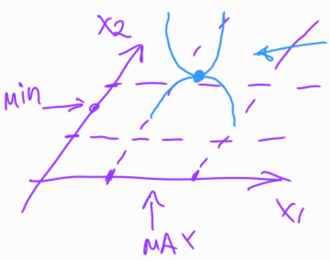
2) $\nabla^2 f(\tilde{x}) \succeq 0$ (min)

$\nabla^2 f(\tilde{x}) \preceq 0$ (max)

(по критерию Сильвестра)



$x \ge 0$



седловая точка

2) Условная оптимизация

а)

$$f(x) \to \min_{x\in\mathbb{R}^n}$$

$$\boxed{g(x)=0} \quad S$$

$\forall x\in S$: $\boxed{f(x) = L(x,\lambda)}$

$\forall \lambda$:

функция Лагранжа

$$\boxed{L(x,\lambda) = f(x) + \lambda\cdot g(x)}$$

$g(x) \to 0$

заменяем задачу усл. опт. на задачу безусл. опт.

„штраф за нарушение условия"

если $x\notin S \Rightarrow g(x)\neq 0$

$\lambda g(x)$ - штраф

Пример:

$$\underbrace{c^Tx}_{f(x)} \to \min_{x\in\mathbb{R}^n}$$

$\boxed{Ax=b}$ | m штук $A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$

$$L(x,\lambda) = \underbrace{c^Tx}_{f(x)} + \lambda_1\cdot(Ax-b)_1 + \lambda_2\cdot(Ax-b)_2 + \ldots + \lambda_m(Ax-b)_m =$$

$$= \underset{1\times n\ n\times 1}{c^Tx} + \underset{1\times m}{\lambda^T}\underset{m\times 1}{(Ax-b)}, \quad \lambda\in\mathbb{R}^m$$

$x, \lambda \to \boxed{L} \to$ выход $\min_{x,\lambda}$

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases} \qquad \begin{cases} \dfrac{\partial(f(x)+\lambda\cdot g(x))}{\partial x} = 0 \\ \dfrac{\partial(f(x)+\lambda\cdot g(x))}{\partial \lambda} = 0 \end{cases}$$

$$\begin{cases} \dfrac{\partial f}{\partial x} + \lambda\cdot\dfrac{\partial g}{\partial x} = 0 \\ g(x) = 0 \end{cases} \longrightarrow \tilde{x}, \tilde{\lambda} :$$

$\tilde{x}$ — экстремум исходной функции $f(x)$

$$L = c^Tx + \lambda^T(Ax-b) = c^Tx + \lambda^TAx - \lambda^Tb = c^Tx + \langle A^T\lambda, x\rangle - \lambda^Tb$$

$A = A^T$, $A\in S^n_{++}$

$$\begin{cases} \dfrac{\partial L}{\partial x} = 0 \\ \dfrac{\partial L}{\partial \lambda} = 0 \end{cases} \quad \begin{cases} c + \lambda^T\cdot\nabla(Ax-b) = 0 \\ Ax = b \end{cases} \quad \begin{cases} c + A^T\lambda = 0 \\ Ax = b \end{cases} \longrightarrow \begin{cases} A^T\cdot\lambda = -c \\ x^* = A^{-1}b \end{cases}$$

$$\begin{cases} \lambda^* = -A^{-1}\cdot c \\ x^* = A^{-1}\cdot b \end{cases}$$

$f(x)$ - издержки предприятия

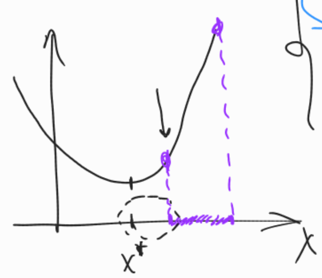
$g(x)$ — командировочные

$g(x) = 100$к руб

$g(x) - 100$кр $= 0$

x - как функция предпр.

