1. Условия оптимальности.
2. Оптимизация с ограничениями - равенствами (неравенствами)
3. Решение системы ЛУ.

Optimality conditions. KKT

Background

Extreme value (Weierstrass) theorem

замкнутое, ограниченное мн-во

Let $S \subset \mathbb{R}^n$ be compact set and $f(x)$ continuous function on S. So that, the point of the global minimum of the function $f(x)$ on S exists.

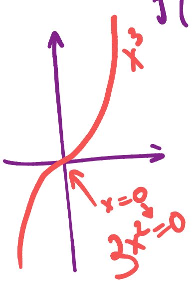
Есть $f(x)$.

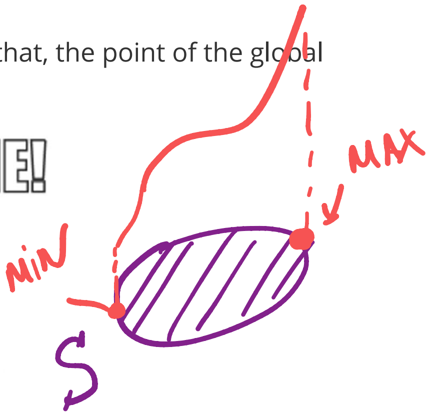
Экстремумы?

$f'(x) = 0$ → $x^* = \dots$

$f''(x) > 0$

$x \geq 0$





Lagrange multipliers

Consider simple yet practical case of equality constraints:

$$
\begin{split}
& f(x) \to \min\limits_{x \in \mathbb{R}^n} \\
\text{s.t. } & h_i(x) = 0, i = 1, \ldots, m
\end{split}
$$

The basic idea of Lagrange method implies switch from conditional to unconditional optimization through increasing the dimensionality of the problem:

$$
L(x, \lambda) = f(x) + \sum\limits_{i=1}^m \lambda_i h_i(x) \to \min\limits_{x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m}
$$

General formulations and conditions

(критерий) целевая функция → $\boxed{f(x) \to \min\limits_{x \in S}}$ ← S - бюджетное мн-во

We say that the problem has a solution if the following set **is not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

Unconstrained optimization

Безусловная оптимизация

General case

Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ be a twice differentiable function.

$$
\boxed{f(x) \to \min\limits_{x \in \mathbb{R}^n}} \tag{UP}
$$

Необходимые усл.

If x^* - is a local minimum of $f(x)$, then:

$$
\boxed{\nabla f(x^*) = 0} \tag{UP:Necessary}
$$

$$\nabla f(x^*) = 0 \qquad \text{(UP:Necessary)}$$

$\forall y \in \mathbb{R}^n, y \neq 0: \; y^T X \cdot y > 0$

$\lambda_1, \dots, \lambda_n > 0 \iff \det > 0$

положит. опред. матрицы

достаточное условие

If $f(x)$ at some point x^* satisfies the following conditions:

$$H_f(x^*) = \nabla^2 f(x^*) \succeq (\preceq) 0, \quad \boxed{\nabla f(x^*) = 0} \qquad \text{(UP:Sufficient)}$$

then (if necessary condition is also satisfied) x^* is a local minimum(maximum) of $f(x)$.

Peano surface

контр пример для $\nabla^2 f \succeq 0$

Convex case

It should be mentioned, that in **convex** case (i.e., $f(x)$ is convex) necessary condition becomes sufficient. Moreover, we can generalize this result on the class of non-differentiable convex functions.

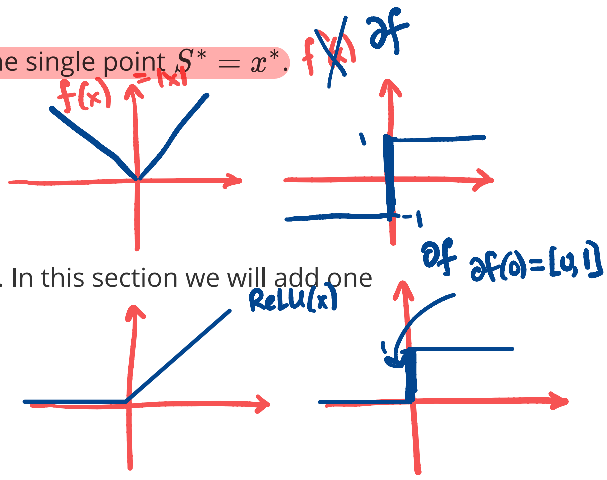
Let $f(x) : \mathbb{R}^n \to \mathbb{R}$ - convex function, then the point x^* is the solution of (UP) if and only if:

$$\boxed{0_n \in \partial f(x^*)}$$

One more important result for convex constrained case sounds as follows. If $f(x) : S \to \mathbb{R}$ - convex function defined on the convex set S, then:

- Any local minima is the global one.
- The set of the local minimizers S^* is convex.
- If $f(x)$ - strongly convex function, then S^* contains only one single point $S^* = x^*$.

субградиент



Optimization with equality conditions

Intuition

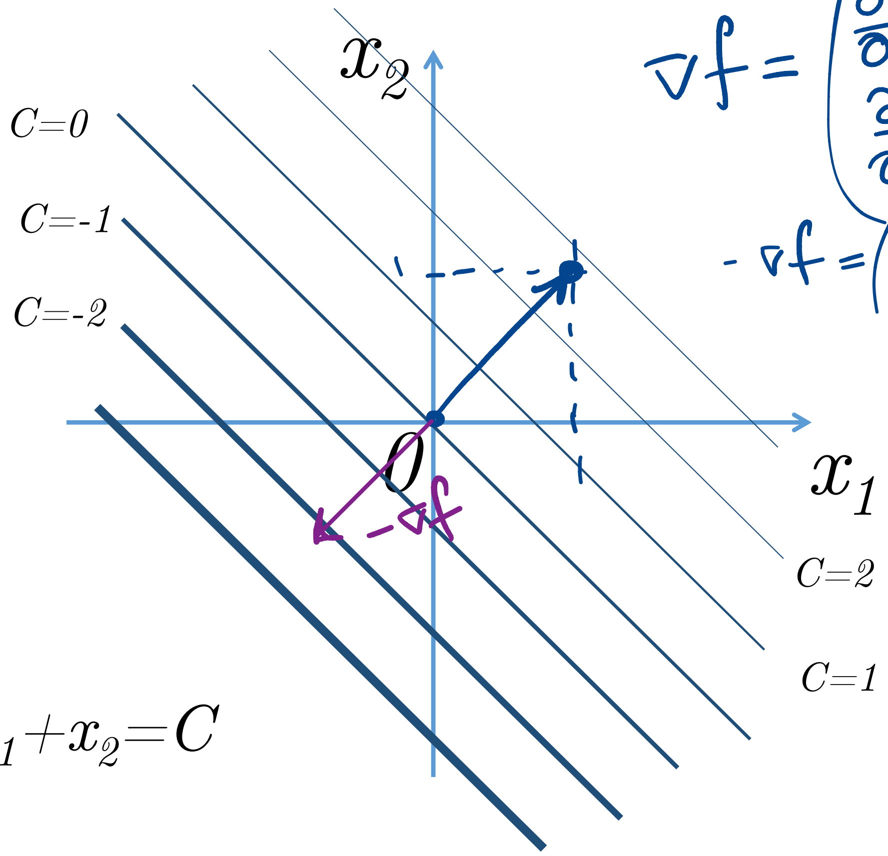
Things are pretty simple and intuitive in unconstrained problem. In this section we will add one equality constraint, i.e.

$$\boxed{\begin{aligned} &f(x) \to \min_{x \in \mathbb{R}^n} \\ &\text{s.t. } h(x) = 0 \end{aligned}}$$

We will try to illustrate approach to solve this problem through the simple example with $f(x) = x_1 + x_2$ and $h(x) = x_1^2 + x_2^2 - 2$

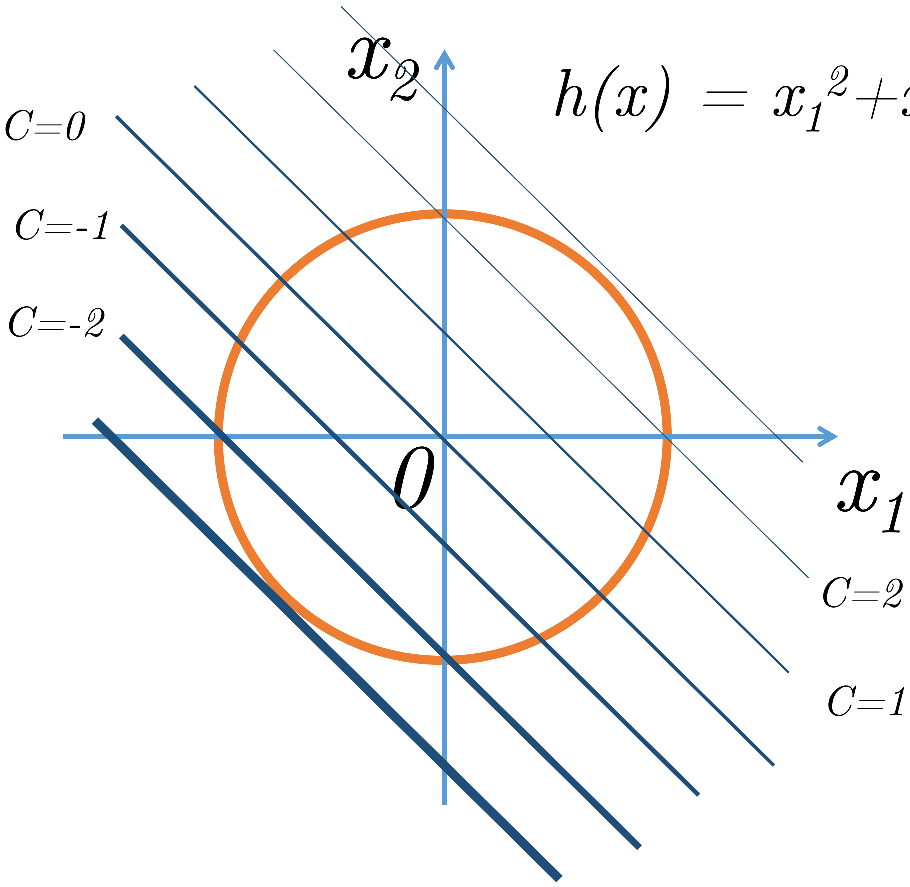
$$\nabla f = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

