

# 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:

$$\begin{aligned} f(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } h_i(x) &= 0, i = 1, \dots, p \end{aligned}$$

$$x \in S$$

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

$$L(x, \nu) = f(x) + \sum_{i=1}^m \nu_i h_i(x) \rightarrow \min_{x \in \mathbb{R}^n, \nu \in \mathbb{R}^p}$$

## General formulations and conditions

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

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

### Optimization on the general set $S$ .

Direction  $d \in \mathbb{R}^n$  is a feasible direction at  $x^* \in S \subseteq \mathbb{R}^n$  if small steps along  $d$  do not take us outside of  $S$ .

Consider a set  $S \subseteq \mathbb{R}^n$  and a function  $f: \mathbb{R}^n \rightarrow \mathbb{R}$ . Suppose that  $x^* \in S$  is a point of local minimum for  $f$  over  $S$ , and further assume that  $f$  is continuously differentiable around  $x^*$ .

1. Then for every feasible direction  $d \in \mathbb{R}^n$  at  $x^*$  it holds that  $\nabla f(x^*)^\top d \geq 0$

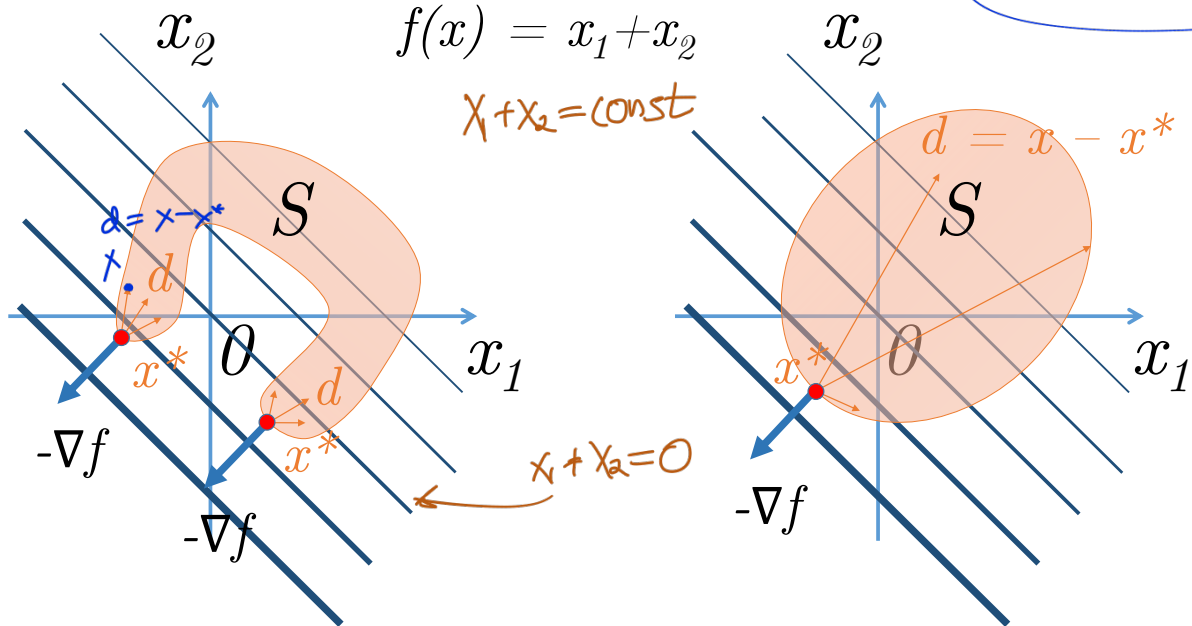
2. If, additionally,  $S$  is convex then

$$d = x - x^*$$

$S$  - non convex

$$\nabla f(x^*)^\top (x - x^*) \geq 0, \forall x \in S.$$

$S$  - convex



## Unconstrained optimization

### General case

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

$$f(x) \rightarrow \min_{x \in \mathbb{R}^n} \quad (\text{UP})$$

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

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

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

$$H_f(x^*) = \nabla^2 f(x^*) \succ (\prec) 0,$$

$$\forall y \neq 0 : y^\top \nabla^2 f(x^*) y > 0 \quad (\text{UP:Nec.})$$

$$y^\top \nabla^2 f(x^*) y > 0 \quad (\text{UP:Suff.})$$

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

Note, that if  $\nabla f(x^*) = 0$ ,  $\nabla^2 f(x^*) = 0$ , i.e. the hessian is positive *semidefinite*, we cannot be sure if  $x^*$  is a local minimum (see [Peano surface](#)  $f(x, y) = (2x^2 - y)(y - x^2)$ ).

### Convex case

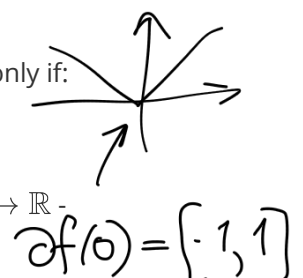
It should be mentioned, that in **convex** case (i.e.,  $f(x)$  is convex) necessary condition becomes sufficient. Moreover, we can generalize this result on the class of non-differentiable convex functions.

Let  $f(x) : \mathbb{R}^n \rightarrow \mathbb{R}$  - 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  $f(x) : S \rightarrow \mathbb{R}$  - 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  $f(x)$  - strictly or strongly (different cases 😊) convex function, then  $S^*$  contains only one single point  $S^* = x^*$ .

