



# Lecture 5

---

Numerical methods for solving optimization problems

# Gradient methods for solving optimization problems



---

General plan of all gradient methods:

- Select some initial point belonging to the domain of the function.
- Find the direction of the gradient at the selected point.
- Move from the initial point to the next one, taking a step of length  $\alpha$  in the direction of the gradient, if the problem of maximizing the objective function is solved, or in the direction of the antigradient, if the problem of minimizing the objective function is solved.
- The minimum (maximum) of the objective function will be achieved when the gradient becomes zero. When solving practical problems, the stopping condition is usually specified as a certain criterion, and each subtype of gradient methods assumes its own stopping criterion.

# Gradient Descent Method

Objective: to decrease the values of the objective function:

$$f(x^{k+1}) < f(x^k), \forall k \geq 0$$

General plan of descent methods:

1. Choose an initial approximation  $x^0$ .

2. Choose a vector  $g^k \neq 0: \forall \alpha > 0$

$$F_k(\alpha) = f(x^k + \alpha g^k) < f(x^k)$$

3. Determine the step length  $\alpha_k > 0 : F_k(\alpha_k) = f(x^k + \alpha_k g^k) < f(x^k)$

# Gradient Descent Method

4. As another approximation, take  $x^{k+1} = x^k + \alpha_k g^k$ .

5. Check the fulfillment of the iteration termination criterion. If the criterion is fulfilled, then set  $x^* \approx x^{k+1}$ . Otherwise, return to point 2.

Thus, the essence of the gradient method is to construct a sequence  $\{x_k\}_{k=0}^{\infty}$  according to the rule  $x^{k+1} = x^k - \alpha_k f'(x^k)$ ,  $\alpha_k > 0$ ,  $k = 0, 1, \dots, N$ .

Let us define on the ray directed along the antigradient the function  $F_k(\alpha) = f(x^k - \alpha_k f'(x^k))$ ,  $\alpha_k$  we find from the problem of one-dimensional minimization  $F_k(\alpha) \rightarrow \min_{\alpha \geq 0}$ .

# Gradient Descent Method

Let us consider the application of the steepest gradient descent method to quadratic functions

1. Calculate the gradient descent direction  $g^k = -(Ax^k + b)$ .  $f(x) = \frac{1}{2}x^T Ax + b^T x + c$

2. Determine the step  $\alpha_k$  from the condition  $F_k(\alpha) \rightarrow \min_{\alpha \geq 0}$ :

$$F_k(\alpha) = f(x^k + \alpha g^k) = f(x^k) - \alpha \|g^k\|^2 + \frac{1}{2} \alpha^2 (Ag^k, g^k)$$

$$F'_k(\alpha) = -\|g^k\|^2 + \alpha (Ag^k, g^k) = 0 \quad \alpha_k = \frac{\|g^k\|^2}{(Ag^k, g^k)} > 0$$