λ - стоимость выход за бюджет на командировки



Least Squares:

$$\underset{m\times n}{A}\ \underset{n\times 1}{x} = \underset{m\times 1}{b}$$

$m<n$ — недоопр. / underdeter. — Решений много

$$\|x\|^2 \to \min$$
$$Ax=b$$

$$L(x,\lambda) = \underbrace{x^Tx}_{f(x)} + \lambda^T(Ax-b)$$

$$\begin{cases} \frac{\partial L}{\partial x} = 2x + A^T\lambda = 0 \\ \frac{\partial L}{\partial \lambda} = Ax - b = 0 \end{cases}$$

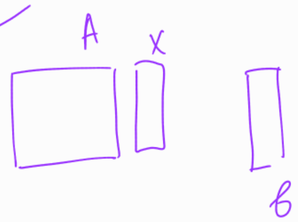
$$\begin{cases} x = -\frac{1}{2}A^T\lambda \\ A\cdot\left(-\frac{1}{2}A^T\lambda\right) - b = 0 \end{cases}$$

$$\begin{cases} x = -\frac{1}{2}A^T\lambda \\ -\frac{1}{2}\underbrace{A\cdot A^T}_{\substack{m\times n\, n\times m \\ m\times m}}\lambda - b = 0 \end{cases} \qquad \begin{cases} x = -\frac{1}{2}A^T\lambda \\ \tilde\lambda = -2\cdot(AA^T)^{-1}\cdot b \end{cases}$$

$$\boxed{\tilde x = A^T\cdot(AA^T)^{-1}\cdot b}$$
$$\boxed{\tilde\lambda = -2(AA^T)^{-1}b}$$

$m=n$ — Решение одно

$$\det A \neq 0$$
$$\boxed{x^* = A^{-1}b}$$



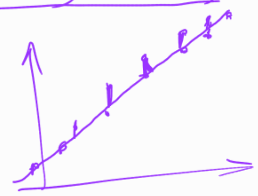
$m>n$ — Overdeterm. — Решений нет

$$\|Ax-b\|_2^2 \to \min_{x\in\mathbb{R}}$$

$$A^T(Ax-b) = 0 \quad (\nabla f = 0)$$

$$\underset{n\times m}{A^T}\ \underset{m\times n}{A}x = A^Tb$$

$$\boxed{\tilde x = (\underset{n\times n}{A^TA})^{-1}\cdot A^Tb}$$

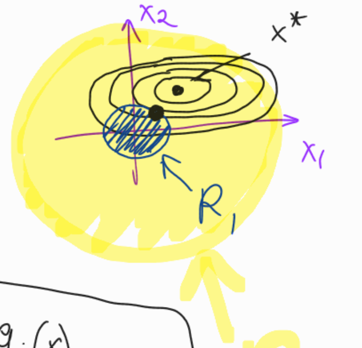


граница

$$h(x)=0$$

$$\min f(x)$$
$$h(x)\le 0$$

ограничения-неравенства чуть усложняют задачу



$$x_1^2 + x_2^2 - R^2 \le 0$$

$$R_1 < R_2$$

KKT

$$\min f(x)$$
$$h(x)=0$$
$$g_i(x)\le 0$$

$$\Rightarrow \quad L(x,\lambda,\mu) = f(x) + \sum_j \lambda_i h_i(x) + \sum_i \mu_i g_i(x)$$