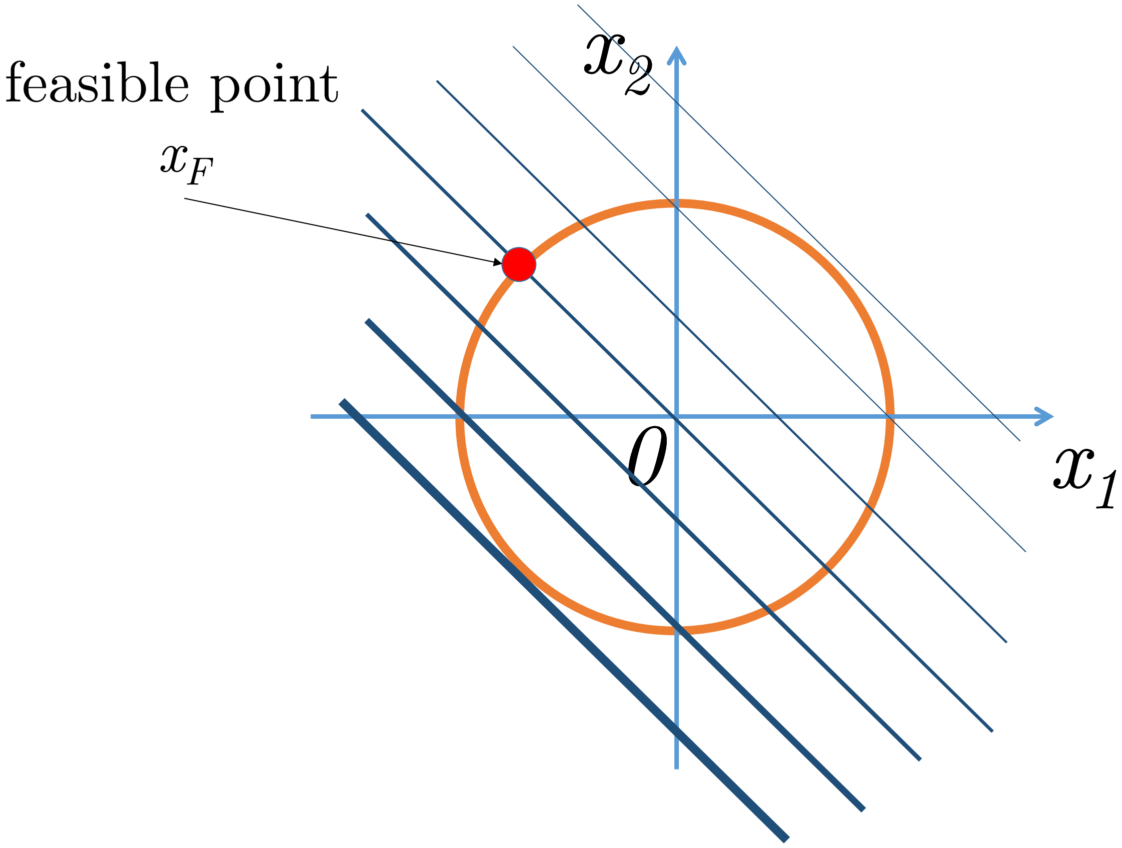
$$-\nabla f = \begin{pmatrix} -1 \\ -1 \end{pmatrix}$$