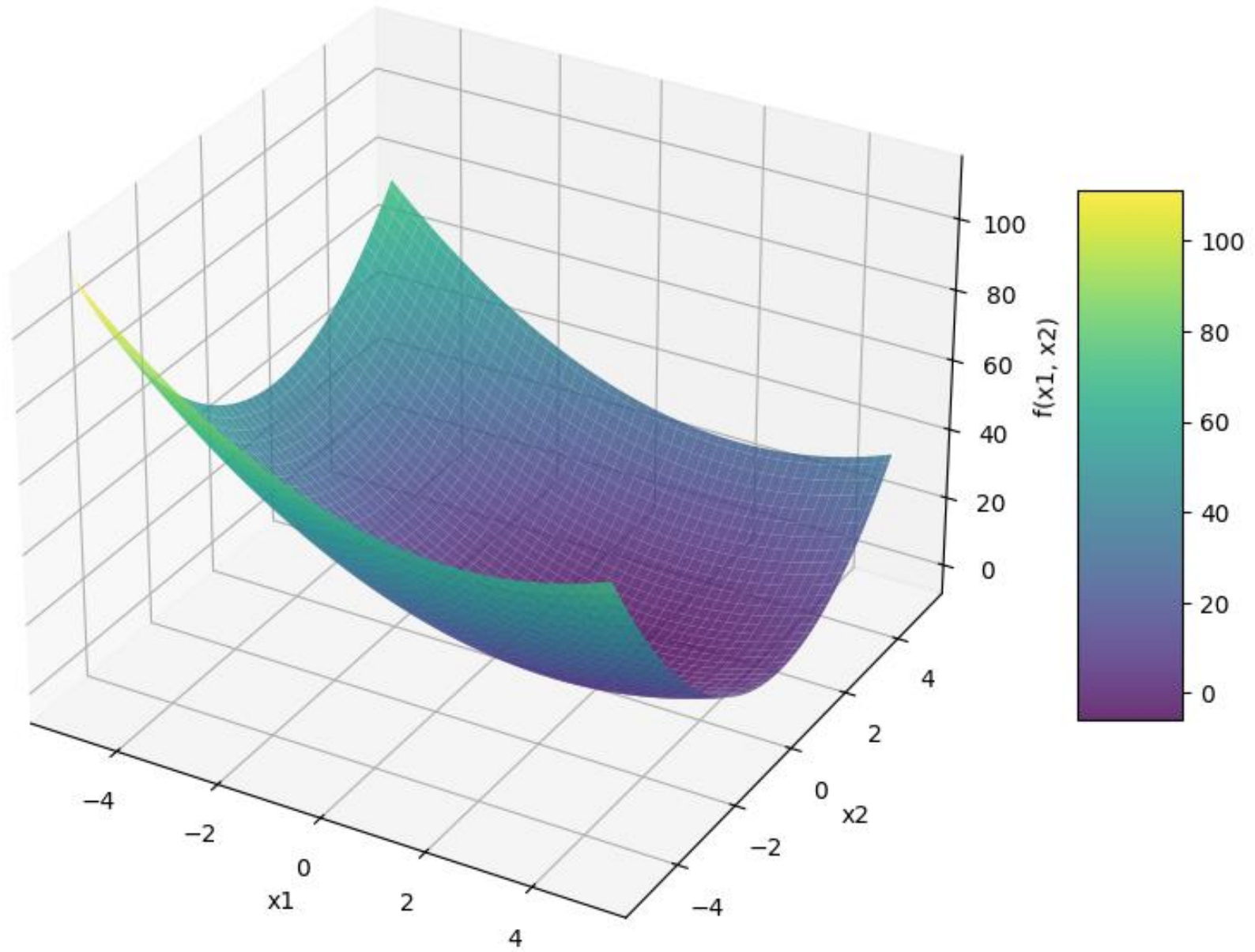
3. We proceed to the next iteration  $x^{k+1} = x^k + \alpha_k g^k$ . We check the fulfillment of the iteration termination criterion. If the criterion is fulfilled, then we assume  $x^* \approx x^{k+1}$ , otherwise we proceed to point 1. As a result, we obtained a sequence  $\{x^k\}$  of decreasing function  $f(x)$ .