1) $\nabla_x L(\tilde x,\tilde\lambda,\tilde\mu) = 0$
2) $\nabla_\lambda L(\tilde x,\tilde\lambda,\tilde\mu) = 0$
3) $\mu_i \geq 0$
4) $\mu_i\cdot g_i(x) = 0$
5) $g_i(x)\le 0$

$$\longrightarrow \tilde x, \tilde\lambda, \tilde\mu$$

- Функция Лагранжа
- Если БУ → градиент

  Усл. опт. с рав. → $L(x,\lambda)$

  Усл. опт. с нерав. → $L(x,\lambda,\mu)$ + ККТ

fmin.xyz
Theory / Optimality conditions

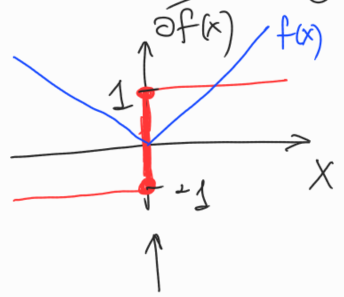
- Если $f(x)$ – выпуклая и $x \in S$ – выпуклое мн-во – ВЫПУКЛАЯ ЗАДАЧА

  бюджетное

  Если задача выпуклая → то любой ЛОКАЛЬНЫЙ экстремум – ГЛОБАЛЬНЫЙ

- Если $f(x)$ – не гладкая, то $\nabla \to \partial$

$$\cancel{\nabla f(\tilde{x}) = 0} \implies 0 \in \partial f(\tilde{x})$$



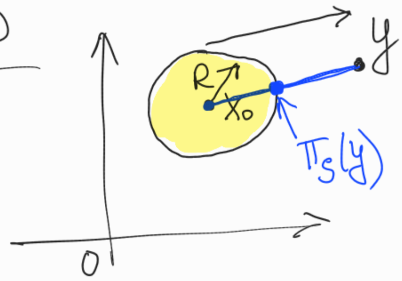
$f(x) = |x|$

$\partial f = [-1, 1]$

$g = -\frac{2}{3} \in \partial f$

Проекции для метода проекции (суб)градиента

① Шар



$$S = \{x : \|x - x_0\| \leq R\}$$

$$\Pi_S(y) = \, ? = \pi$$

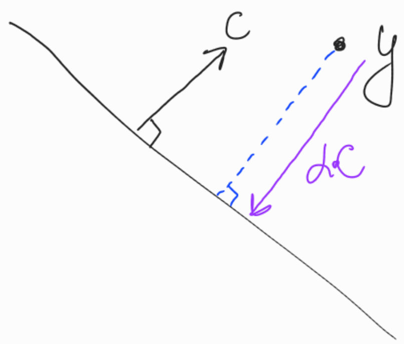
Точка π лежит на отрезке от x_0 до y: $\pi = x_0 + \lambda(y - x_0)$

π лежит на границе: $\|\pi - x_0\| = R$

$$\Pi_S(y) \to \boxed{\pi = x_0 + \frac{y - x_0}{\|y - x_0\|} \cdot R}, \quad y \notin S$$

$$\to \pi = y, \quad y \in S$$

② Гиперплоскость



$$S = \{x : c^T x = b\}$$

$$\Pi_S(y) = ? = \Pi$$

$$\Pi = y + \alpha \cdot c$$

$$\hat{\ } \; 0$$

$\Pi \in S$: $\quad c^T \Pi = b$

$$c^T(y + \alpha c) = b \;=\; c^T y + \alpha \cdot c^T c = b$$

$$\alpha c^T c = b - c^T y$$

$$\Rightarrow \boxed{\alpha = \frac{b - c^T y}{c^T c}}$$

$$\boxed{\Pi_S(y) = \begin{cases} y + \dfrac{b - c^T y}{c^T c} \cdot c, & y \notin S \\ y, & y \in S \end{cases}}$$

$$\Pi = y + \frac{b - c^T y}{c^T c} \cdot c = \frac{c^T c \cdot y + b \cdot c - y^T c \cdot c}{c^T c} =$$

③ Линейное подпространство

$$S = \{x \in \mathbb{R}^n : Ax = b, \; A \in \mathbb{R}^{m \times n}, \; b \in \mathbb{R}^m\}$$

мно-во решений системы лин. ур-ий

$$\Pi = b + AA^T(x - b)$$

Почему l_1 – спарсность

$$\|x\|_1 = \sum_i |x_i|$$