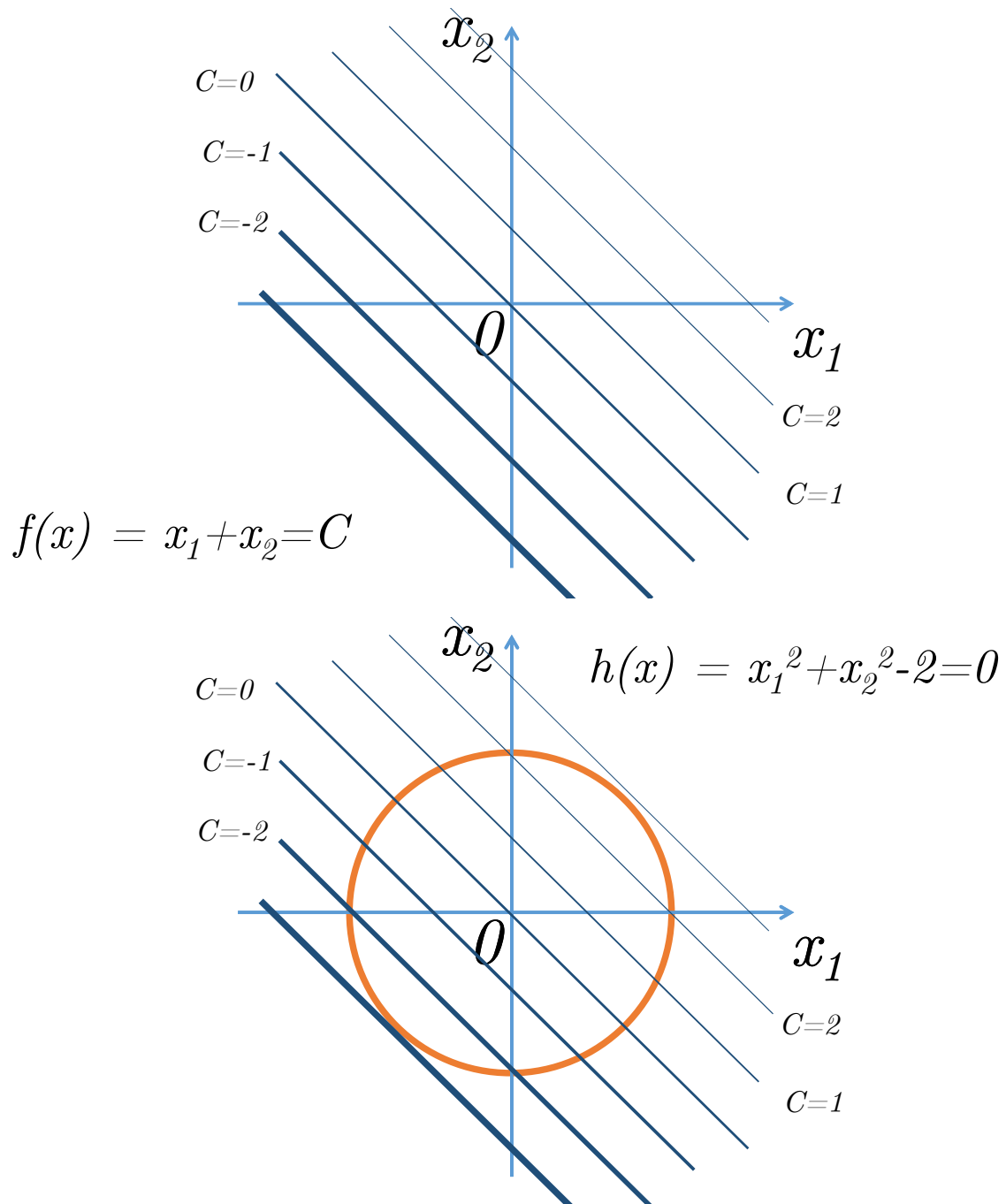
## Optimization with equality conditions

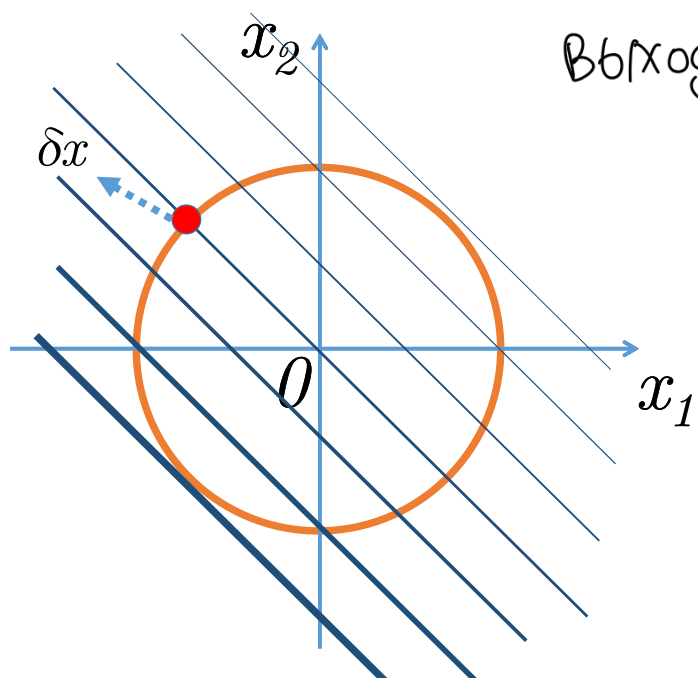
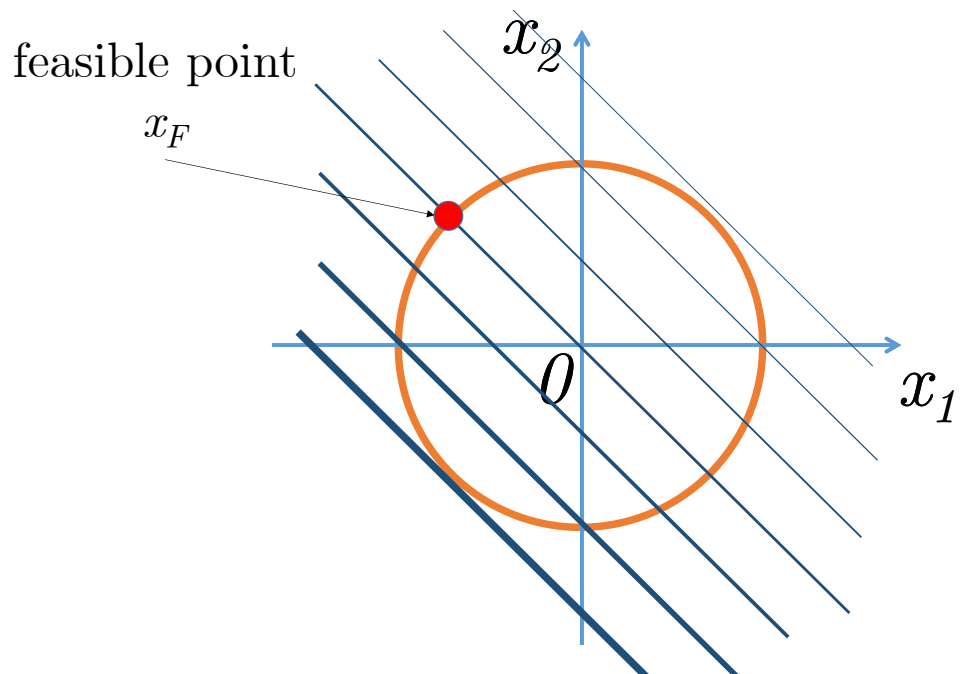
### Intuition