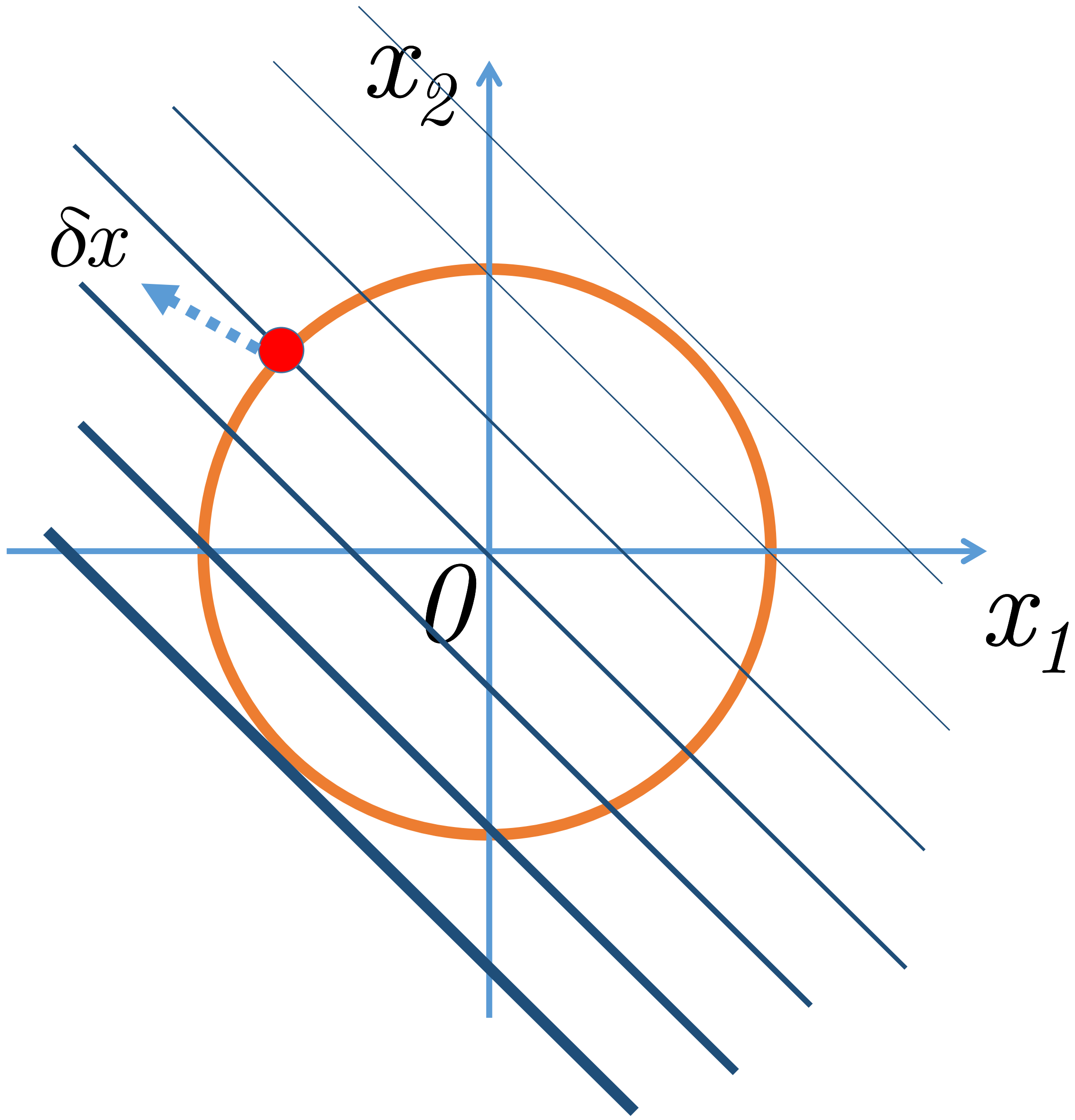


$$f(x) = x_1 + x_2 = C$$

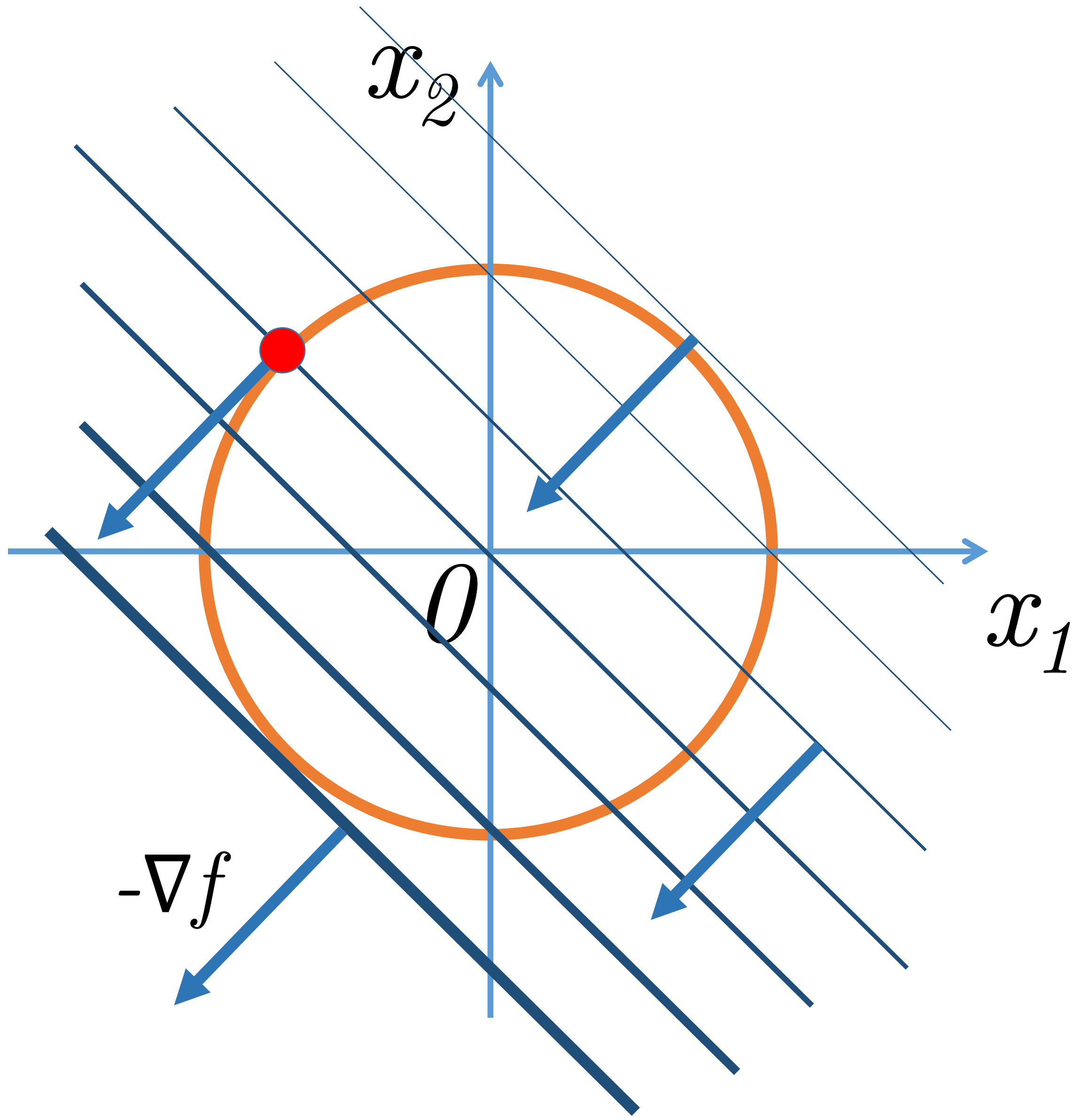
$$h(x) = x_1^2 + x_2^2 - 2 = 0$$




feasible point
x_F
x_2
x_1
0



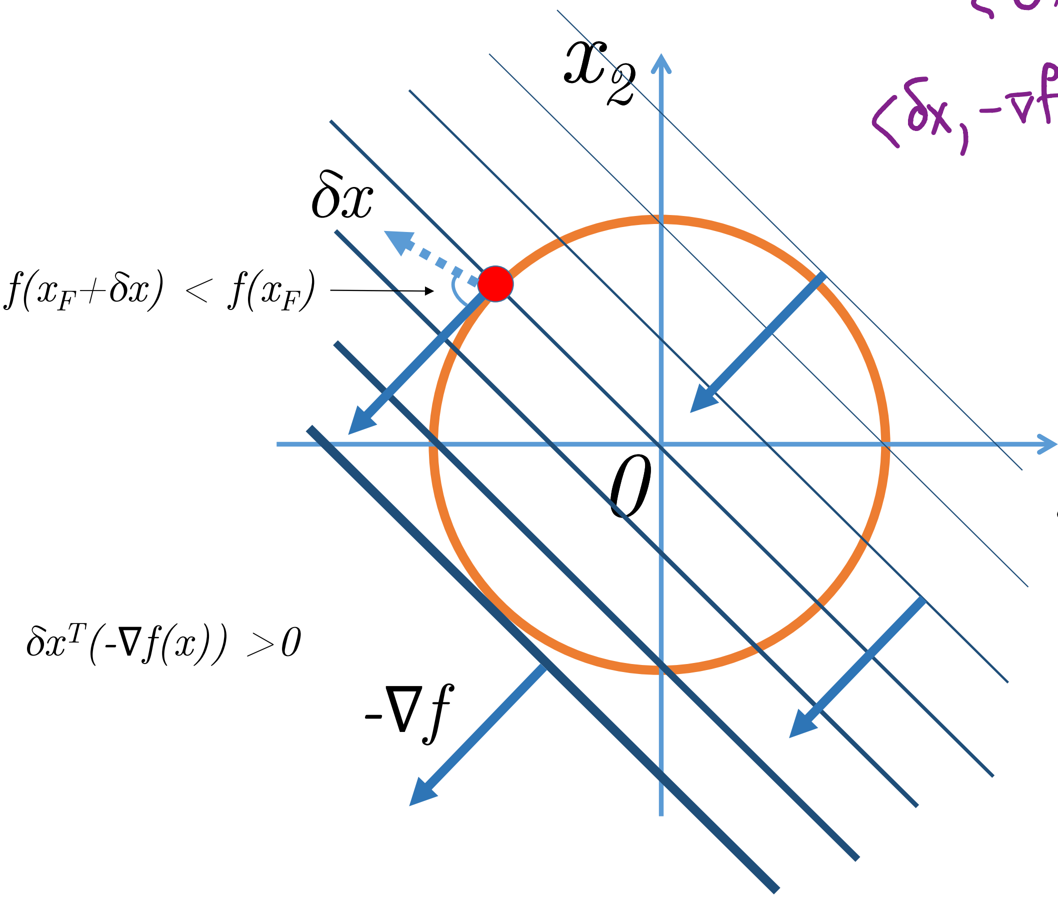
x_2
δx
0
x_1



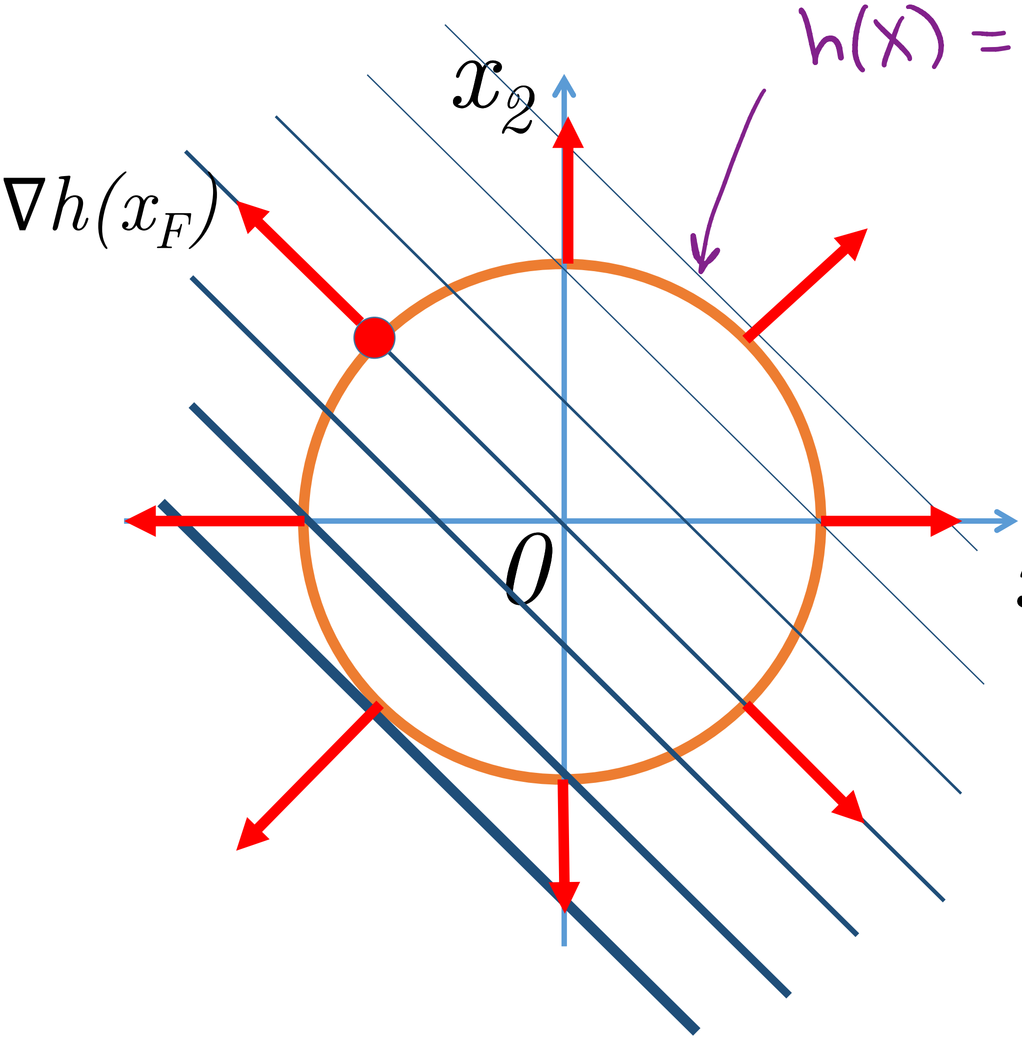
x_2
x_1
0
$-\nabla f$

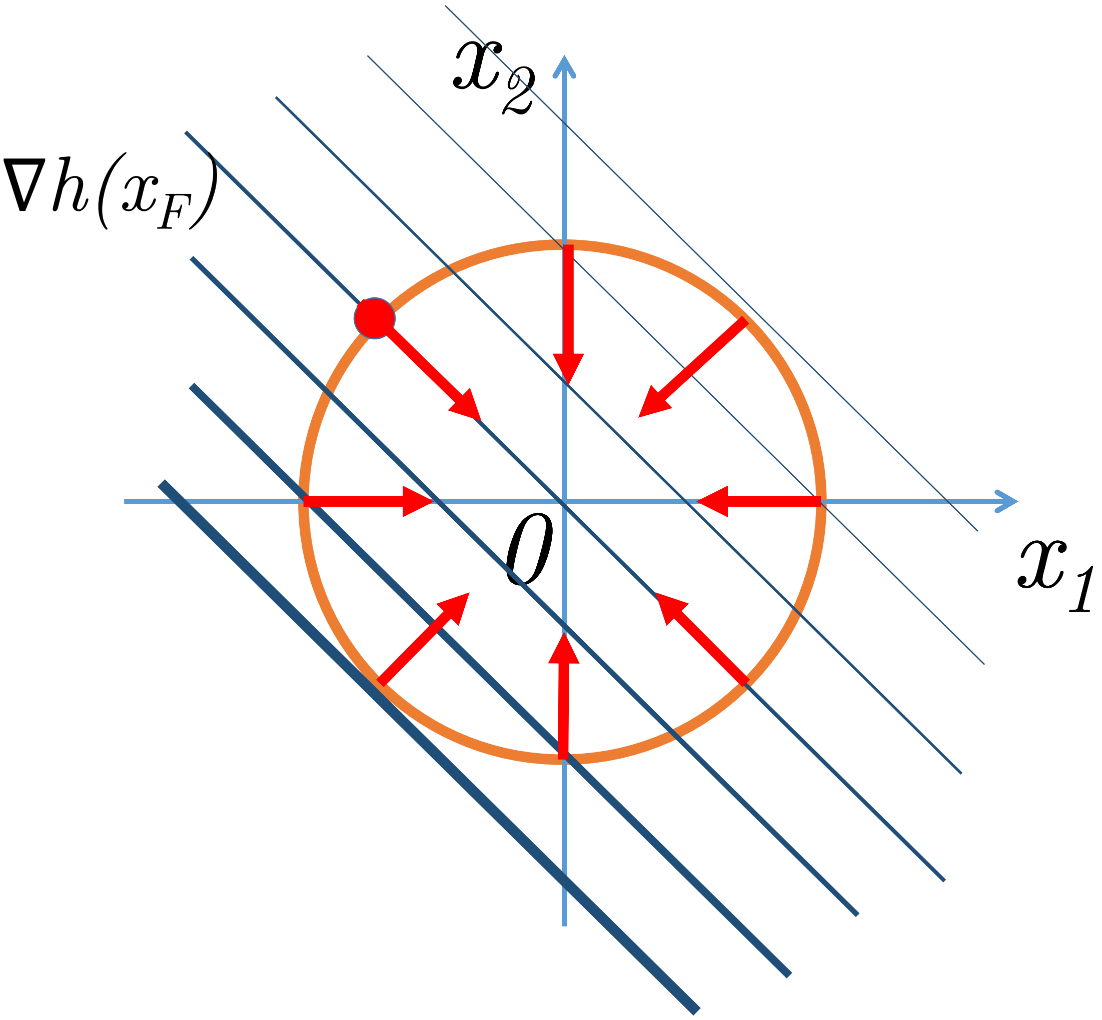
Если $\langle \delta x, -\nabla f \rangle \geq 0$ $f\downarrow\downarrow$

