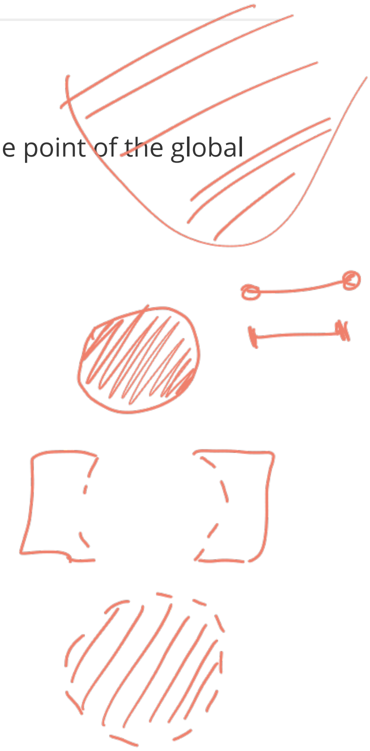
Optimality conditions. KKT

Background

Extreme value (Weierstrass) theorem

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.



Lagrange multipliers

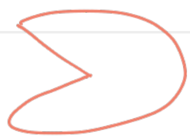
Consider simple yet practical case of equality constraints:

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

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



$$
f(x) \to \min\limits_{x \in S}
$$

objective budget set

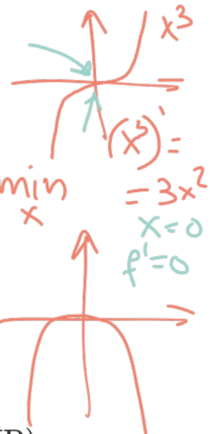
We say that the problem has a solution if the following set **is not empty**: $x^* \in S$, in which the minimum or the infimum of the given function is achieved.

x^3

$(x^3)' = 3x^2$

$x = 0$

$f' = 0$



Unconstrained optimization

$S = R^n$ $\quad f \to \min\limits_x$

General case

Недостаточно

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

$$
\boxed{f(x) \to \min\limits_{x \in \mathbb{R}^n}} \qquad \boxed{\nabla f(x^*) = 0} \quad ?? \quad \text{(UP)}
$$

MIN MAX

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

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

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

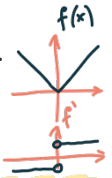
① $\nabla f(x^*) = 0$
② $\nabla^2 f(x^*) \succeq 0$
min

$\nabla f(x^*) = 0$ max
$\nabla^2 f(x^*) \preceq 0$
$(-\nabla^2 f(x^*) \succeq 0)$

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

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

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



ВАЖНО:
$\min\limits_{x \in \mathbb{R}^n} f(x)$
1. $\nabla f(x^*) = 0$
$f(x) = \frac{1}{m} \sum_{i=1}^{m} f_i(x)$
m - кол-во данных
$= 0$
[max, min, saddle point
2. проверяем дост. усл.
$\nabla^2 f(x^*)$? $\succeq 0$

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:

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

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 $S^* = x^*$.

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.

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

$f: \mathbb{R}^2 \to \mathbb{R}^1$

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

$x_1 + x_2 \to \min\limits_{x_1, x_2}$
$x_1^2 + x_2^2 - 2 = 0$

ЛАГРАНЖИАН:

$$L(x, \lambda) = f(x) + \lambda \cdot h(x)$$

$$\boxed{L(x, \lambda) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2)}$$

если $f: \mathbb{R}^n \to \mathbb{R}$
$h: \mathbb{R}^m \to \mathbb{R}$
$L: \mathbb{R}^{n+m} \to \mathbb{R}$
$\nabla L \in \mathbb{R}^{m+n}$
$\nabla^2 L \in \mathbb{R}^{(m+n)\times(m+n)}$

решаем задачу $\boxed{\min\limits_{x, \lambda} L(x, \lambda)}$

$\begin{cases} \nabla_x L = 0 \\ \nabla_\lambda L = 0 \end{cases}$ необх. условия

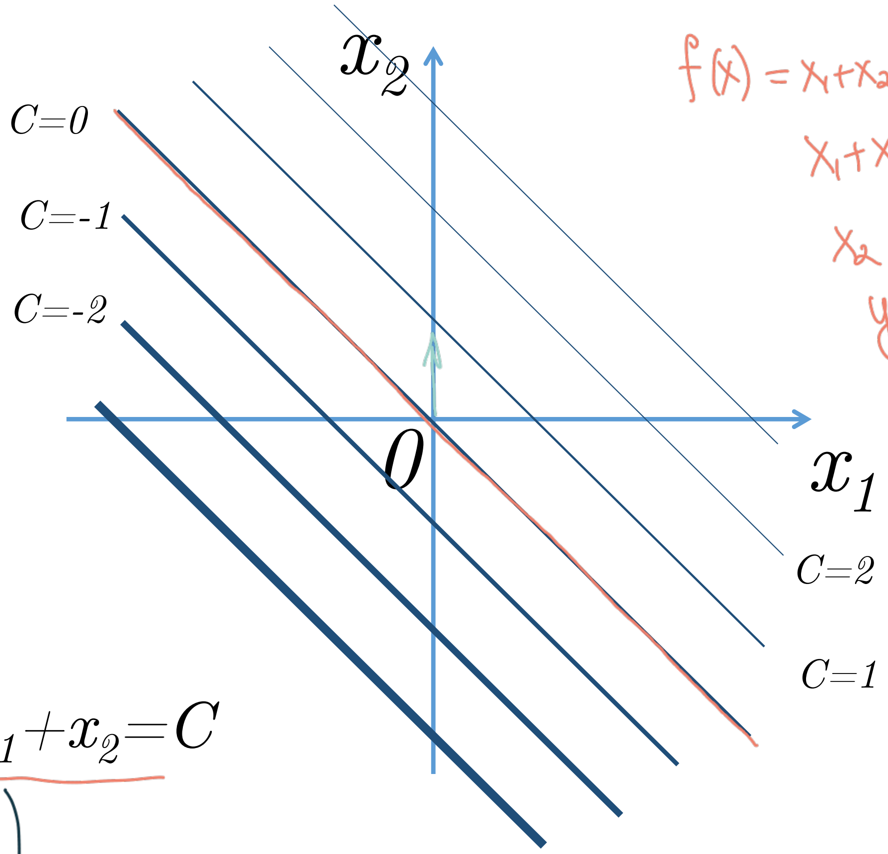
$\nabla^2 L$??