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

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

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



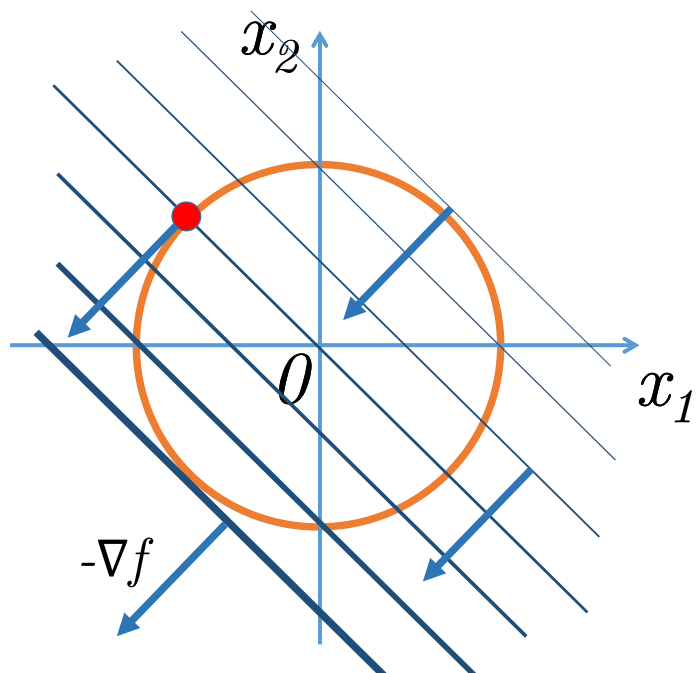


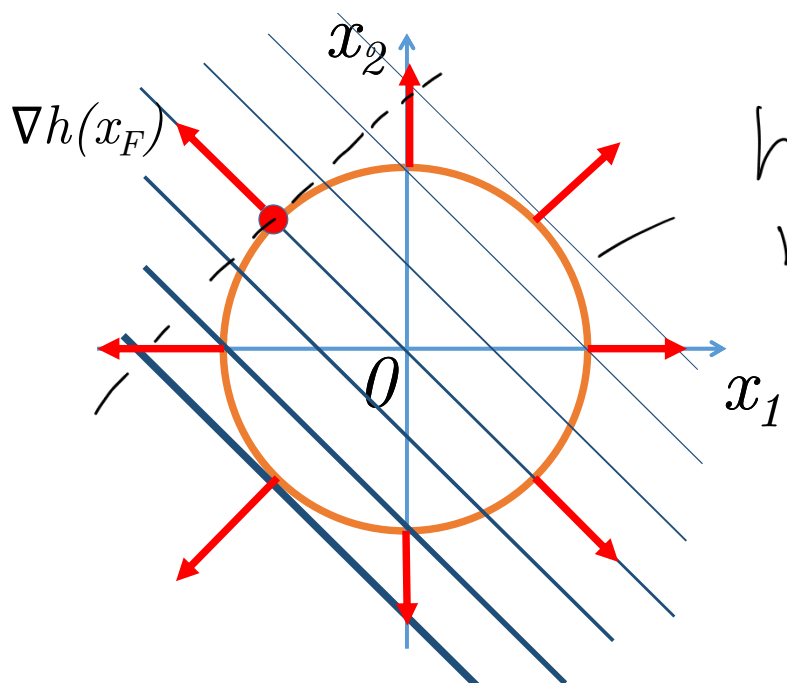
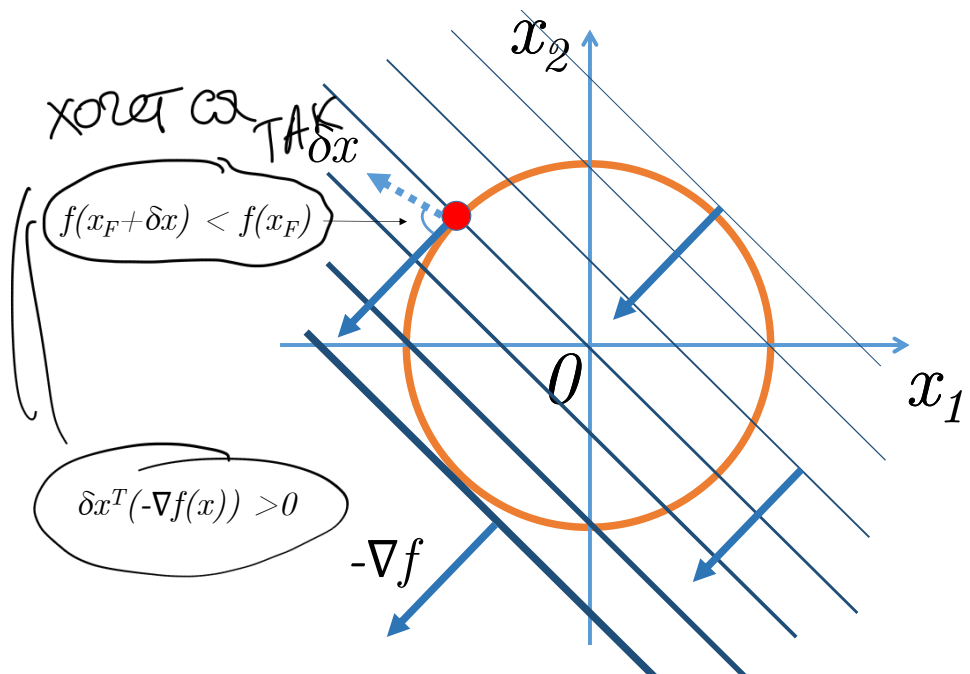
Выходим