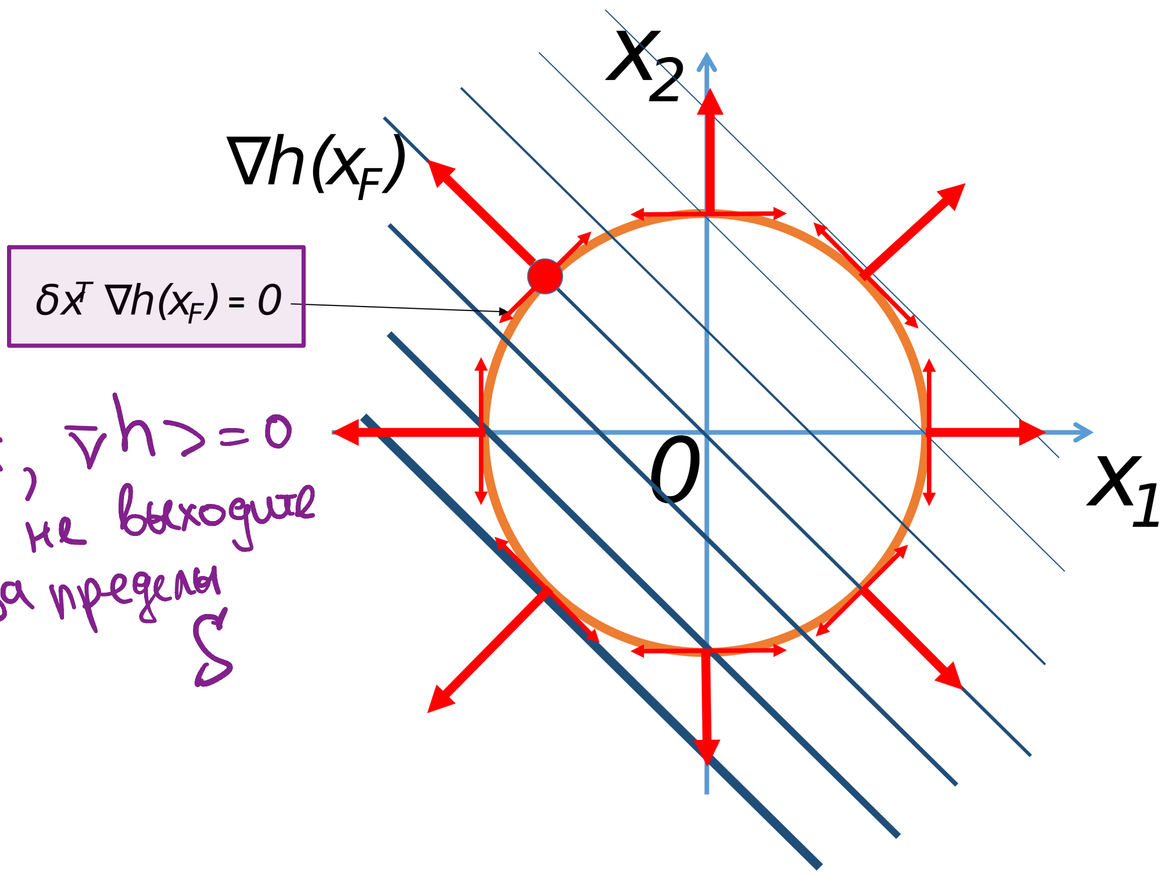
$\langle \delta x, -\nabla f \rangle \leq 0$, $f \uparrow\uparrow$



$h(x) = 0$ (бюджетное мн-во)







$\langle \delta x, \nabla h \rangle = 0$
вы не выходите за пределы S

Generally: in order to move from x_F along the budget set towards decreasing the function, we need to guarantee two conditions:

$$\boxed{\langle \delta x, \nabla h(x_F) \rangle = 0}$$ НЕ выходим из бюджета

$$\boxed{\langle \delta x, -\nabla f(x_F) \rangle \geqslant 0}$$ не делаем хуже

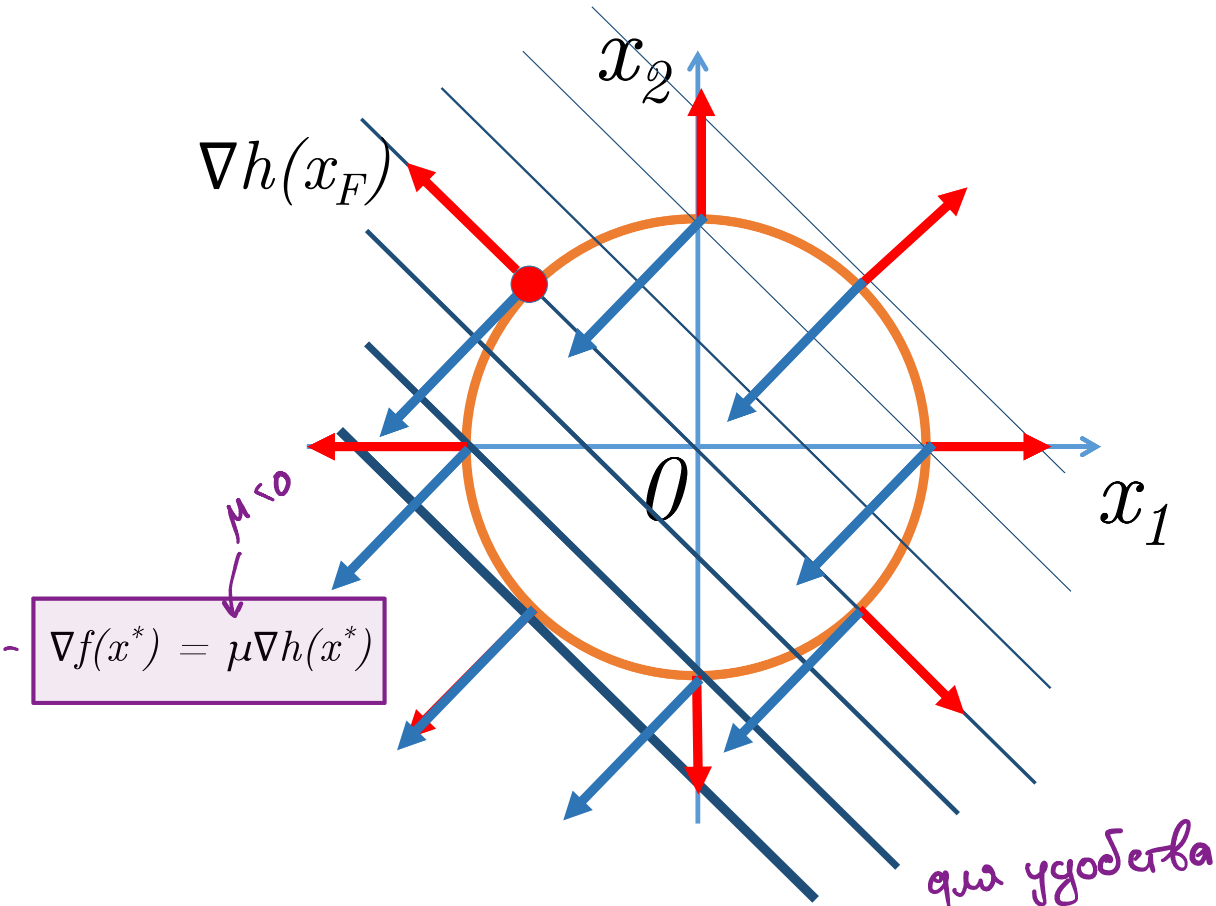
Let's assume, that in the process of such a movement we have come to the point where

условие локального оптимума

$$\delta x \cdot \boxed{\nabla f(x) = \lambda \nabla h(x)} \quad \delta x \cdot$$

$$\langle \delta x, -\nabla f(x) \rangle = -\langle \delta x, \lambda \nabla h(x) \rangle = 0$$

Then we came to the point of the budget set, moving from which it will not be possible to reduce our function. This is the local minimum in the limited problem :)



для удобства

So let's define a Lagrange function (just for our convenience):

функция Лагранжа

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

$L(x, \lambda) : \mathbb{R}^{n+1} \to \mathbb{R}$

$\mathbb{R}^{n+m} \to \mathbb{R}$

Then the point x^* be the local minimum of the problem described above, if and only if:

n - кол-во переменных

m - кол-во огранич.

необх.

$$\boxed{\begin{split} & \nabla_x L(x^*, \lambda^*) = 0 \text{ that's written above} \\ & \nabla_\lambda L(x^*, \lambda^*) = 0 \text{ condition of being in budget set} \end{split}}$$

достаточное $\rightarrow$ $$\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \quad \forall y \underset{\neq 0}{\in} \mathbb{R}^n : \nabla h(x^*)^\top y = 0$$

$\nabla L = 0 \in \mathbb{R}^{n+m}$

$$\begin{pmatrix} \frac{\partial L}{\partial x_1} \\ \frac{\partial L}{\partial x_2} \\ \frac{\partial L}{\partial \lambda} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

We should notice that $L(x^*, \lambda^*) = f(x^*)$.

General formulation

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } h_i(x) = 0, \ i = 1, \ldots, m$$

Solution

$\nabla f(x) + \lambda \cdot \nabla h(x) = 0$

$\nabla f = -\lambda \nabla h(x)$

$h(x) = 0$

min $f(x)$
$h_i(x) = 0, \; i = \overline{1,m}$

$$\boxed{L(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) = f(x) + \lambda^\top h(x)}$$

Let $f(x)$ and $h_i(x)$ be twice differentiable at the point x^* and continuously differentiable in some neighborhood x^*. The local minimum conditions for $x \in \mathbb{R}^n, \lambda \in \mathbb{R}^m$ are written as

$$
\begin{aligned}
&\nabla_x L(x^*, \lambda^*) = 0 \\
&\nabla_\lambda L(x^*, \lambda^*) = 0 \\
&\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0, \quad \forall y \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0
\end{aligned}
$$

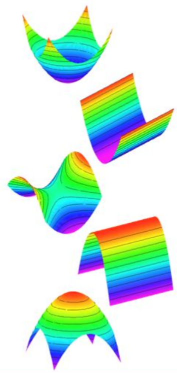
$\nabla L \in \mathbb{R}^{n+m}$ $\quad \nabla^2 L = \mathbb{R}^{(n+m)\times(n+m)}$

$\nabla_x L \in \mathbb{R}^n$ $\quad \nabla^2_{xx} L = \mathbb{R}^{n \times n}$

$\nabla^2 L$ и $\nabla^2_{xx} L$?