$f(x) = x_1 + x_2 = \text{const}$

$x_1 + x_2 = \cancel{0}\ 1$

$x_2 = 1 - x_1$

$y(x) = 1 - x$

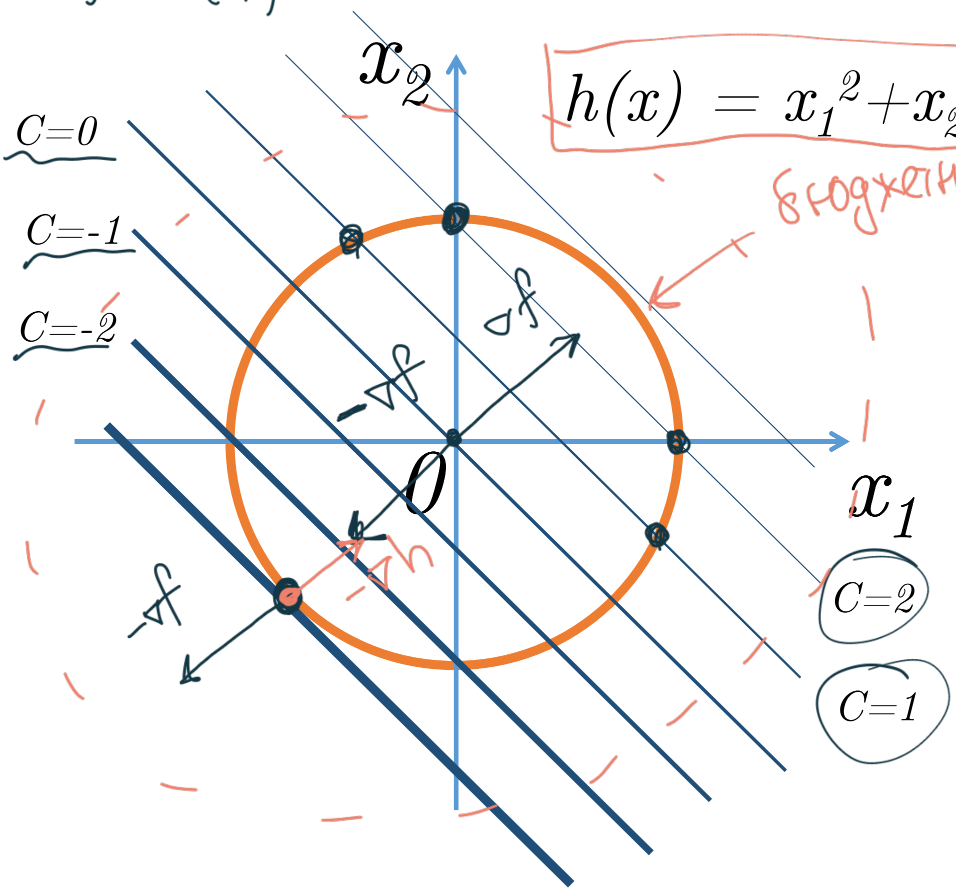


$f(x) = x_1 + x_2 = C$

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

$\boxed{h(x) = x_1^2 + x_2^2 - 2 = 0}$