из

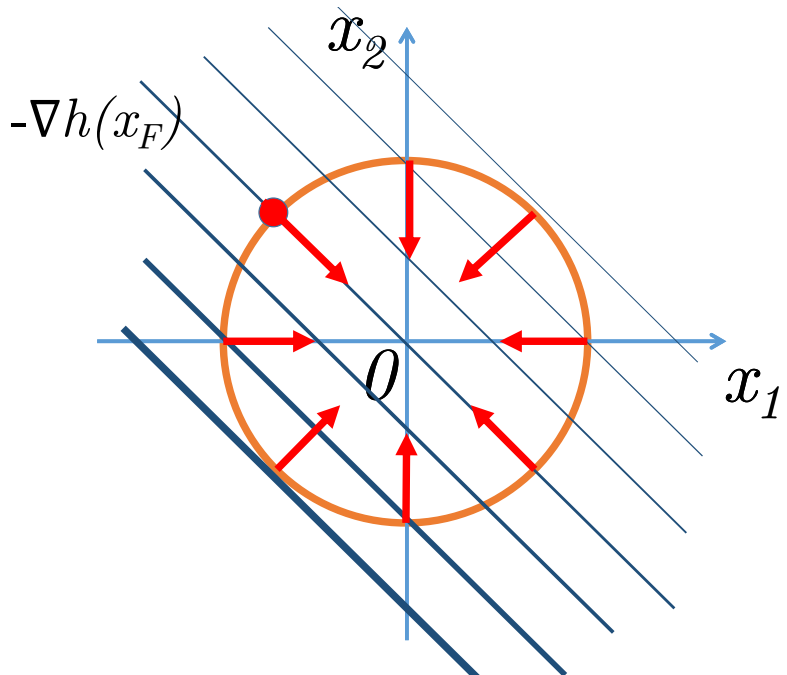
$S$

:(:(:(:(:(





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



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

$$\langle \delta x, \nabla h(x_F) \rangle = 0 \quad \text{БЮДЖЕТНОСТЬ}$$

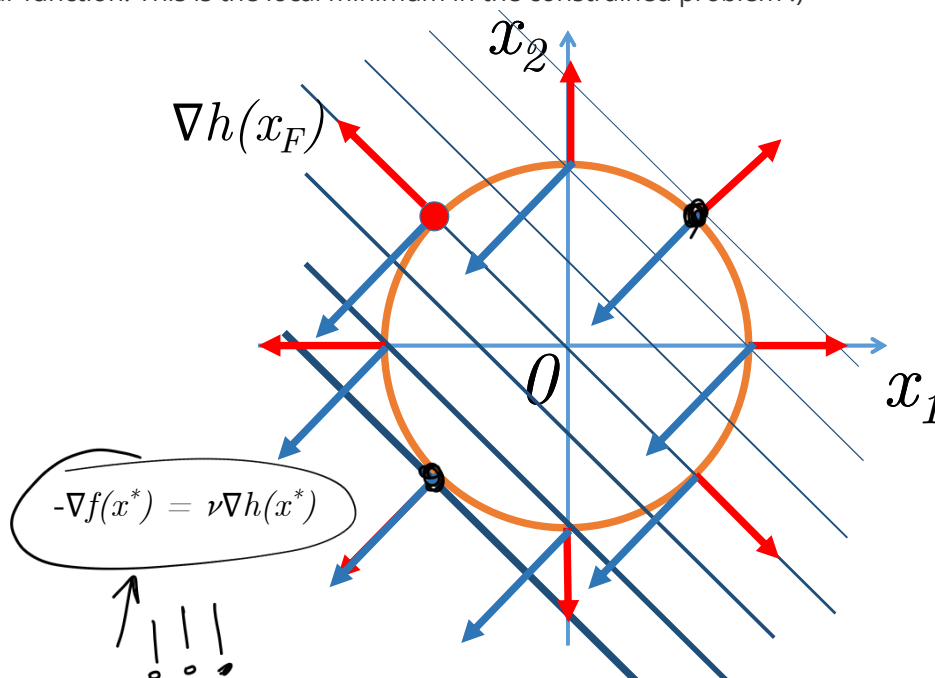
$$\langle \delta x, -\nabla f(x_F) \rangle > 0 \quad \text{УБЫВАНИЕ}$$

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

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

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

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 constrained problem :)



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

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

$$\nabla f(x) + \nu \nabla h(x) = 0$$

$$\nabla f = -\nu \nabla h$$

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

**Necessary conditions**

$$\begin{cases} \nabla_x L(x^*, \nu^*) = 0 \text{ that's written above} \\ \nabla_\nu L(x^*, \nu^*) = 0 \text{ budget constraint} \end{cases} \rightarrow h(x^*) = 0$$

**Sufficient conditions**

$$\begin{cases} \langle y, \nabla_{xx}^2 L(x^*, \nu^*) y \rangle \geq 0, \\ \forall y \neq 0 \in \mathbb{R}^n : \nabla h(x^*)^\top y = 0 \end{cases}$$

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

## General formulation

$$\begin{aligned} & f(x) \rightarrow \min_{x \in \mathbb{R}^n} \\ & \text{s.t. } h_i(x) = 0, \quad i = 1, \dots, p \end{aligned} \quad (\text{ECP})$$

Solution

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

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

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

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

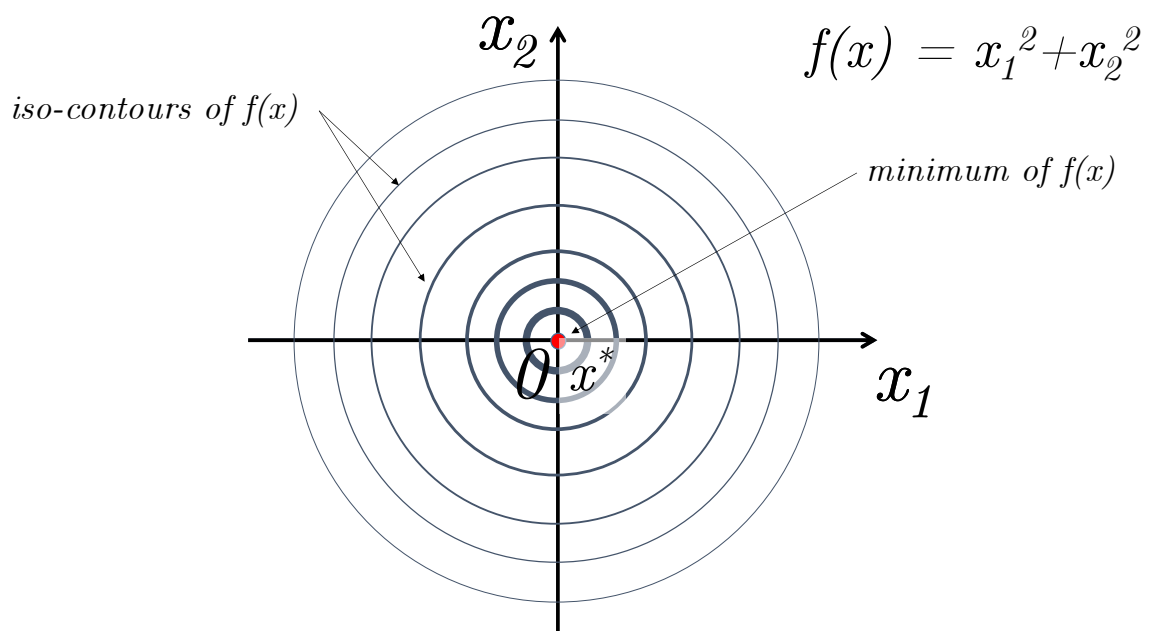
$y^\top H y$	$\lambda_i$	Definiteness H	Nature $x^*$	
$> 0$		Positive d.	Minimum	
$\geq 0$		Positive semi-d.	Valley	
$\neq 0$		Indefinite	Saddlepoint	
$\leq 0$		Negative semi-d.	Ridge	
$< 0$		Negative d.	Maximum	