Depending on the behavior of the Hessian, the critical points can have a different character.

$\mathbf{y}^T \mathbf{H} \mathbf{y}$	λ_i	**Definiteness H**	**Nature x***
>0		Positive d.	Minimum
≥ 0		Positive semi-d.	Valley
$\neq 0$		Indefinite	Saddlepoint
≤ 0		Negative semi-d.	Ridge
<0		Negative d.	Maximum



Optimization with inequality conditions

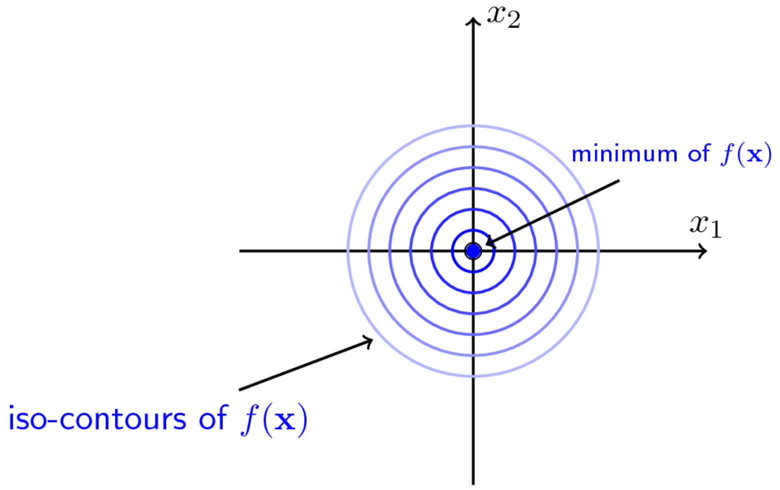
Example

$$f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

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

$$\text{s.t. } g(x) \leq 0$$

Tutorial example - Cost function



$$\boxed{f(\mathbf{x}) = x_1^2 + x_2^2}$$

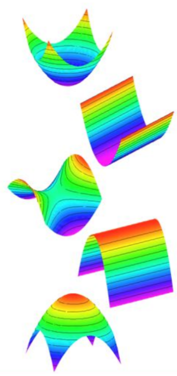
$$L(x, \nu) = f(x) + \sum_{i=1}^{p} \nu_i h_i(x) = f(x) + \nu^\top h(x)$$

Let $f(x)$ and $h_i(x)$ be twice differentiable at the point x^* and continuously differentiable in some neighborhood x^*. The local minimum conditions for $x \in \mathbb{R}^n, \nu \in \mathbb{R}^m$ are written as

$$\begin{aligned}
&\text{ECP: Necessary conditions} \\
&\nabla_x L(x^*, \nu^*) = 0 \\
&\nabla_\nu L(x^*, \nu^*) = 0 \\
&\text{ECP: Sufficient conditions} \\
&\langle y, \nabla^2_{xx} L(x^*, \nu^*) y \rangle > 0, \\
&\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y = 0
\end{aligned}$$

Depending on the behavior of the Hessian, the critical points can have a different character.

$\mathbf{y}^T\mathbf{H}\mathbf{y}$ λ_i	**Definiteness H**	**Nature x***
>0	Positive d.	Minimum
≥ 0	Positive semi-d.	Valley
$\neq 0$	Indefinite	Saddlepoint
≤ 0	Negative semi-d.	Ridge
<0	Negative d.	Maximum



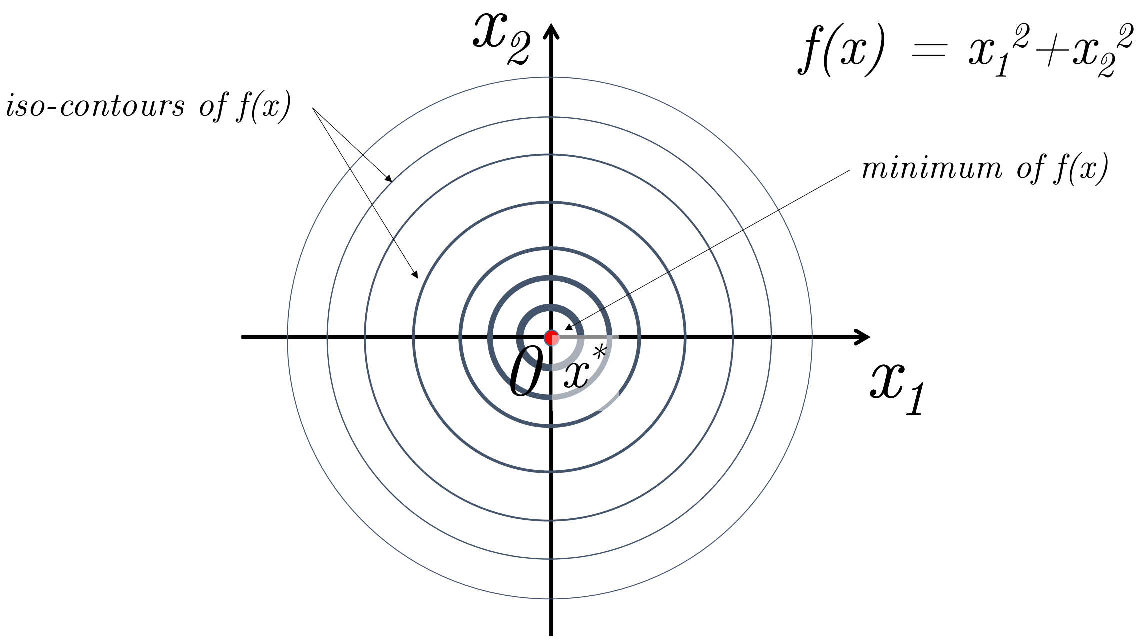
Optimization with inequality conditions

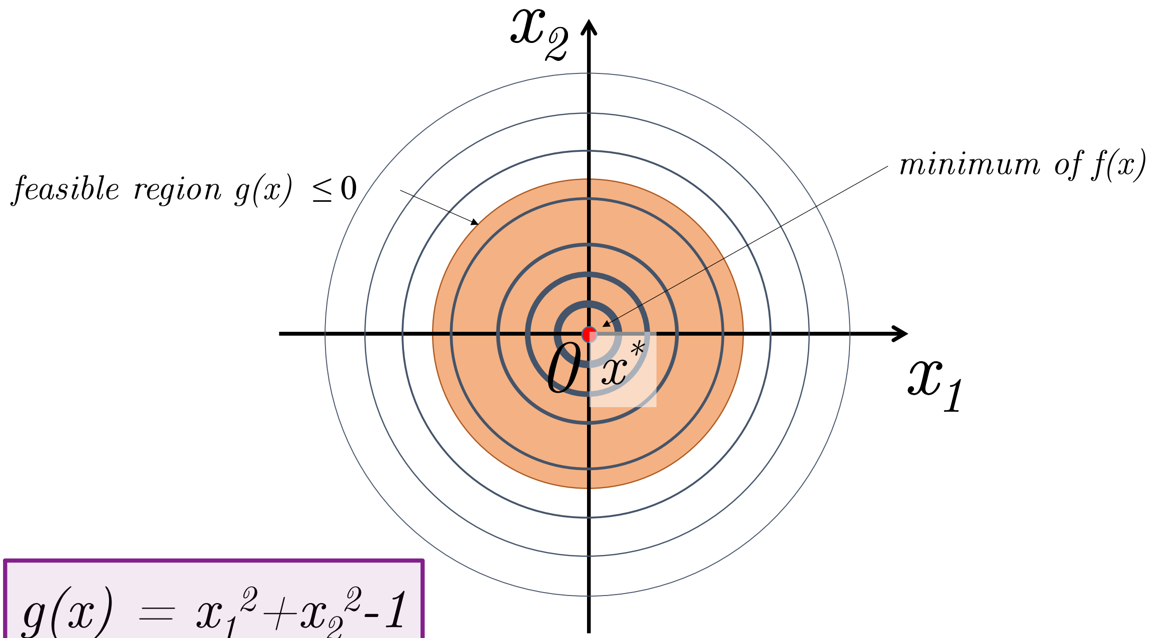
Example

$$\boxed{f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1}$$

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

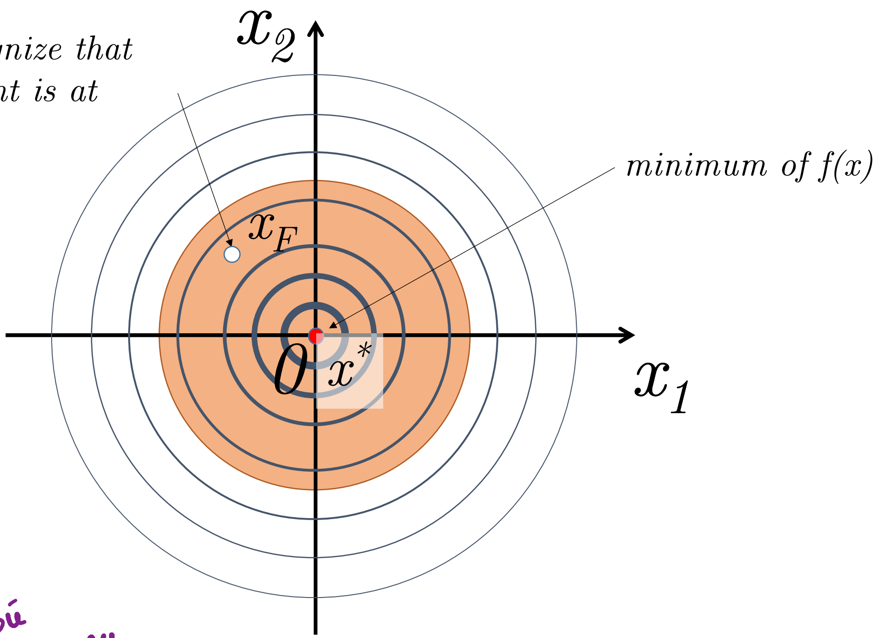
$$\text{s.t. } g(x) \leq 0$$





$$g(x) = x_1^2 + x_2^2 - 1$$

How can we recognize that some feasible point is at local minimum?