бюджетное мн-во



в т. оптимума

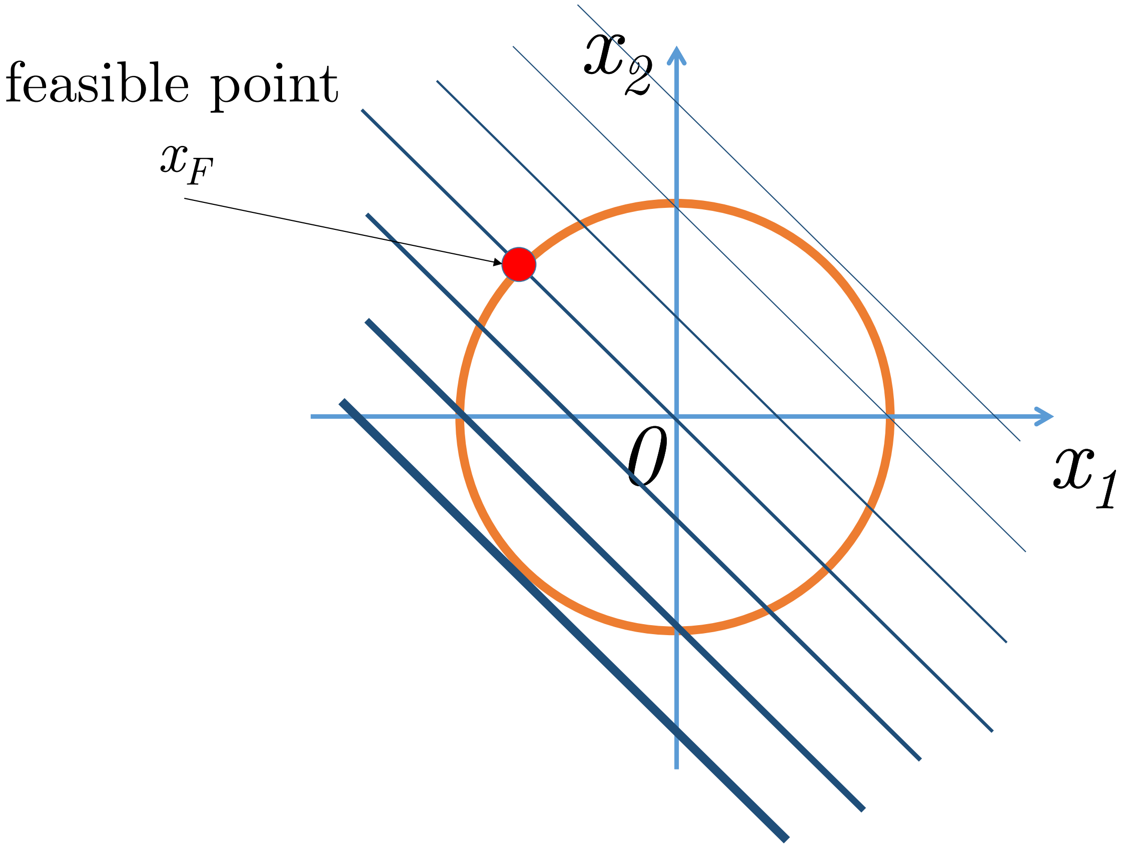
$\nabla f \sim \nabla h$

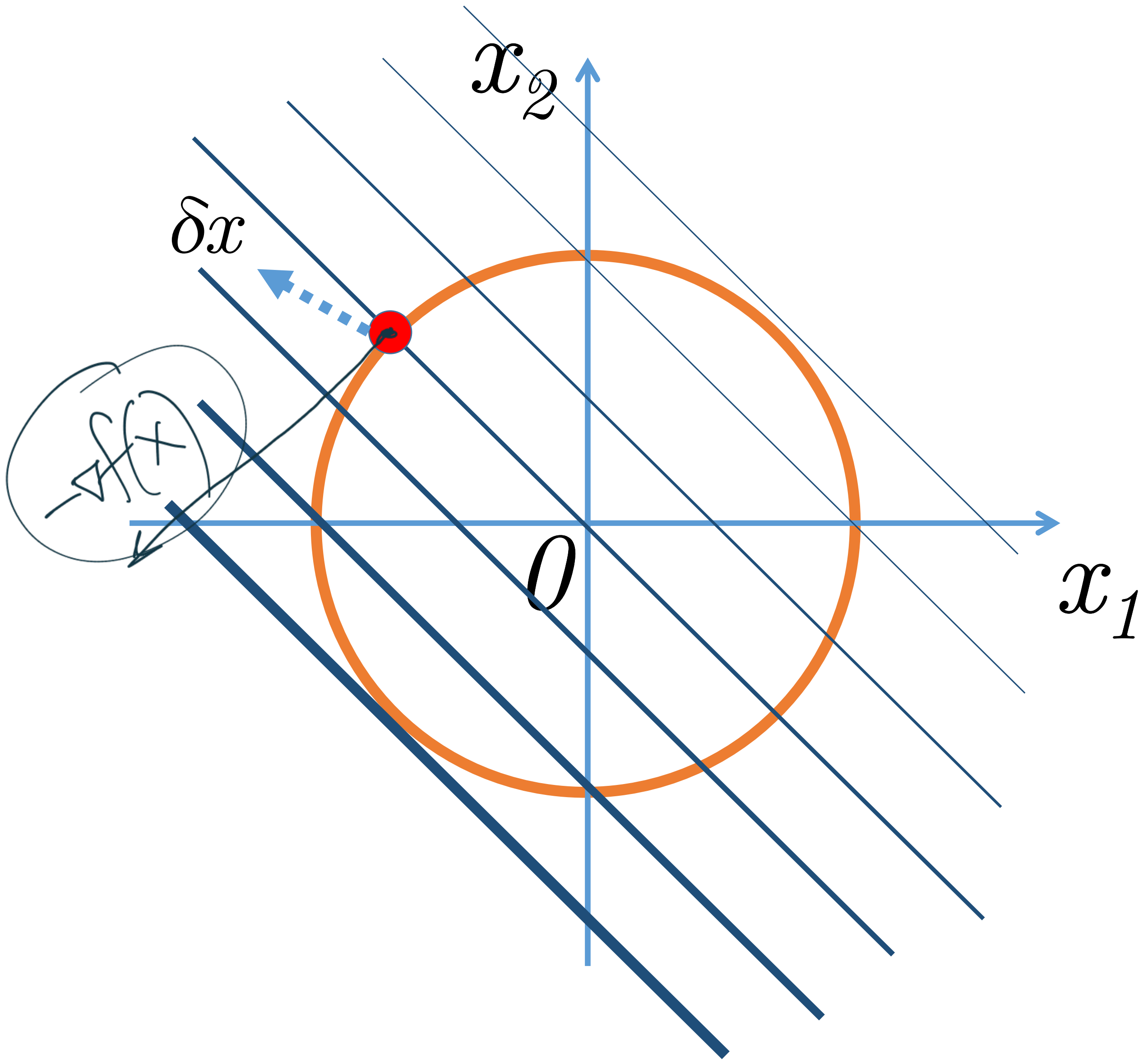
параллельны

$\nabla f = -\lambda \nabla h$

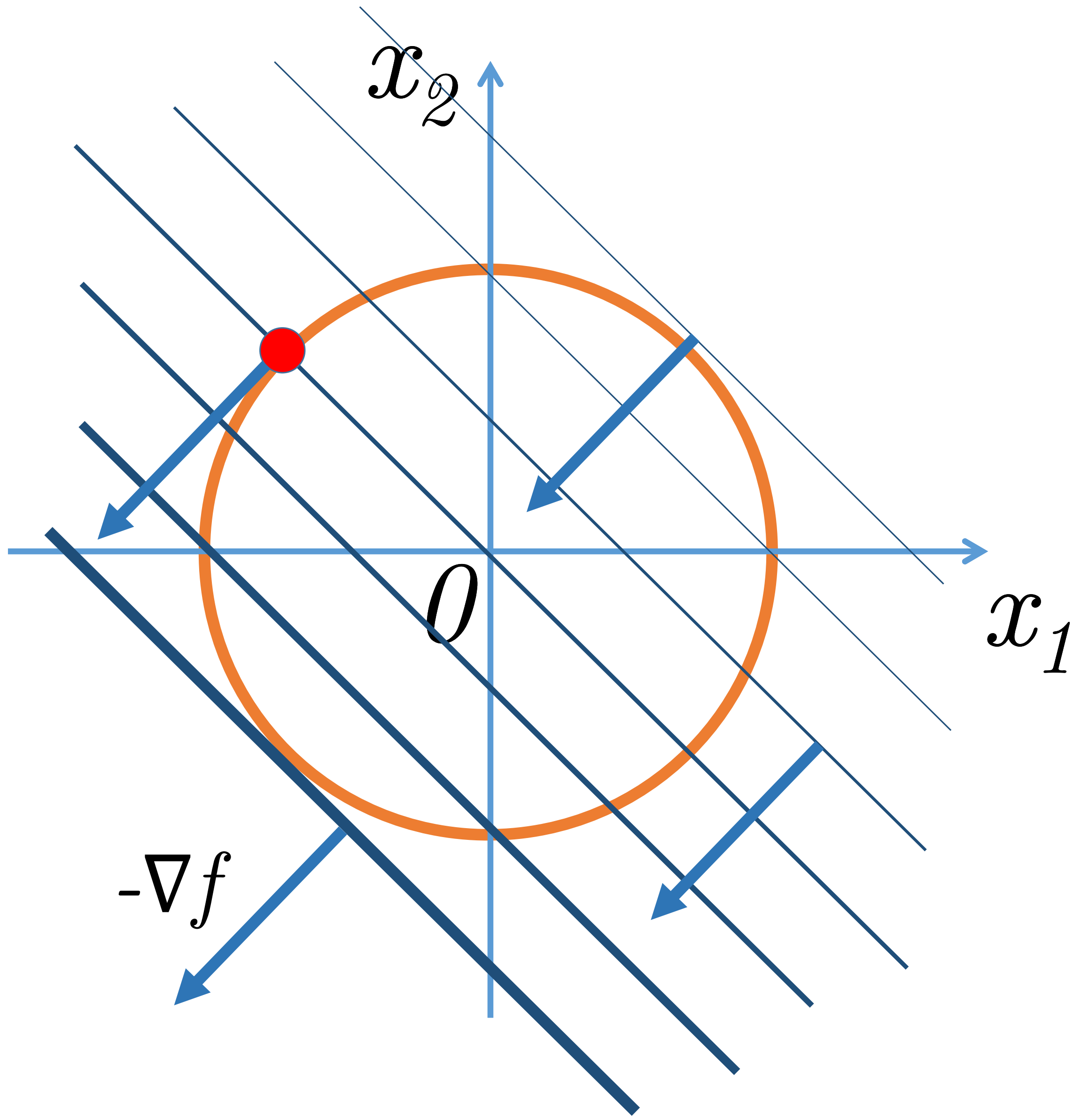
$\boxed{\nabla f + \lambda \cdot \nabla h = 0}$