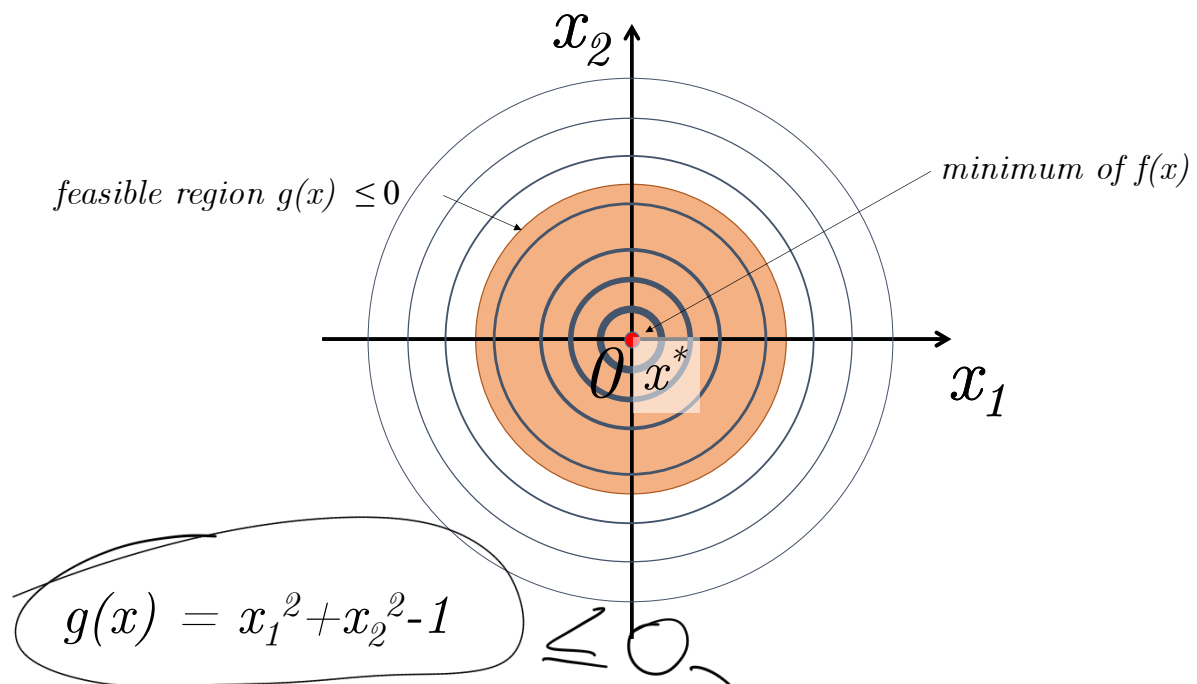
## Optimization with inequality conditions

### Example

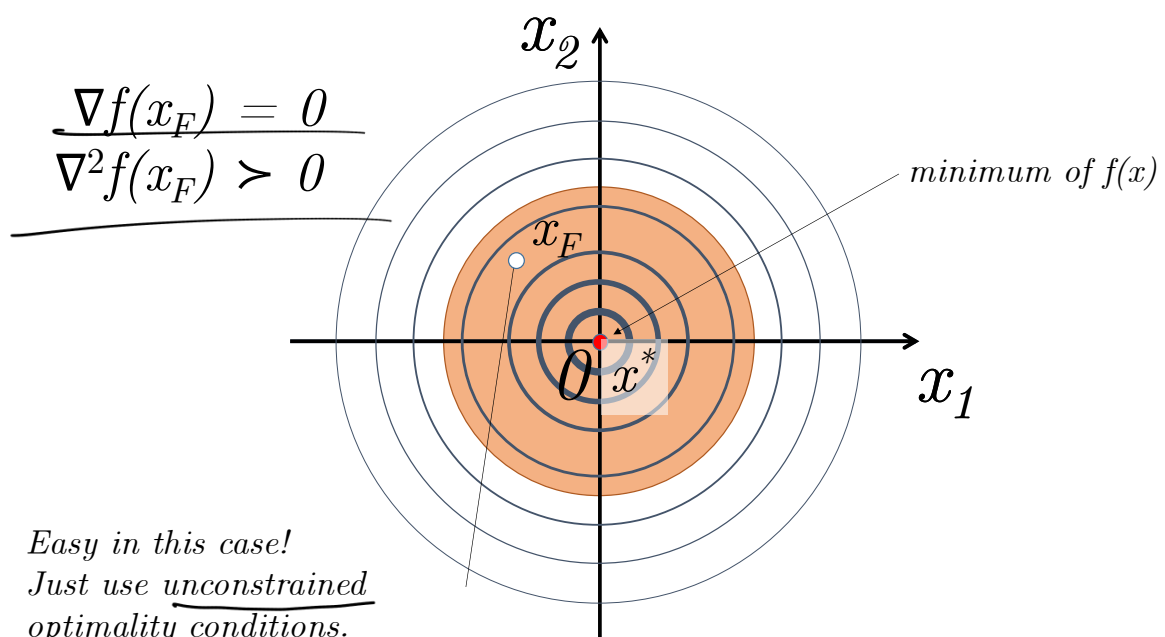
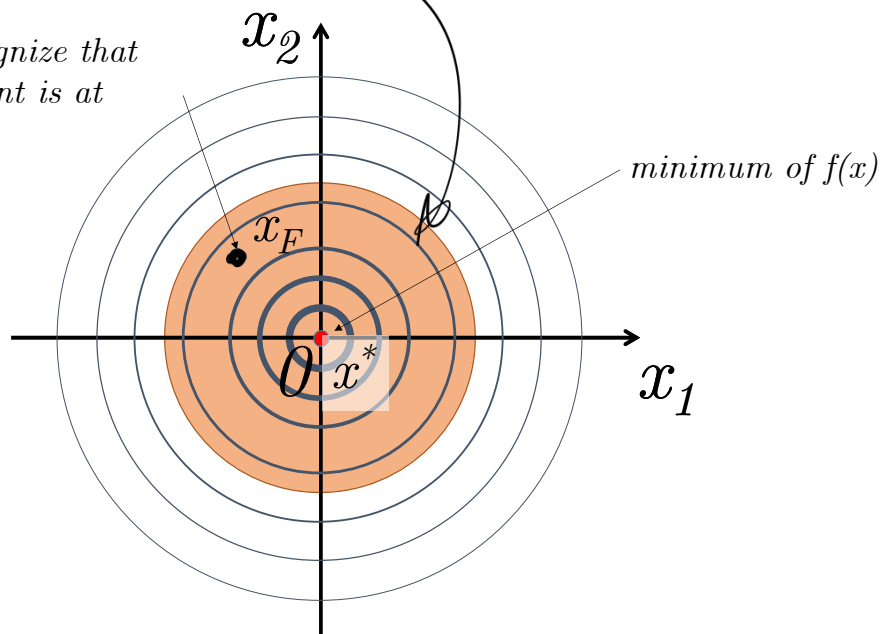
$$f(x) = x_1^2 + x_2^2 \quad g(x) = x_1^2 + x_2^2 - 1$$