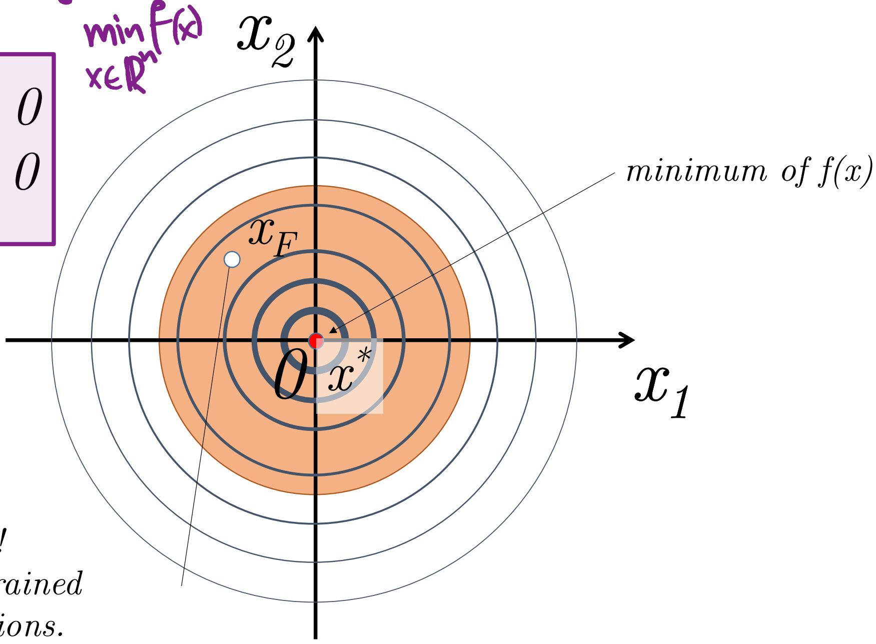


для безусловной задачи

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

$$\nabla f(x_F) = 0$$
$$\nabla^2 f(x_F) \succ 0$$

$g(x_F) \leqslant 0$

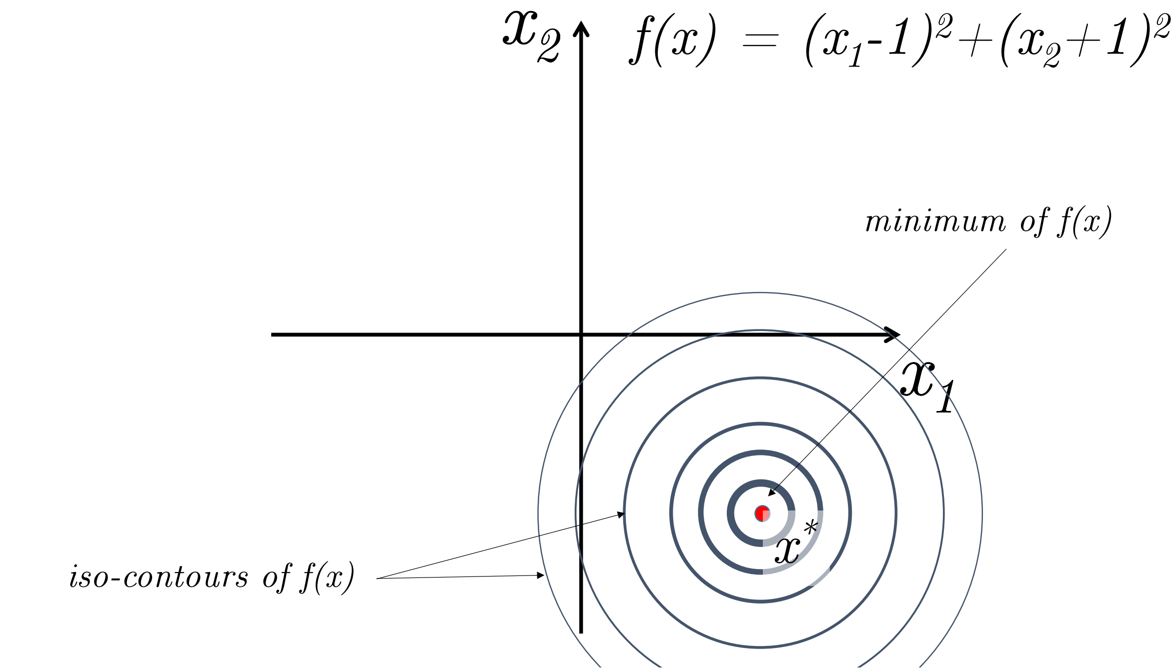


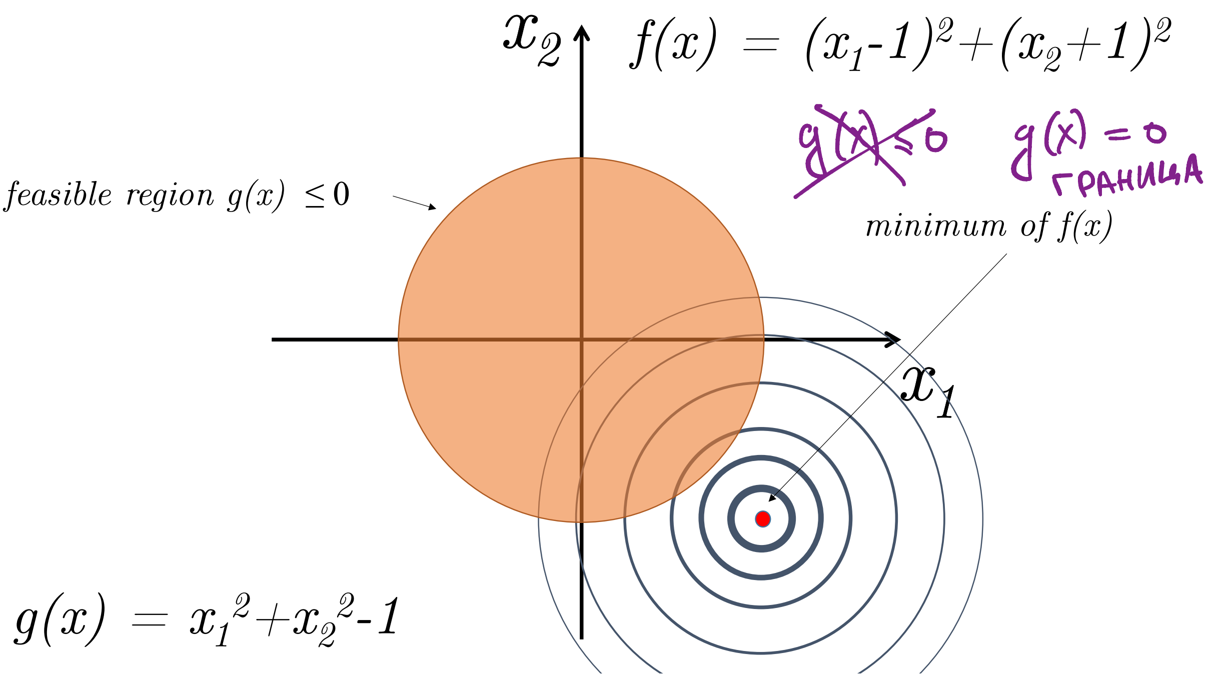
Easy in this case!
Just use unconstrained optimality conditions.

Thus, if the constraints of the type of inequalities are inactive in the constrained problem, then don't worry and write out the solution to the unconstrained problem. However, this is not the whole story 🤔. Consider the second childish example

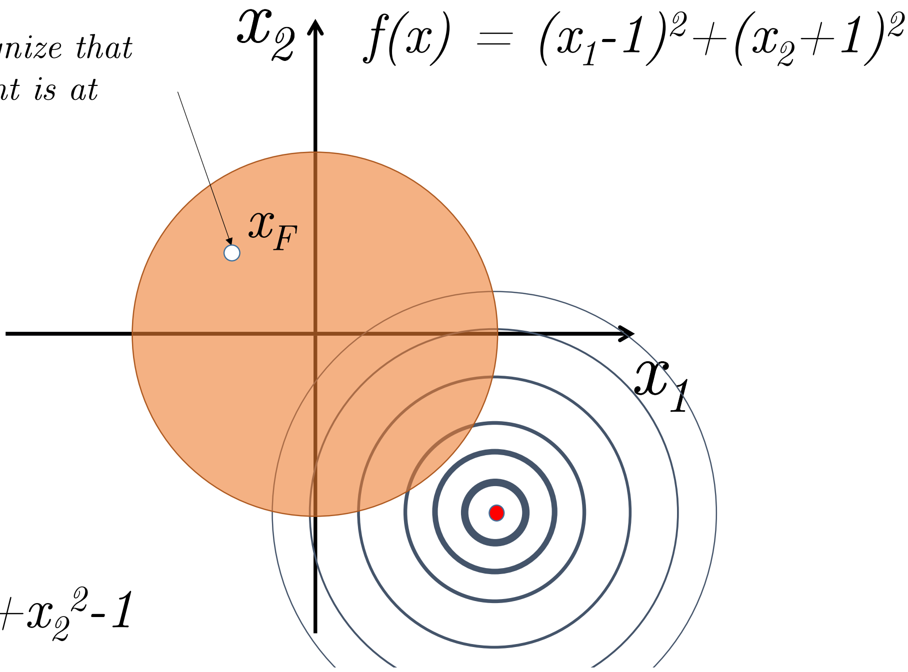
$$
f(x) = (x_1 - 1)^2 + (x_2 + 1)^2 \quad g(x) = x_1^2 + x_2^2 - 1
$$