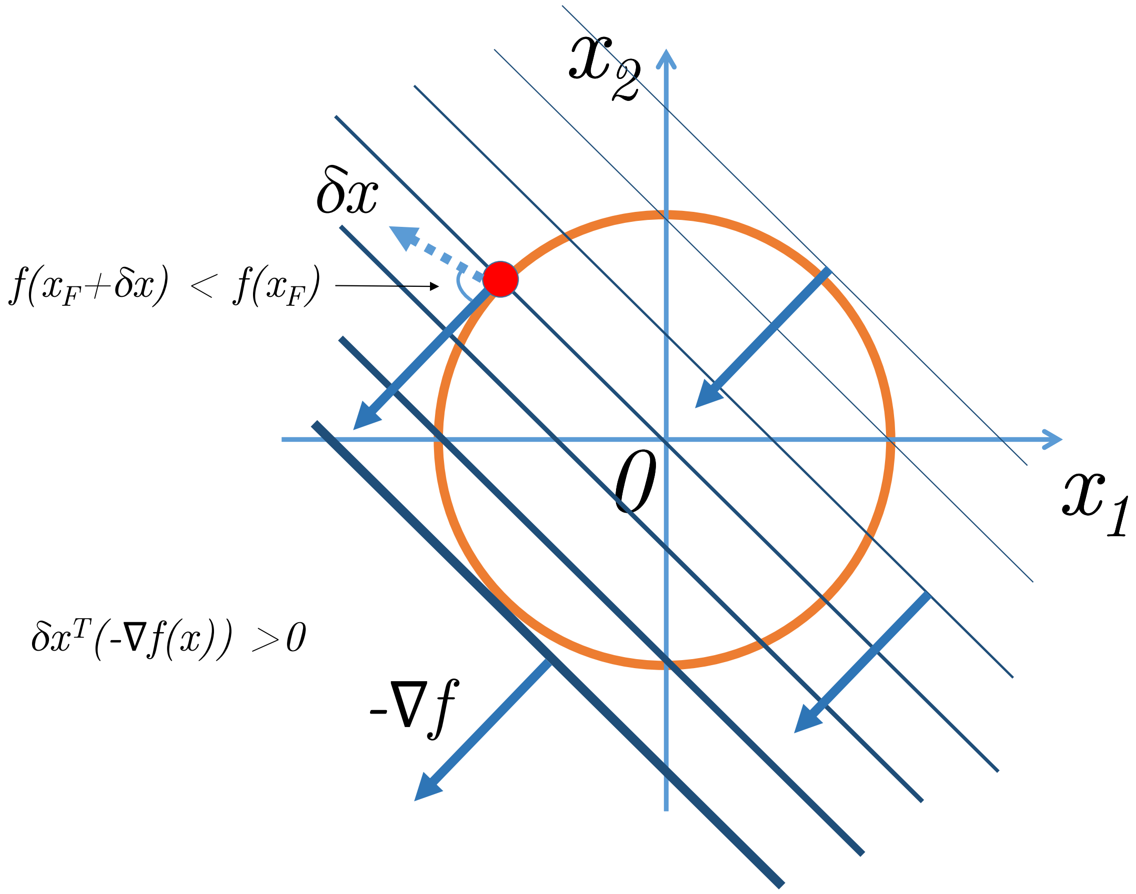
$\nabla_x L = 0$


feasible point
x_F
x_2
x_1
0

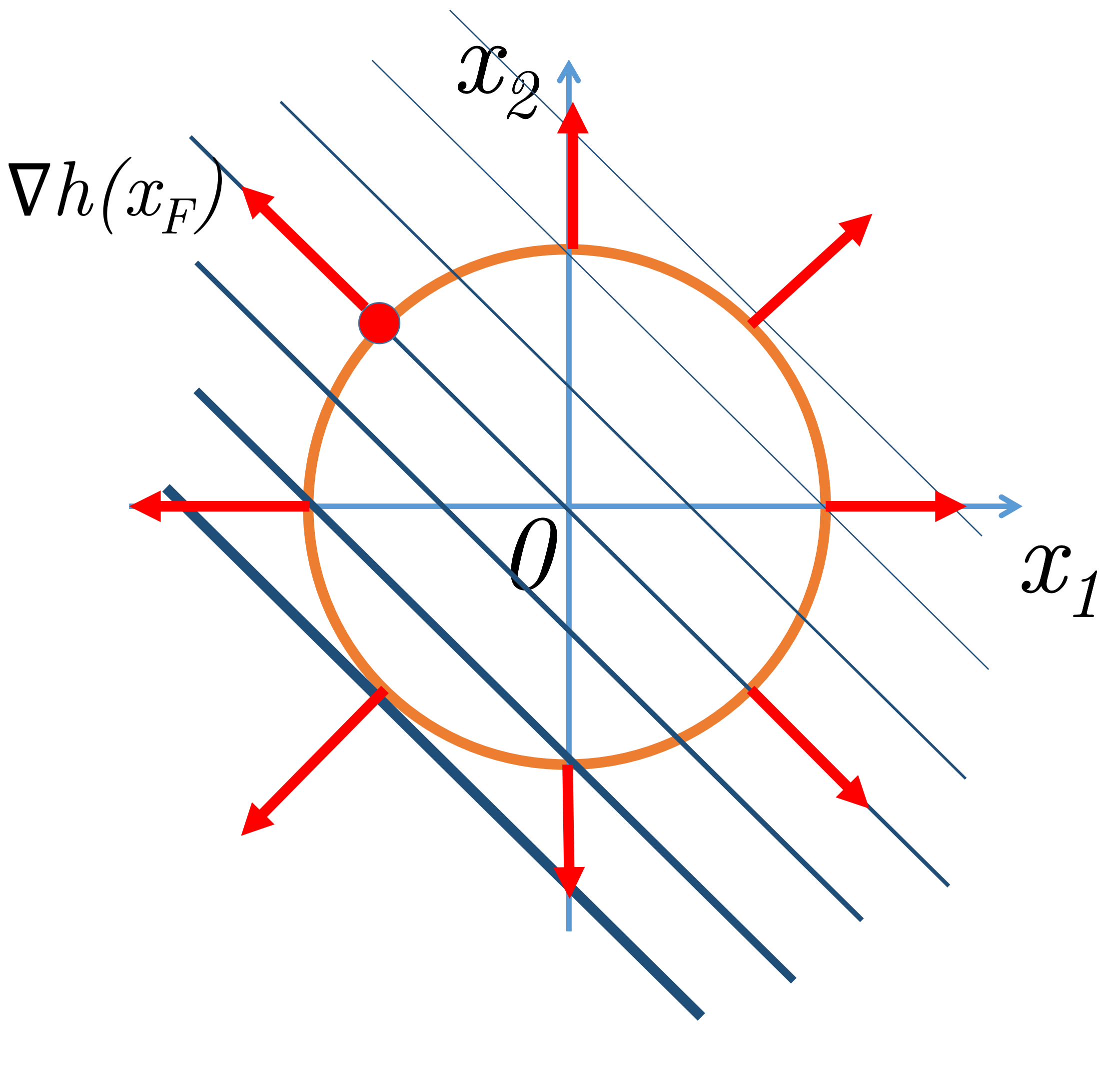


x_2
δx
$-\nabla f(x)$
0
x_1

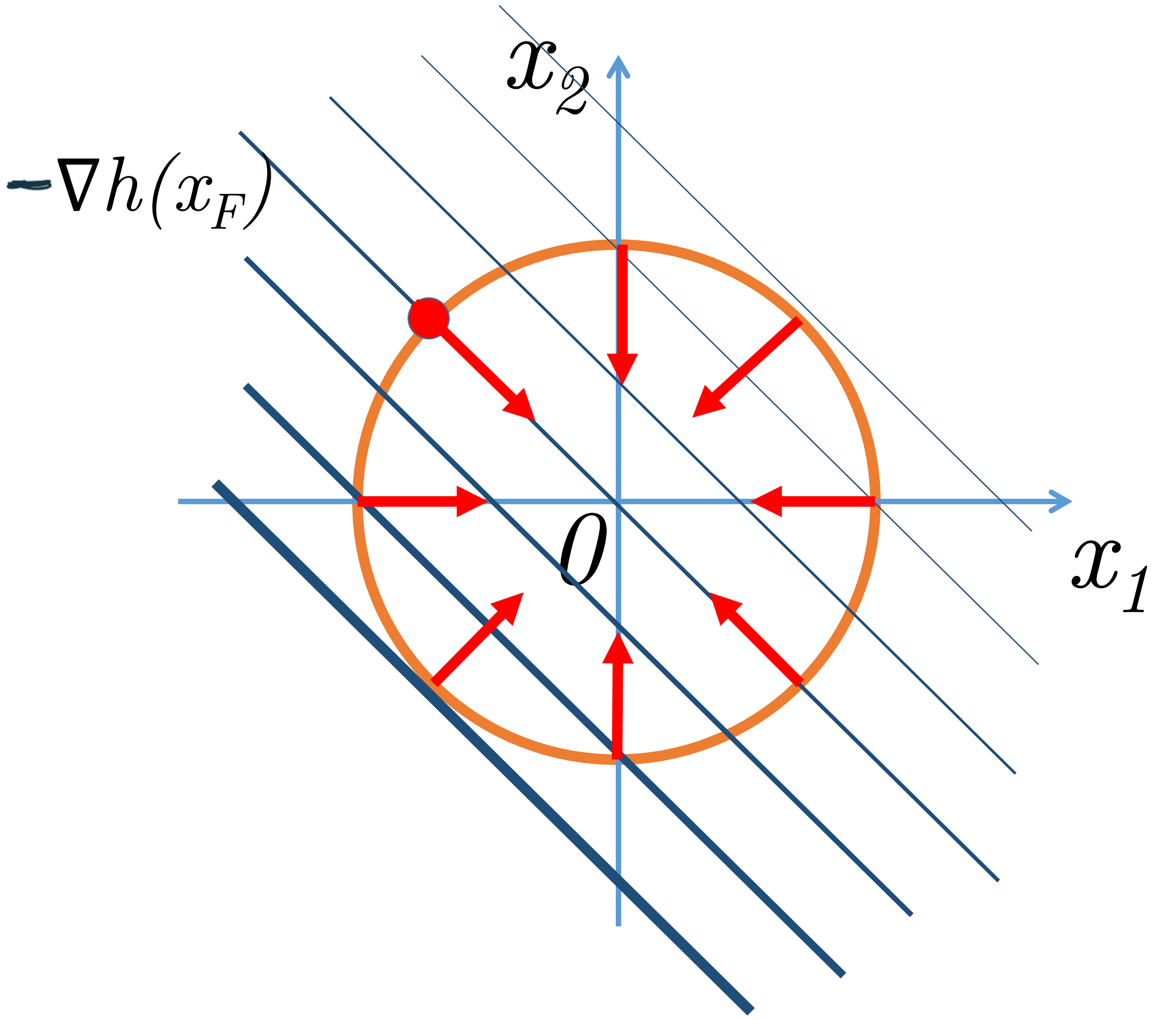

x_2
x_1
0
$-\nabla f$

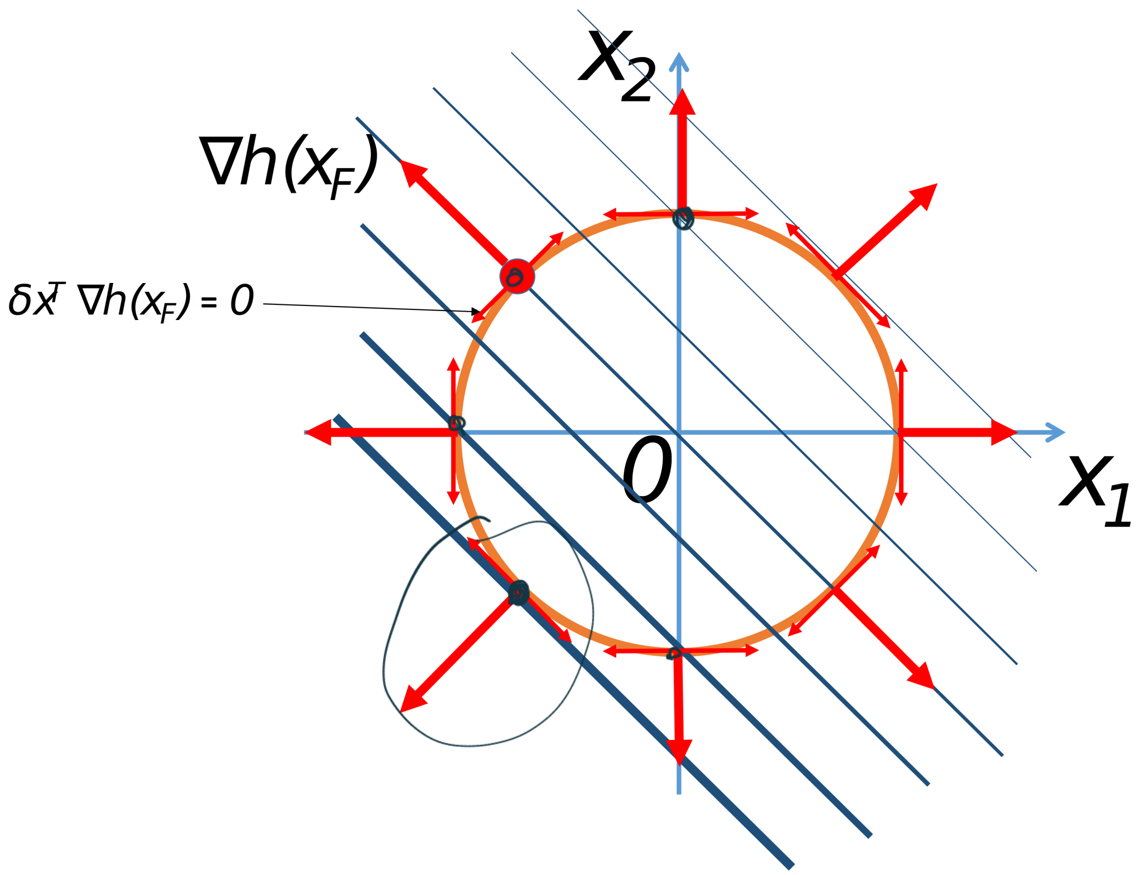


x_2
δx
$f(x_F+\delta x) < f(x_F)$
0
x_1
$\delta x^T(-\nabla f(x)) > 0$
$-\nabla f$



x_2
$\nabla h(x_F)$
0
x_1





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

$$\langle \delta x, \nabla h(x_F) \rangle = 0$$

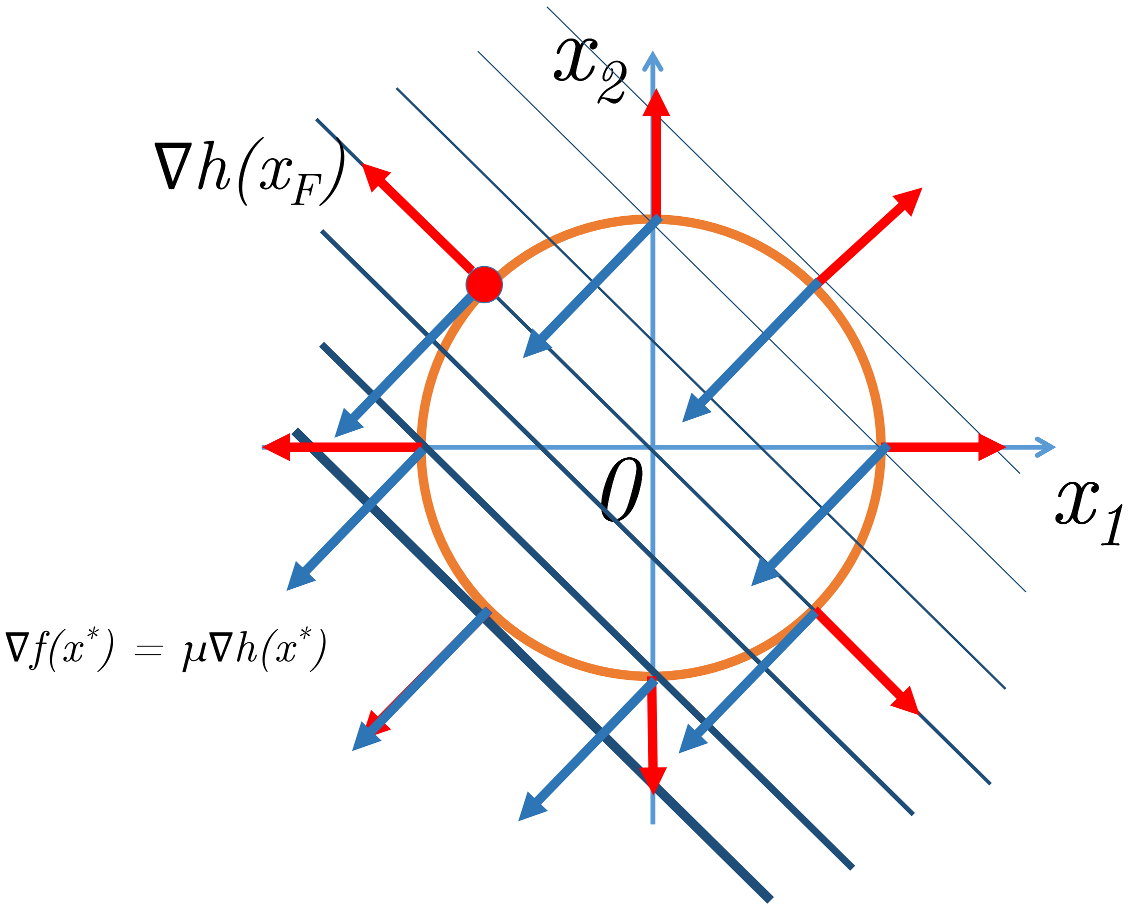
$$\langle \delta x, -\nabla f(x_F) \rangle > 0$$

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

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

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

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 :)



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

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

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

$$\begin{aligned}
&\nabla_x L(x^*, \lambda^*) = 0 \text{ that's written above} \\
&\nabla_\lambda L(x^*, \lambda^*) = 0 \text{ condition of being in budget set} \\
&\langle y, \nabla^2_{xx} L(x^*, \lambda^*) y \rangle \geq 0, \quad \forall y \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0
\end{aligned}$$

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

$Ax = b$ $\quad A \in \mathbb{R}^{m \times n}$

m линейных огранич.