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





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

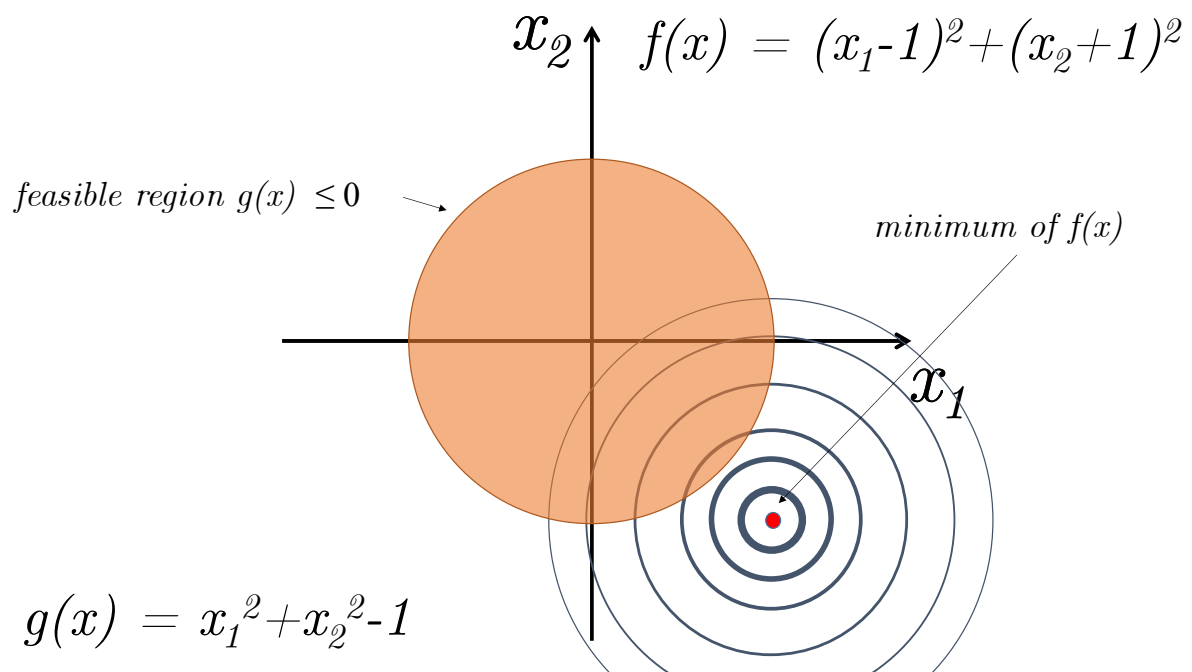
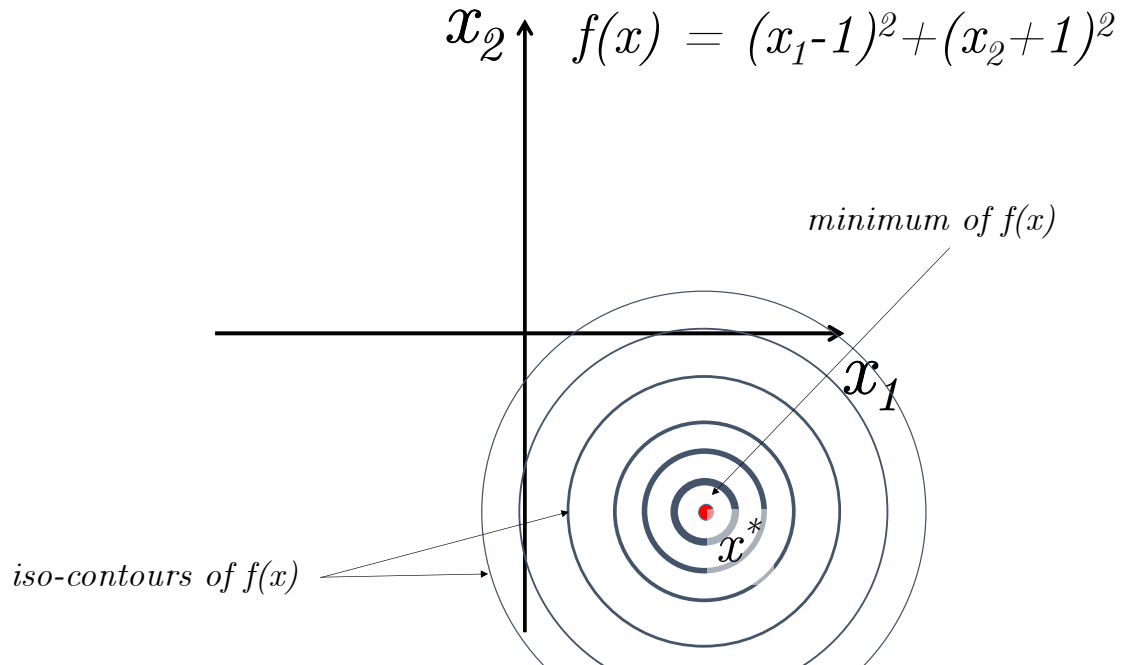