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

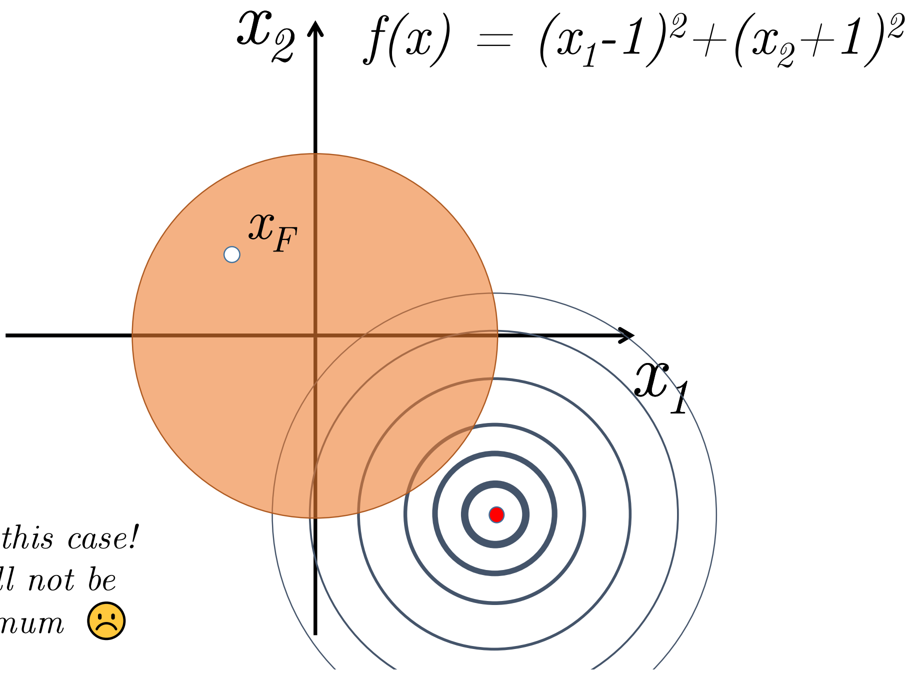




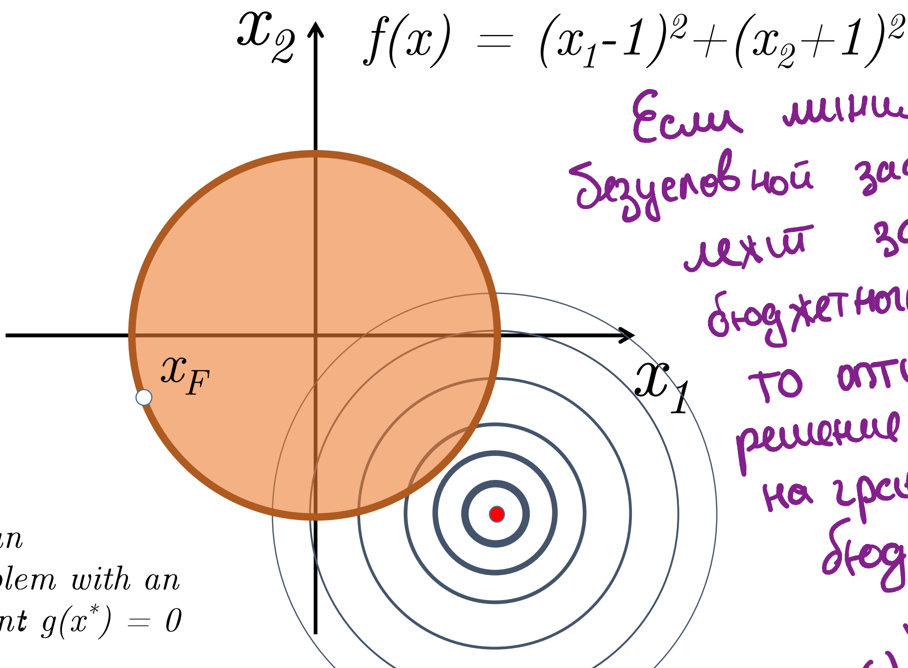
How can we recognize that some feasible point is at local minimum?



$g(x) = x_1^2+x_2^2-1$



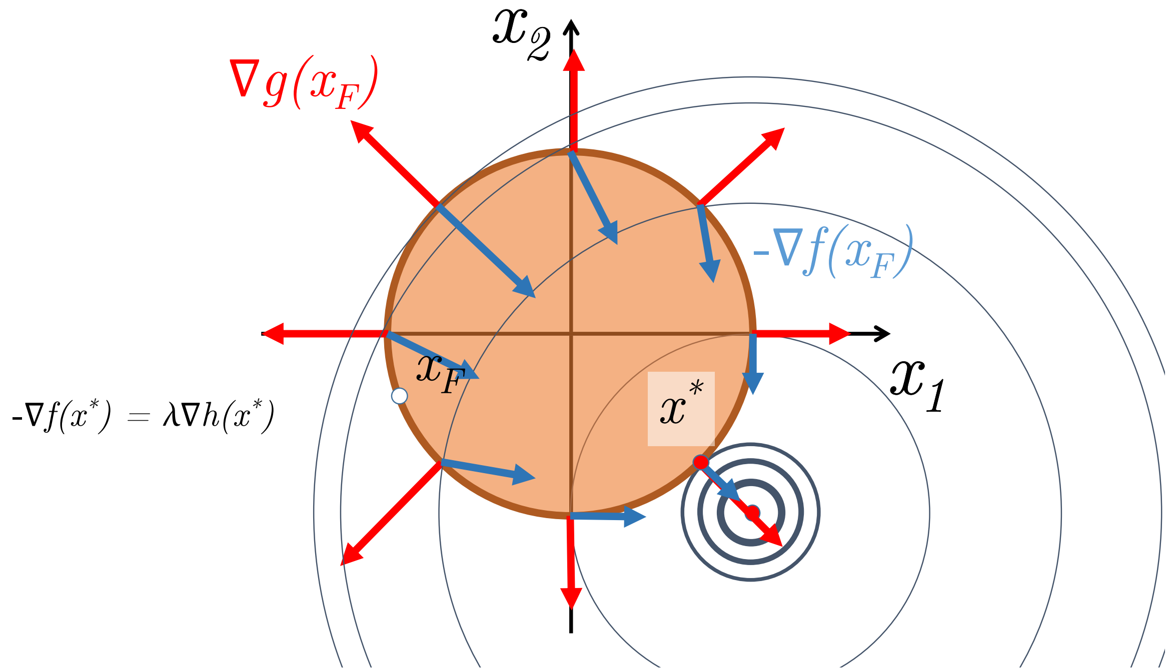
Not very easy in this case! Even gradient will not be zero at local optimum ☹

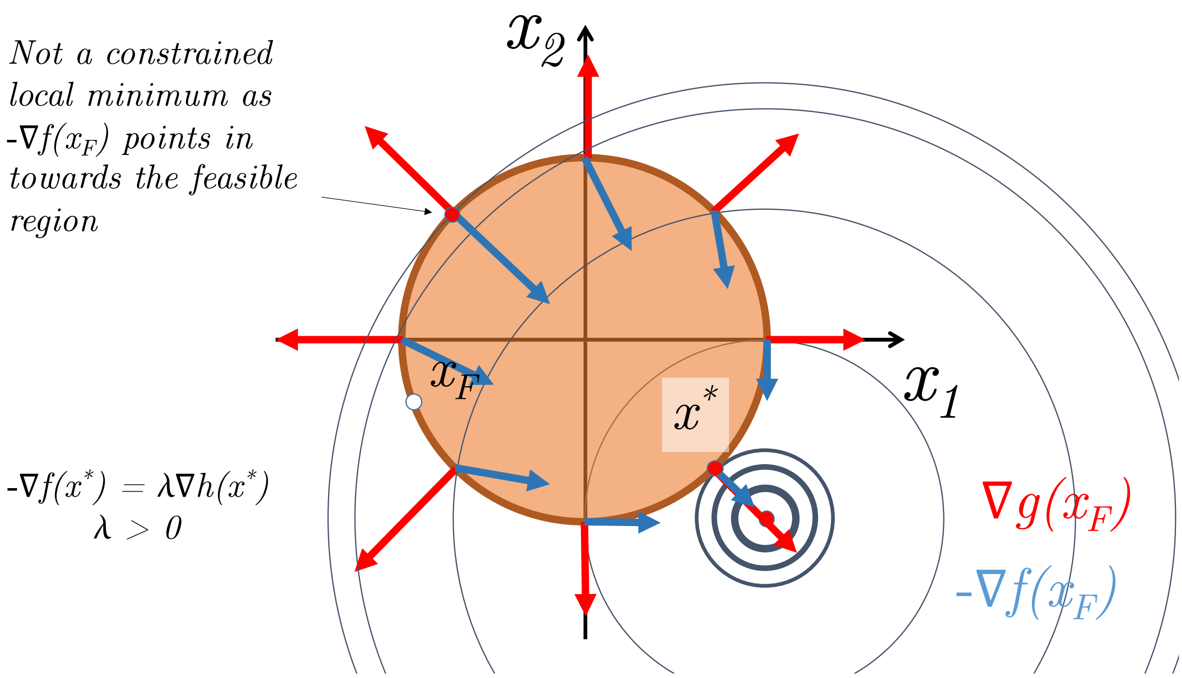


Если минимум безусловной задачи лежит за пределами бюджетного мн-ва, то оптимальное решение лежит на границе бюджетного мн-ва

$g(x^*) = 0$

Effectively have an optimization problem with an equality constraint $g(x^*) = 0$





So, we have a problem:

$$
\boxed{\begin{aligned} & f(x) \to \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } g(x) \leq 0 \end{aligned}}
$$

Two possible cases:

1

$g(x) \leq 0$ **is inactive.** $g(x^*) < 0$	$g(x) \leq 0$ **is active.** $g(x^*) = 0$
$g(x^*) < 0$ внутри S $\nabla f(x^*) = 0$ $\nabla^2 f(x^*) > 0$	Necessary conditions $g(x^*) = 0$ на границе S $-\nabla f(x^*) = \lambda \nabla g(x^*)$, $\lambda > 0$ Sufficient conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0,$ $\forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$

Combining two possible cases, we can write down the general conditions for the problem:

$$
\begin{split}
& f(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & g(x) \leq 0
\end{split}
$$

Let's define the Lagrange function:

$$
L(x, \lambda) = f(x) + \lambda g(x)
$$

Then x^* point - local minimum of the problem described above, if and only if:

1) $\lambda > 0$, $g(x^*) = 0$ границ.

2) $\lambda = 0$, $g(x^*) < 0$ внутр.

$$
\begin{split}
& (1) \ \nabla_x L(x^*, \lambda^*) = 0 \\
& (2) \ \lambda^* \geq 0 \\
& (3) \ \lambda^* g(x^*) = 0 \\
& (4) \ g(x^*) \leq 0 \\
& (5) \ \langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle > 0 \\
& \forall y \neq 0 \in \mathbb{R}^n : \nabla g(x^*)^\top y \leq 0
\end{split}
$$

$\nabla f(x) = -\lambda \nabla g(x)$

бюджетное мн-во

← ПОКА ЗАБЬЁМ

It's noticeable, that $L(x^*, \lambda^*) = f(x^*)$. Conditions $\lambda^* = 0, (1), (4)$ are the first scenario realization, and conditions $\lambda^* > 0, (1), (3)$ - the second.