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

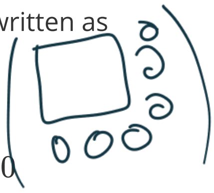
$h(x) = 0$ $L(x, \lambda) = f(x)$

если $x \in S$, то

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

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

гессиан лагранжиана



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

$y^T H y$ 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

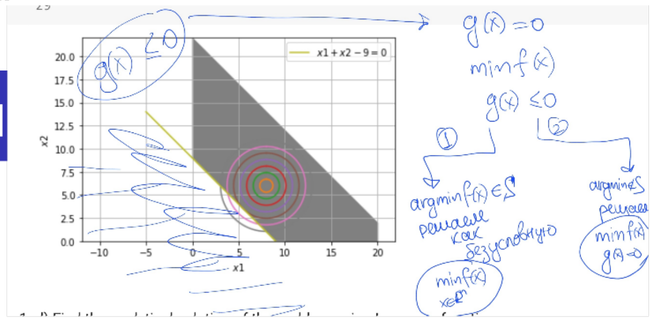
Optimization with inequality conditions

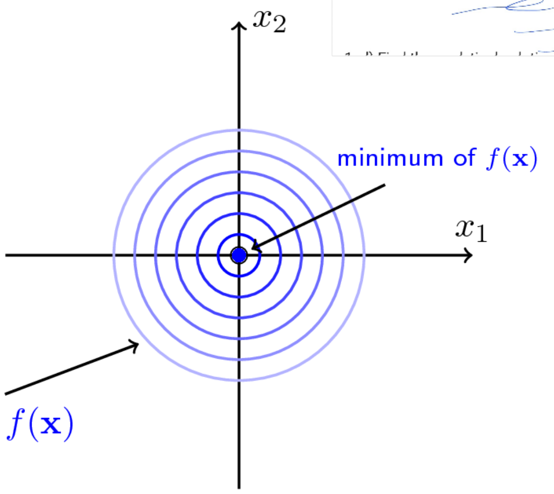
Example

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

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

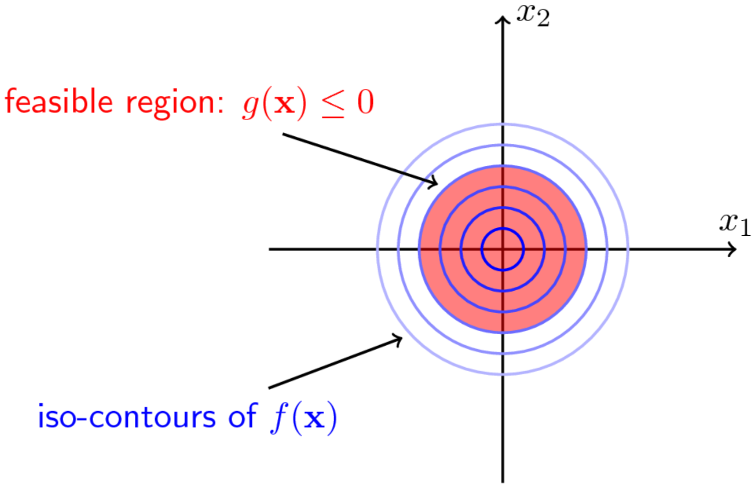
Tutorial





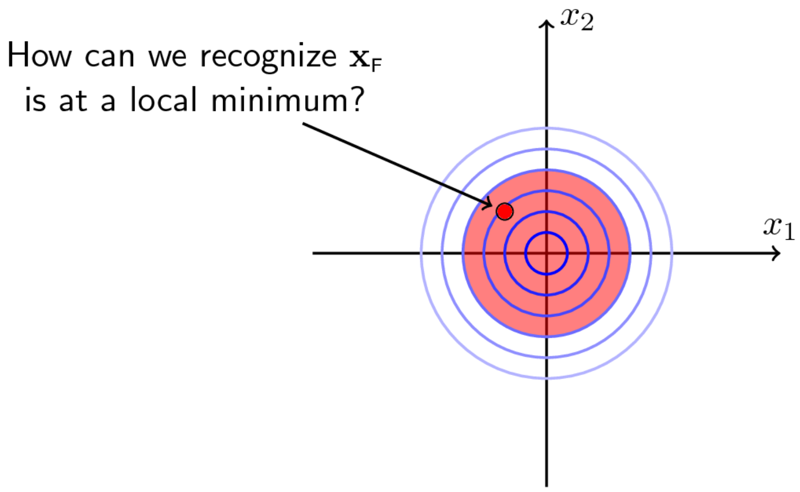
$$
\boxed{f(\mathbf{x}) = x_1^2 + x_2^2}
$$

Tutorial example - Feasible region

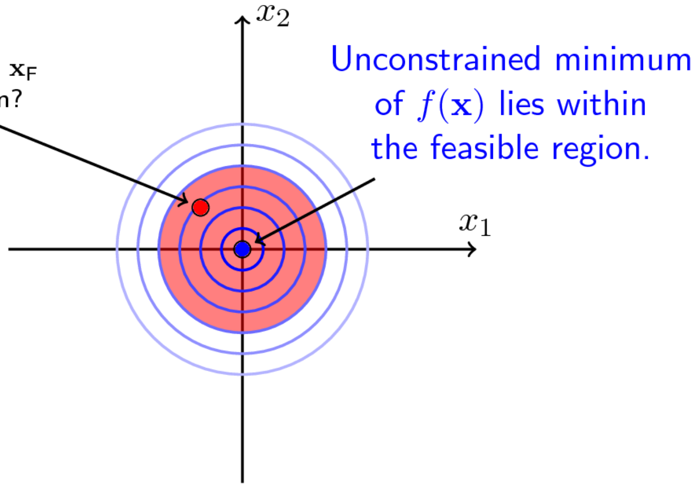


$$\boxed{g(\mathbf{x}) = x_1^2 + x_2^2 - 1}$$

How do we recognize if $\mathbf{x}_F$ is at a local optimum?



Remember $\mathbf{x}_F$ denotes a feasible point.



$\therefore$ Necessary and sufficient conditions for a constrained local minimum are the same as for an unconstrained local minimum.

$$\boxed{\nabla_{\mathbf{x}} f(\mathbf{x}_F) = \mathbf{0} \text{ and } \nabla_{\mathbf{xx}} f(\mathbf{x}_F) \text{ is positive definite}}$$

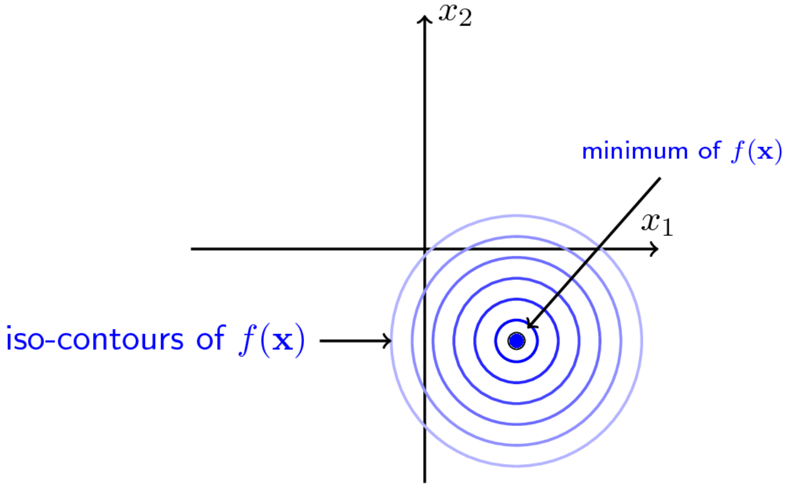
Thus, if the constraints of the type of inequalities are inactive in the UM problem, then don't worry and write out the solution to the UM problem. However, this is not a heal-all :) Consider the second childish example

$$f(x) = (x_1 - 1.1)^2 + (x_2 + 1.1)^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