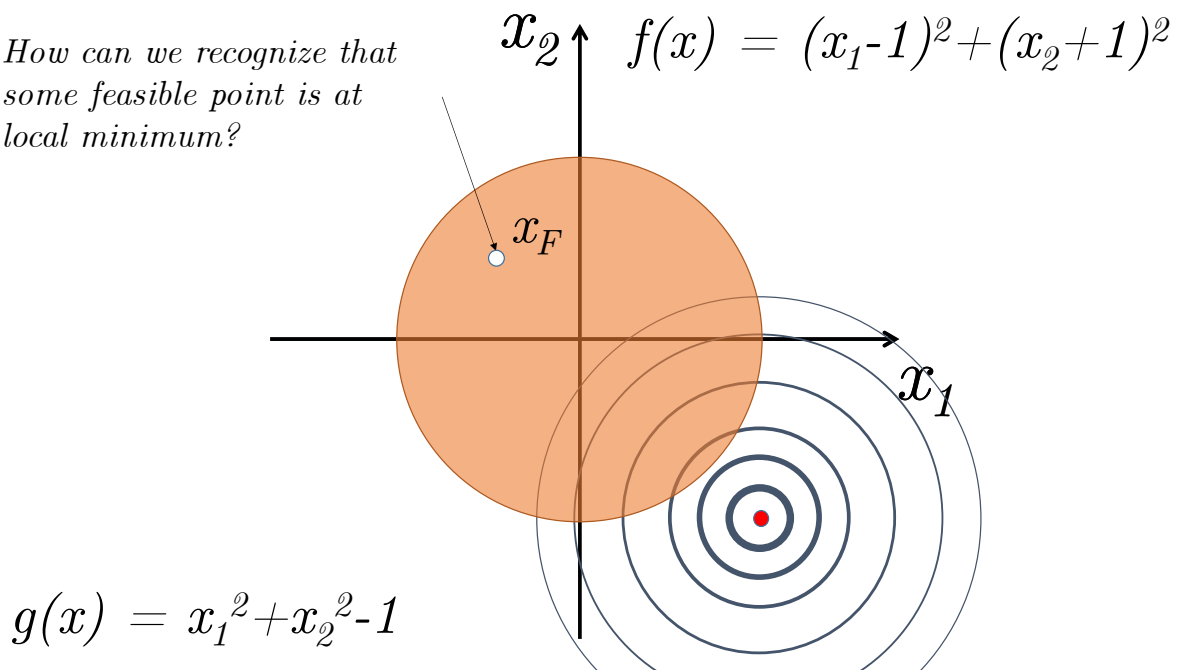
Thus, if the constraints of the type of inequalities are inactive in the constrained problem, then don't worry and write out the solution to the unconstrained problem. However, this is not the whole story ☹️. Consider the second childish example

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

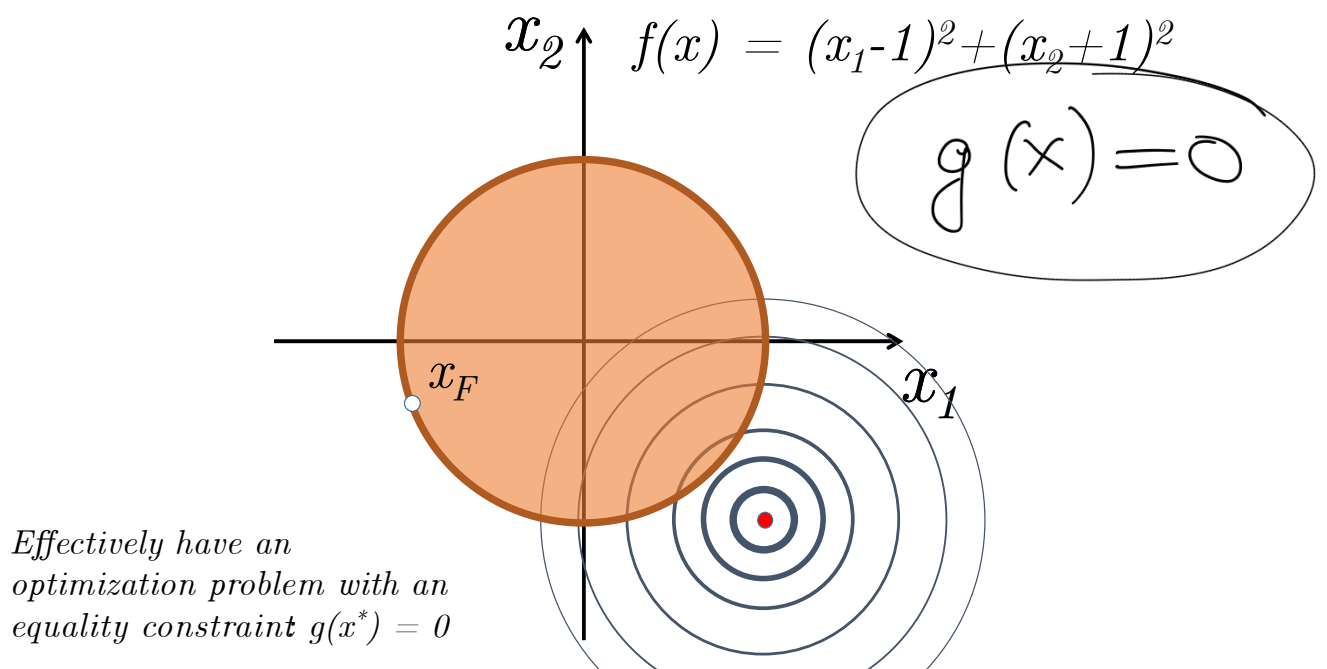
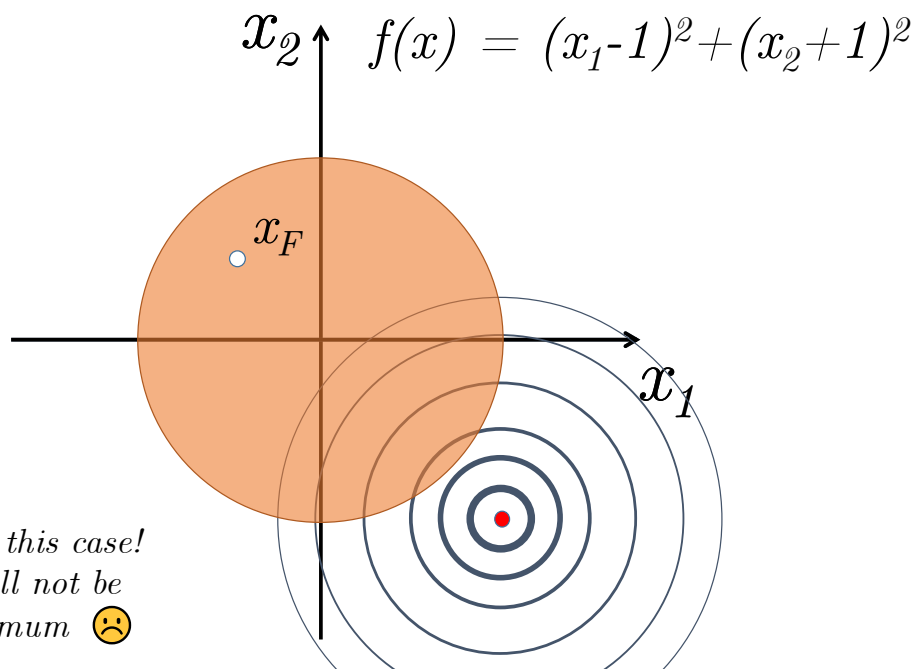
$$\begin{aligned} f(x) &\rightarrow \min_{x \in \mathbb{R}^n} \\ \text{s.t. } g(x) &\leq 0 \end{aligned}$$