ОБЩАЯ задача математического программирования

General formulation

$$
\begin{split}
& f_0(x) \to \min_{x \in \mathbb{R}^n} \\
\text{s.t. } & f_i(x) \leq 0, \ i = 1, \ldots, m \\
& h_i(x) = 0, \ i = 1, \ldots, p
\end{split}
$$

This formulation is a general problem of mathematical programming.

λ_i, ν_i - множители Лагранжа

The solution involves constructing a Lagrange function:

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

Karush-Kuhn-Tucker conditions

Necessary conditions

Let x^*, (λ^*, ν^*) be a solution to a mathematical programming problem with zero duality gap (the optimal value for the primal problem p^* is equal to the optimal value for the dual problem d^*).Let also the functions f, f_i, h_i be differentiable.

- $\nabla_x L(x^*, \lambda^*, \nu^*) = 0$
- $\nabla_\nu L(x^*, \lambda^*, \nu^*) = 0$
- $\lambda^*_i \geq 0, i = 1, \ldots, m$
- $\lambda^*_i f_i(x^*) = 0, i = 1, \ldots, m$
- $f_i(x^*) \leq 0, i = 1, \ldots, m$

Some regularity conditions

These conditions are needed in order to make KKT solutions necessary conditions. Some of them even turn necessary conditions into sufficient (for example, Slater's). Moreover, if you have regularity, you can write down necessary second order conditions $\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$ with *semi-definite* hessian of Lagrangian.

- **Slater's condition**. If for a convex problem (i.e., assuming minimization, f_0, f_i are convex and h_i are affine), there exists a point x such that $h(x) = 0$ and $f_i(x) < 0$. (Existance of strictly feasible point), than we have a zero duality gap and KKT conditions become necessary and sufficient.
- **Linearity constraint qualification** If f_i and h_i are affine functions, then no other condition is needed.
- For other examples, see wiki.

Sufficient conditions

For smooth, non-linear optimization problems, a second order sufficient condition is given as follows. The solution x^*, λ^*, ν^*, which satisfies the KKT conditions (above) is a constrained local minimum if for the Lagrangian,

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

the following conditions holds:

$$
\begin{aligned}
&\langle y, \nabla^2_{xx} L(x^*, \lambda^*, \nu^*) y \rangle > 0 \\
&\forall y \neq 0 \in \mathbb{R}^n : \nabla h_i(x^*)^\top y \leq 0, \nabla f_j(x^*)^\top y \leq 0 \\
&i = 1, \ldots, p \quad \forall j : f_j(x^*) = 0
\end{aligned}
$$

References

- Lecture on KKT conditions (very intuitive explanation) in course "Elements of Statistical Learning" @ KTH.
- One-line proof of KKT

Example 1

Linear Least squares Write down exact solution of the linear least squares problem:

$$
f(x) = \|Ax - b\|^2 \to \min_{x \in \mathbb{R}^n}, A \in \mathbb{R}^{m \times n}
$$

$f = \langle Ax-b, Ax-b \rangle$

$df = 2\langle Ax-b, d(Ax-b)\rangle =$

$= 2\langle Ax-b, Adx\rangle =$

$= \langle 2A^T(Ax-b), dx\rangle$

$\nabla f = 2A^T(Ax-b) = 0$

$Ax = b$

Consider three cases:

бесконечно много → 1. $m < n$

решение единств → 2. $m = n$

решений нет → 3. $m > n$

(1) Давайте найдём реш. с мин. нормой

$\|x\|^2 \to \min_{x \in \mathbb{R}^n}$

$Ax = b$

$L(x, \lambda) = \|x\|^2 + \lambda^T(Ax - b).$

$\begin{cases} \nabla_x L = 0 \\ \nabla_\lambda L = 0 \end{cases} \quad \begin{cases} d(\langle x,x\rangle + \langle \lambda, Ax-b\rangle) = 0 \\ Ax = b \end{cases} \quad \begin{cases} \langle 2x, dx\rangle + \langle A^T\lambda, dx\rangle = 0 \\ Ax = b \end{cases}$

$\begin{cases} 2x + A^T\lambda = 0 \\ Ax = b \end{cases} \to \begin{cases} x = -\frac{1}{2}A^T\lambda \\ A\cdot(-\frac{1}{2}A^T\lambda) = b \end{cases} \to \begin{cases} x = -\frac{1}{2}A^T\lambda \\ AA^T\cdot\lambda = -2b \end{cases} \quad \begin{cases} \boxed{x = A^T(AA^T)^{-1}b} \\ \lambda = -2(AA^T)^{-1}\cdot b \end{cases}$

$\det A^TA = 0$

$\det AA^T > 0$

Ответ: $x^* = A^Tb$

$x^* = A^T(AA^T)^{-1}b$

$A_{m \times n}$ $\qquad AA^T_{m\times m}$ $\qquad A^TA_{n\times n}$ $\qquad \operatorname{rang} A^TA = m < n$

$\operatorname{rk} AA^T = m$ $\qquad \det A^TA = 0$

② $m > n$ $\qquad \|Ax - b\|_2^2 \to \min\limits_{x\in\mathbb{R}^n}$ $\qquad \nabla f(x) = \underline{2A^T(Ax-b)} = 0$ $\qquad n < m$

$\underbrace{A^TA}_{n\times n} \cdot x = A^Tb$ $\qquad \det A^TA > 0$

A^{-1}, $\lambda(A)$, $\det(A)$ $\to$ только для квадратных матриц $\qquad \cancel{x = A^{-1}b}$

$$\boxed{x^* = (A^TA)^{-1} A^Tb}$$

$x^* = A^+b$

$A^+ = (A^TA)^{-1}A^T$

Ответ: $x^* = A^+b = (A^TA)^{-1}A^Tb$

③ $m = n$ $\qquad x^* = A^{-1}b = A^+b$

$A^+ = \lim\limits_{\alpha\to 0}(A^TA + \alpha\cdot I)^{-1}\cdot A^T = \lim\limits_{\alpha\to 0} A^T(AA^T + \alpha I)^{-1} = A^{-1}$ если $\det A \neq 0$

Ответ: $x^* = A^+b$ $\qquad$ np.linalg.pinv(A)

Концепция барьеров:

$\min f(x)$

$g(x) = 0$

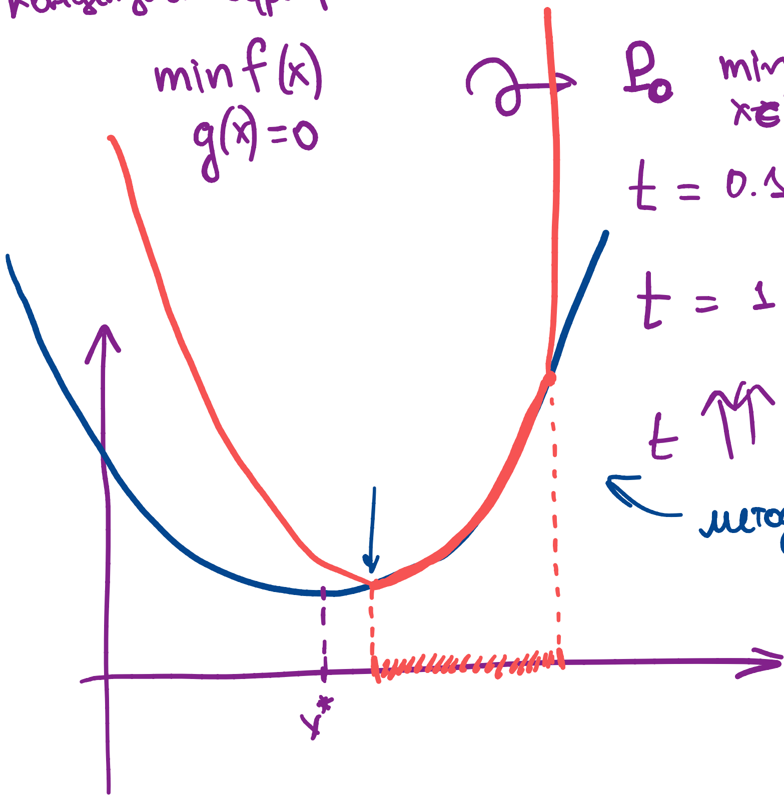
$\to$ P_0 $\quad \min\limits_{x\in\mathbb{R}^n} f(x) + \frac{1}{t}\cdot[g(x)]^2$ $\qquad \cancel{g(x) \le 0}$

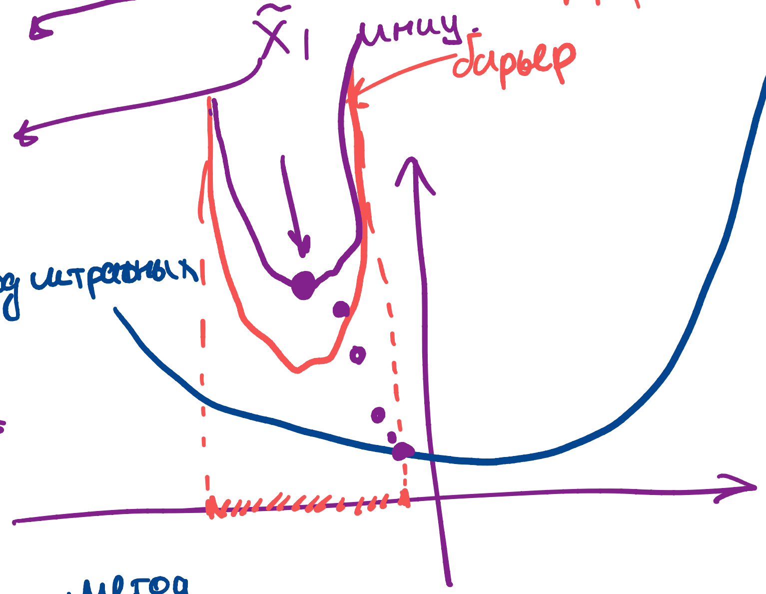
$t = 0.1 \to \tilde{x}_0^*$ инициализируем

$t = 1$ $\qquad \tilde{x}_1$

$t \uparrow\uparrow$

метод штрафных





метод барьеров