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

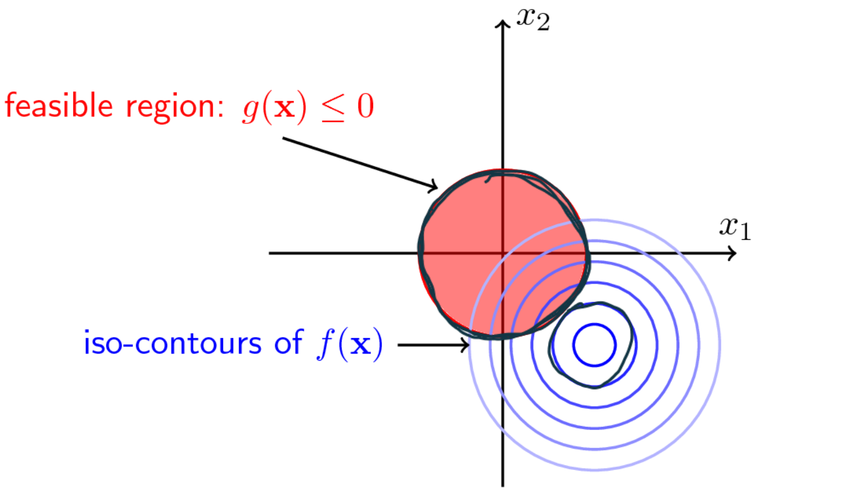
$$\text{s.t. } g(x) \leq 0$$

Tutorial example - Cost function



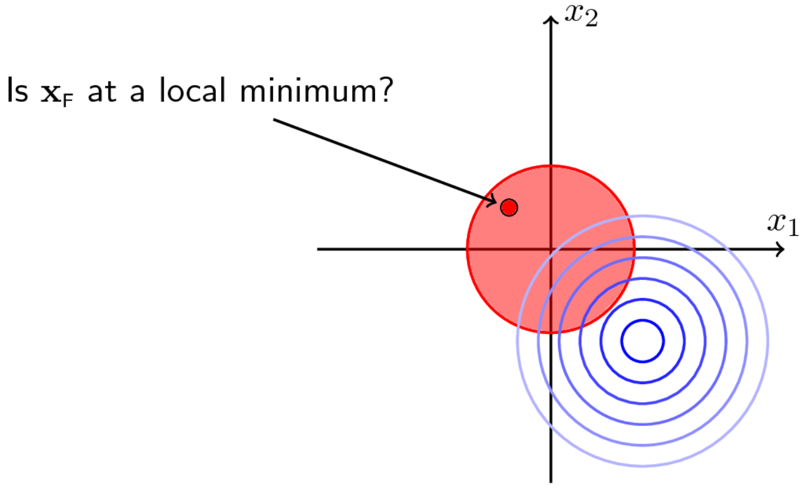
$$f(\mathbf{x}) = (x_1 - 1.1)^2 + (x_2 + 1.1)^2$$

Tutorial example - Feasible region



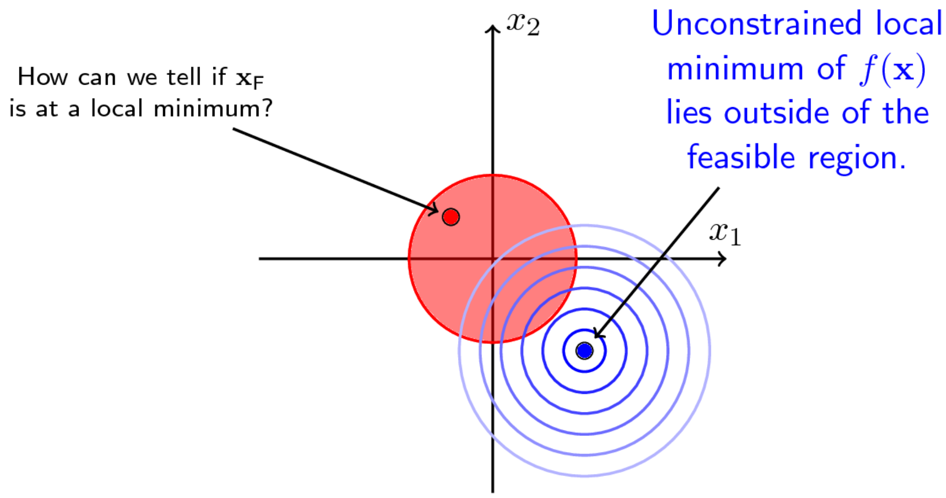
$$g(\mathbf{x}) = x_1^2 + x_2^2 - 1$$

How do we recognize if $\mathbf{x}_F$ is at a local optimum?



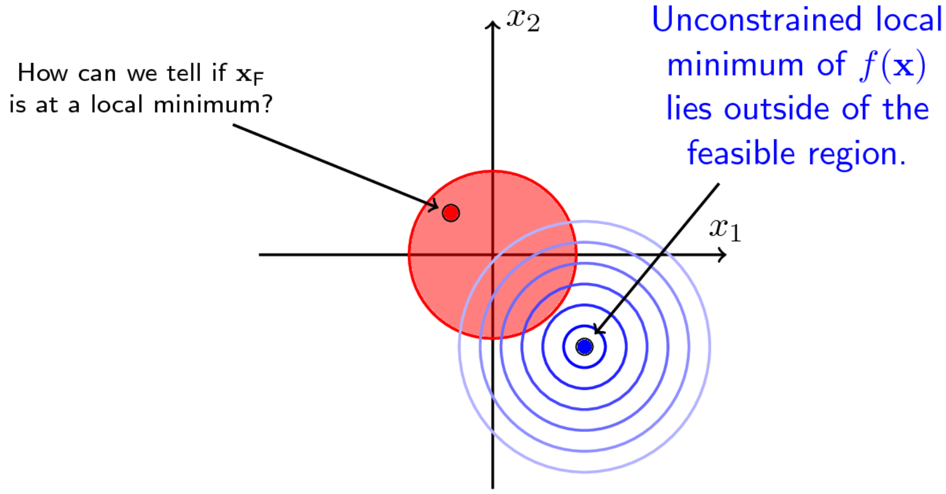
Remember $\mathbf{x}_F$ denotes a feasible point.

How do we recognize if $\mathbf{x}_F$ is at a local optimum?



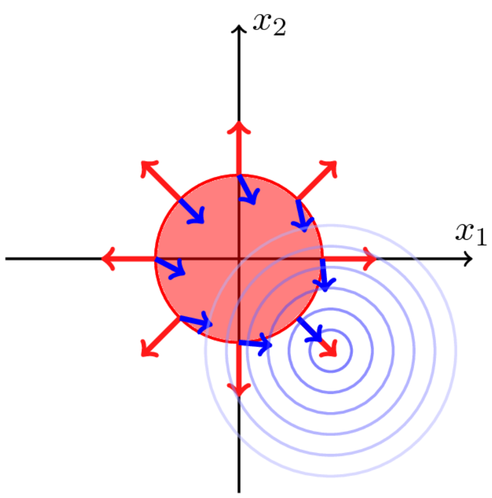
$\therefore$ the constrained local minimum occurs on the surface of the constraint surface.

How do we recognize if $\mathbf{x}_F$ is at a local optimum?



$\therefore$ Effectively have an optimization problem with an **equality constraint**: $g(\mathbf{x}) = 0$.

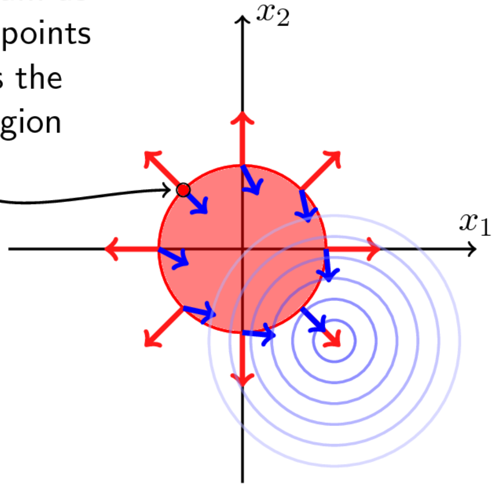
Given an equality constraint



A local optimum occurs when $\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ are parallel:

$$\boxed{-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x})}$$

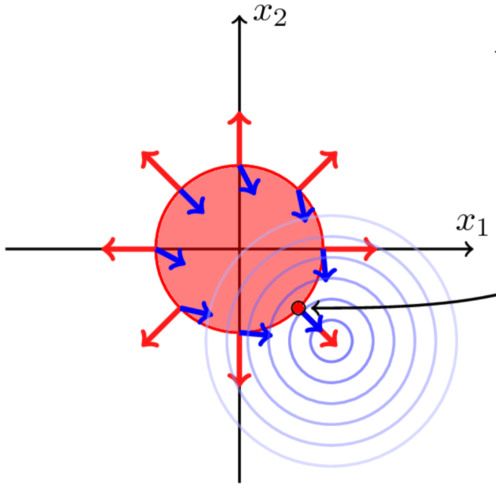
Want a constrained local minimum...