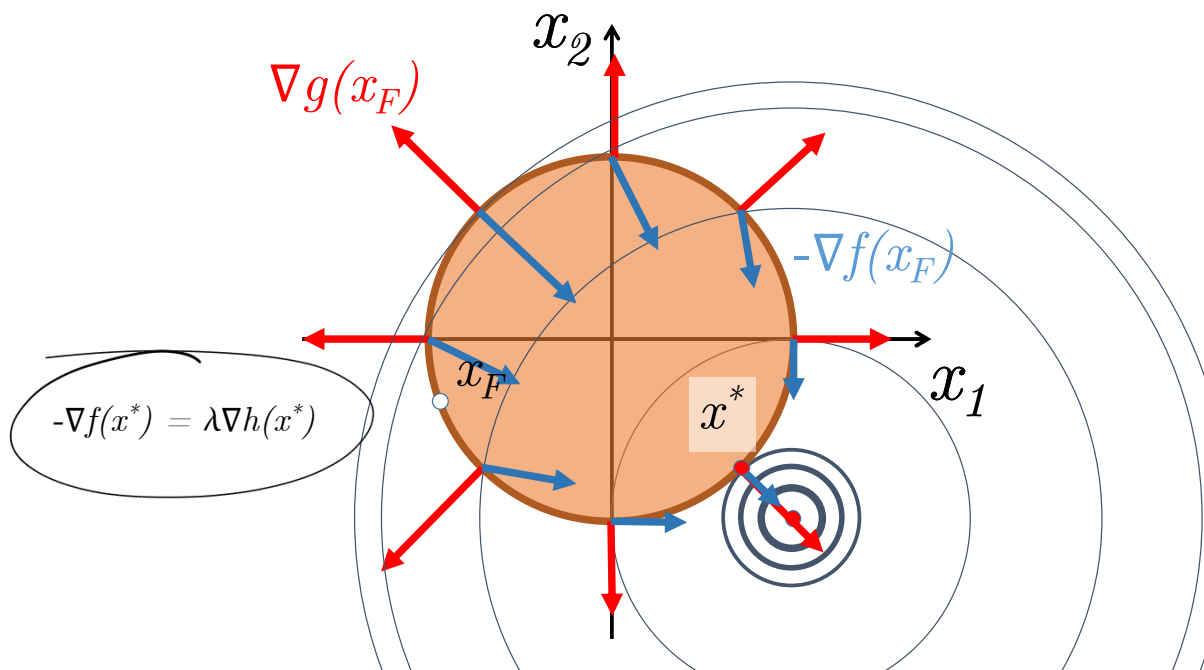
How can we recognize that some feasible point is at local minimum?



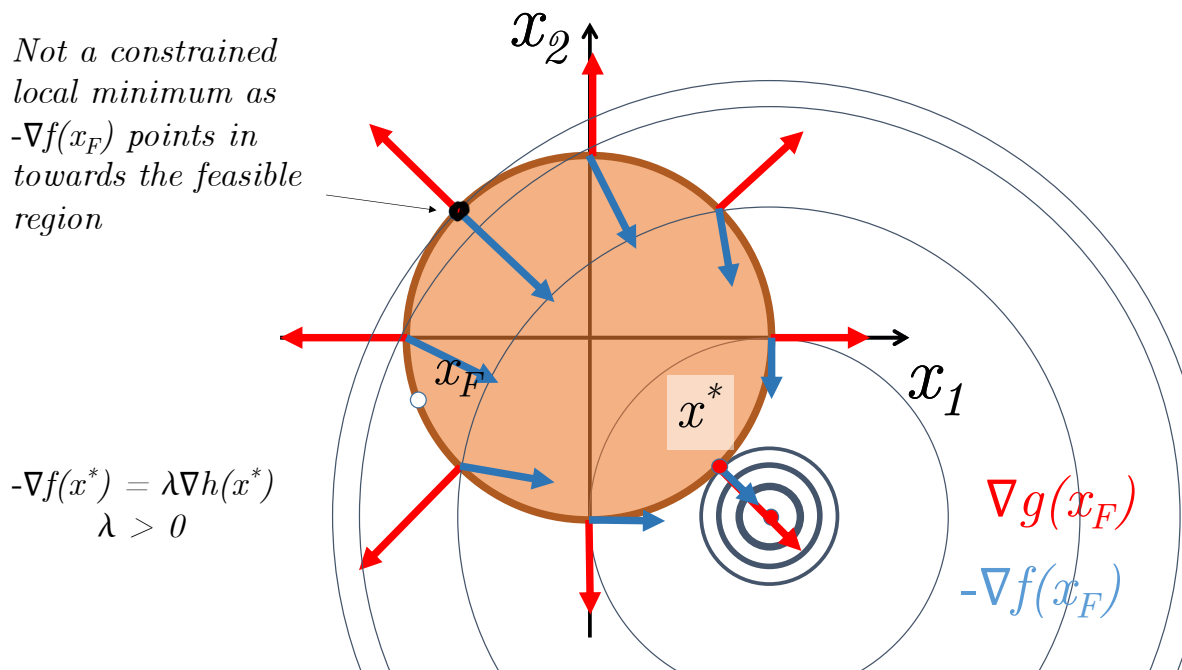
Not very easy in this case!  
Even gradient will not be zero at local optimum 😞



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



Not a constrained local minimum as  $-\nabla f(x_F)$  points in towards the feasible region



So, we have a problem:

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

s.t.  $g(x) \leq 0$

Two possible cases:

by UP

ECP

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

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

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

Let's define the Lagrange function:

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

НЕ АКТУАЛЬНО

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

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

$\lambda = 0, g(x^*) < 0$   
 $\lambda > 0, g(x^*) = 0$   
 АКТУАЛЬНО

гол. унк.

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.

## General formulation

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

This formulation is a general problem of mathematical programming.

The solution involves constructing a Lagrange function:

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

$$\begin{aligned} h_i(x) &= 0 \\ h_i(x) &\leq 0 \\ -h_i(x) &\leq 0 \end{aligned}$$

## Karush-Kuhn-Tucker conditions

### Necessary conditions

Let  $x^*, (\lambda^*, \nu^*)$  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.

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

### 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  $\langle y, \nabla_{xx}^2 L(x^*, \lambda^*, \nu^*) y \rangle \geq 0$  with *semi-definite* hessian of Lagrangian.