∴ Constrained local minimum occurs when $-\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ point in the same direction:

$$\boxed{-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) \quad \textbf{and} \quad \lambda > 0}$$

Want a constrained local minimum...



∴ Constrained local minimum occurs when $-\nabla_{\mathbf{x}} f(\mathbf{x})$ and $\nabla_{\mathbf{x}} g(\mathbf{x})$ point in the same direction:

$$\boxed{-\nabla_{\mathbf{x}} f(\mathbf{x}) = \lambda \nabla_{\mathbf{x}} g(\mathbf{x}) \quad \textbf{and} \quad \lambda > 0}$$

So, we have a problem:

$$f(x) \to \min_{x \in \mathbb{R}^n}$$
$$\text{s.t. } g(x) \le 0$$

Two possible cases:

$$g(x^*) < 0$$

1. $\nabla f(x^*) = 0$

$$\nabla^2 f(x^*) > 0$$

$$g(x^*) = 0$$

2. $-\nabla f(x^*) = \mu \nabla g(x^*), \quad \mu > 0$

$$\langle y, \nabla^2_{xx} L(x^*, \mu^*) y \rangle \geq 0, \quad \forall y \in \mathbb{R}^n : \nabla g(x^*)^\top y = 0$$

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

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

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

если $x \in$ внутр. мн-ве $g(x) \leq 0$

$L(x,\mu) \leq f(x)$

Let's define the Lagrange function:

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

если $g(x)=0 \Rightarrow L = f$

$g<0$, $\mu=0 \Rightarrow L = f$

внутри внутр. мн-ва

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

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

$g=0$, $\mu>0$

$g<0$, $\mu=0$

аналог $\nabla_x L = 0$

достат. условие

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

General formulation

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

$$\text{s.t. } g_i(x) \leq 0, \; i = 1, \ldots, m$$

$$h_j(x) = 0, \; j = 1, \ldots, p$$

This formulation is a general problem of mathematical programming. From now, we only consider $\mathbf{regular}$ tasks. This is a very important remark from a formal point of view. Those wishing to understand in more detail, please refer to Google.

Solution

n p равенств m нер-в

$$L(x, \mu, \lambda) = f(x) + \sum_{j=1}^{p} \lambda_j h_j(x) + \sum_{i=1}^{m} \mu_i g_i(x)$$

множители Лагранжа

Karush-Kuhn-Tucker conditions

Let x^* be a solution to a mathematical programming problem, and the functions f, h_j, g_i are differentiable. Then there are λ^* and μ^* such that the following conditions are carried out:

- $\nabla_x L(x^*, \lambda^*, \mu^*) = 0$ n
- $\nabla_\lambda L(x^*, \lambda^*, \mu^*) = 0$ p
- $\mu_j^* \geq 0$ m
- $\mu_j^* g_j(x^*) = 0$ m
- $g_j(x^*) \leq 0$ m

$n + 3m + p$ ур-ий нер-в

These conditions are sufficient if the problem is regular, i.e. if:

1. the given problem is a convex optimization problem (i.e., the functions f and g_i are convex, h_i are affine) and the Slater condition is satisfied; or
2. strong duality is fulfilled.

References

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

$$Ax = b \qquad A \in \mathbb{R}^{m\times n},\ b \in \mathbb{R}^{m\times n} \quad x = ?$$

(1) $m = n$ кол-во ур-ий = кол-во неизвестных $\Rightarrow$ $\boxed{x^* = A^{-1}\cdot b}$ $\det A \neq 0$, $\operatorname{rank} A = n$

(2) $m < n$ ур-ий < неизвестных $\quad x + y = 0 \Rightarrow$ решений бесконечно много

Недоопределённая система

"подменяем" на $\frac{1}{2}\|x\|_2^2 \to \min\limits_{x\in\mathbb{R}^n}$, $Ax = b$. $\qquad \min f(x)$, $g(x) = 0$, $g(x) = Ax - b = 0$

множитель Лагранжа

(f - вып. S - вып. $\min\limits_{x\in S} f(x)$) Задача выпуклая

Решение:

$$L(x,\lambda) = \frac{1}{2}\langle x, x\rangle + \langle \lambda, Ax - b\rangle$$

$$\nabla_x L = \frac{1}{2}\cdot 2x + A^T\lambda = 0 \Rightarrow \underset{n\times m\ m\times 1}{A^T\lambda} = \underset{n\times 1}{-x} \Rightarrow x = -A^T\lambda$$

$$\nabla_\lambda L = Ax - b = 0$$

$$A\cdot(-A^T\lambda) - b = 0 \Rightarrow \underset{m\times n\ n\times m}{AA^T}\cdot\underset{m\times 1}{\lambda} = \underset{m\times 1}{-b} \Rightarrow \boxed{\lambda = -(AA^T)^{-1}\cdot b} \quad m < n$$

$$x = -A^T\left(-(AA^T)^{-1}\cdot b\right)$$

$$\boxed{x^* = \underbrace{A^T(AA^T)^{-1}}_{A^\dagger}\cdot b} = A^\dagger\cdot b$$

(3) $m > n$ переопределённая система. У исходной задачи нет решений. Подменим на

$$\frac{1}{2}\|Ax - b\|_2^2 \to \min_{x\in\mathbb{R}^n}$$

$$f(x) = \frac{1}{2}\langle Ax - b, Ax - b\rangle$$

$$df = \cancel{\frac{1}{2}}\cdot\cancel{2}\langle Ax - b, A\,dx\rangle = \langle A^T(Ax - b), dx\rangle$$

$$\nabla f = A^TAx - A^Tb = 0$$

$$\underset{n\times m\ m\times n\ n\times 1}{A^TAx} = \underset{n\times m\ m\times 1}{A^Tb} \quad \text{вектор } n\times 1$$

$$n < m \qquad \boxed{x^* = \underbrace{(A^TA)^{-1}A^T}_{A^\dagger}b} = A^\dagger\cdot b$$

$Ax = b$

$x = A^\dagger b$

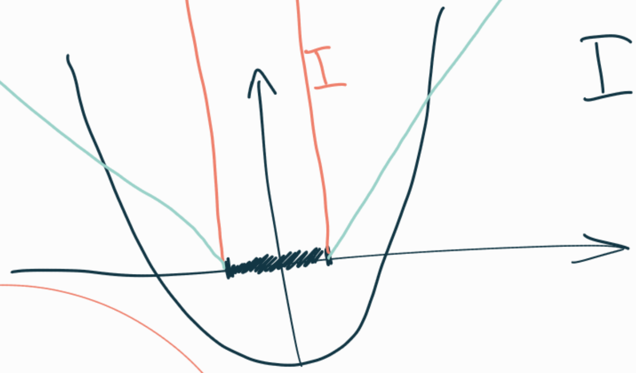
np.linalg.solve(A, b)

np.linalg.pinv(A)

если $\varkappa(A) = \dfrac{\lambda_{max}(A)}{\lambda_{min}(A)}$

$$\min_{x\in S} f(x) \Rightarrow \tilde{f}(x) = f(x) + I(x)$$

$$I(x) = \begin{cases} +\infty, & x\notin S \\ 0, & x\in S \end{cases}$$



в частности Лагранжиан

штрафует линейно

$$\frac{1}{2}\|x\|^2 \to \min$$
$$Ax-b=0$$

$$\|Ax-b\|_2^2 + \lambda\|x\|^2 \to \min_x$$

Ridge Regression

$$\|Ax-b\|_2^2 + \lambda\|x\|_1$$

LASSO Regression