# Gradient Descent Method

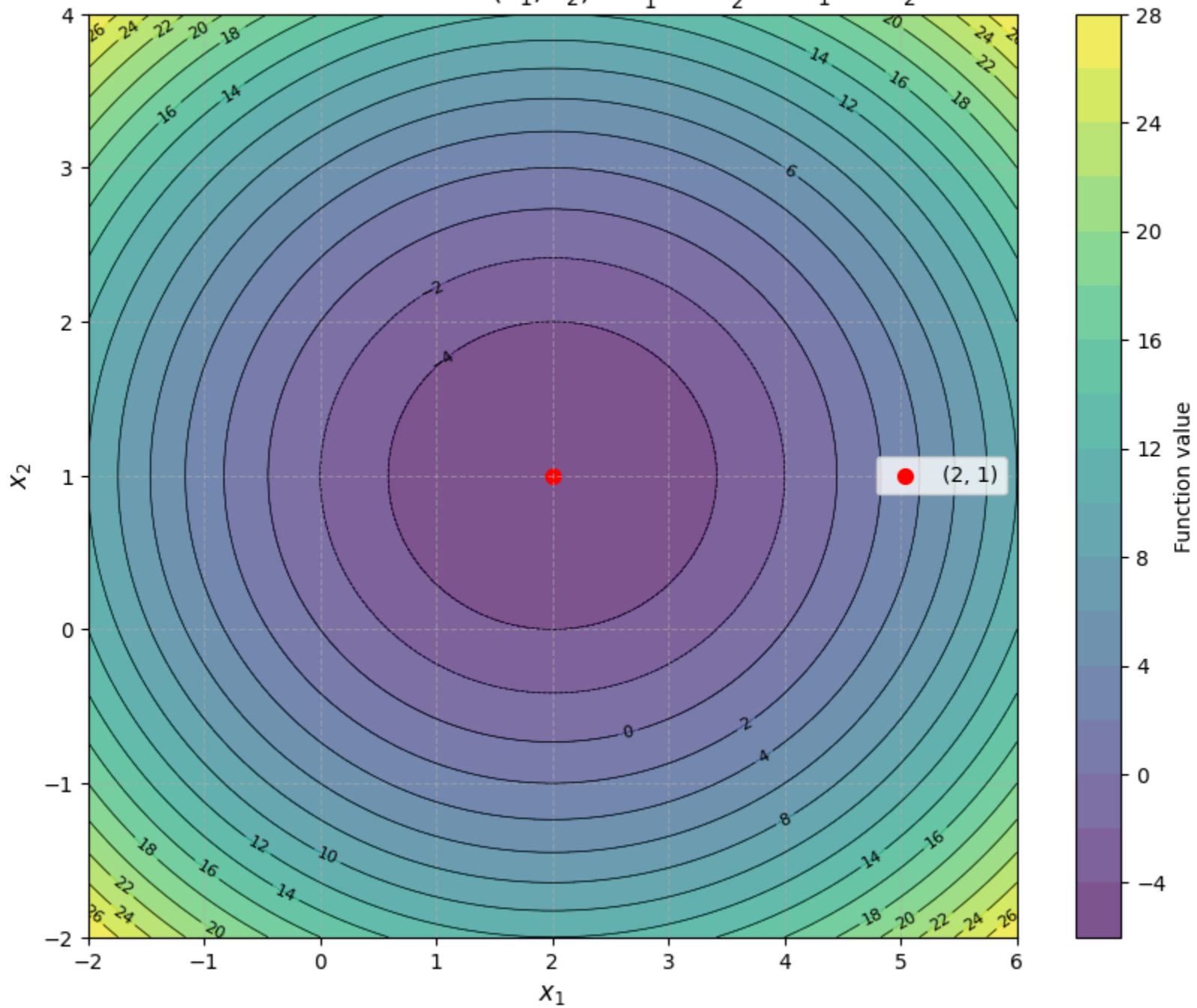


---

**Example.** Find the minimal value of the function  $f(x_1, x_2) = x_1^2 + 2x_2^2 - 4x_1 - 4x_2$ .



Function level lines  $f(x_1, x_2) = x_1^2 + 2x_2^2 - 4x_1 - 4x_2$



# Gradient Descent Method

Let's choose the initial approximation. Let it be the point  $x^0(0; 0)$ .

1. Gradient descent direction  $g^0 = -(Ax^0 + b) = (4; 4)^T$ .

2. Step  $\alpha_0 = \frac{\|g^0\|^2}{(Ag^0, g^0)} = \frac{32}{96} = \frac{1}{3}$

3. Move on to the next iteration  $x^1 = x^0 + \alpha_0 g^0 = (0; 0) + \frac{1}{3}(4; 4)^T = \left(\frac{4}{3}; \frac{4}{3}\right)$ . We check the fulfillment of the iteration termination criterion:  $\|Ax^1 + b\| = \left\| \left(-\frac{4}{3}; \frac{4}{3}\right) \right\| = \frac{\sqrt{32}}{3}$ . The iteration termination criterion is not met, this norm does not tend to zero, which means we continue the solution and move on to point 1 of the algorithm.

# Gradient Descent Method

4.  $g^1 = -(Ax^1 + b) = \left(\frac{4}{3}; -\frac{4}{3}\right)^T$

5.  $\alpha_1 = \frac{\|g^1\|^2}{(Ag^1, g^1)} = \frac{1}{3}$

6.  $x^2 = x^1 + \alpha_1 g^1 = \left(\frac{4}{3}; \frac{4}{3}\right) - \left(\frac{4}{3}; -\frac{4}{3}\right)^T = \left(\frac{16}{9}; \frac{8}{9}\right)^T$ . We check the fulfillment of the iteration termination criterion:  $\|Ax^2 + b\| = \frac{\sqrt{32}}{3^2}$ . . The iteration termination criterion is not fulfilled, this norm does not tend to zero, so we continue the solution and move on to point 1 of the algorithm.

7.  $g^2 = -(Ax^2 + b) = \left(\frac{4}{9}; \frac{4}{9}\right)^T$

8.  $\alpha_2 = \frac{\|g^2\|^2}{(Ag^2, g^2)} = \frac{1}{3}$

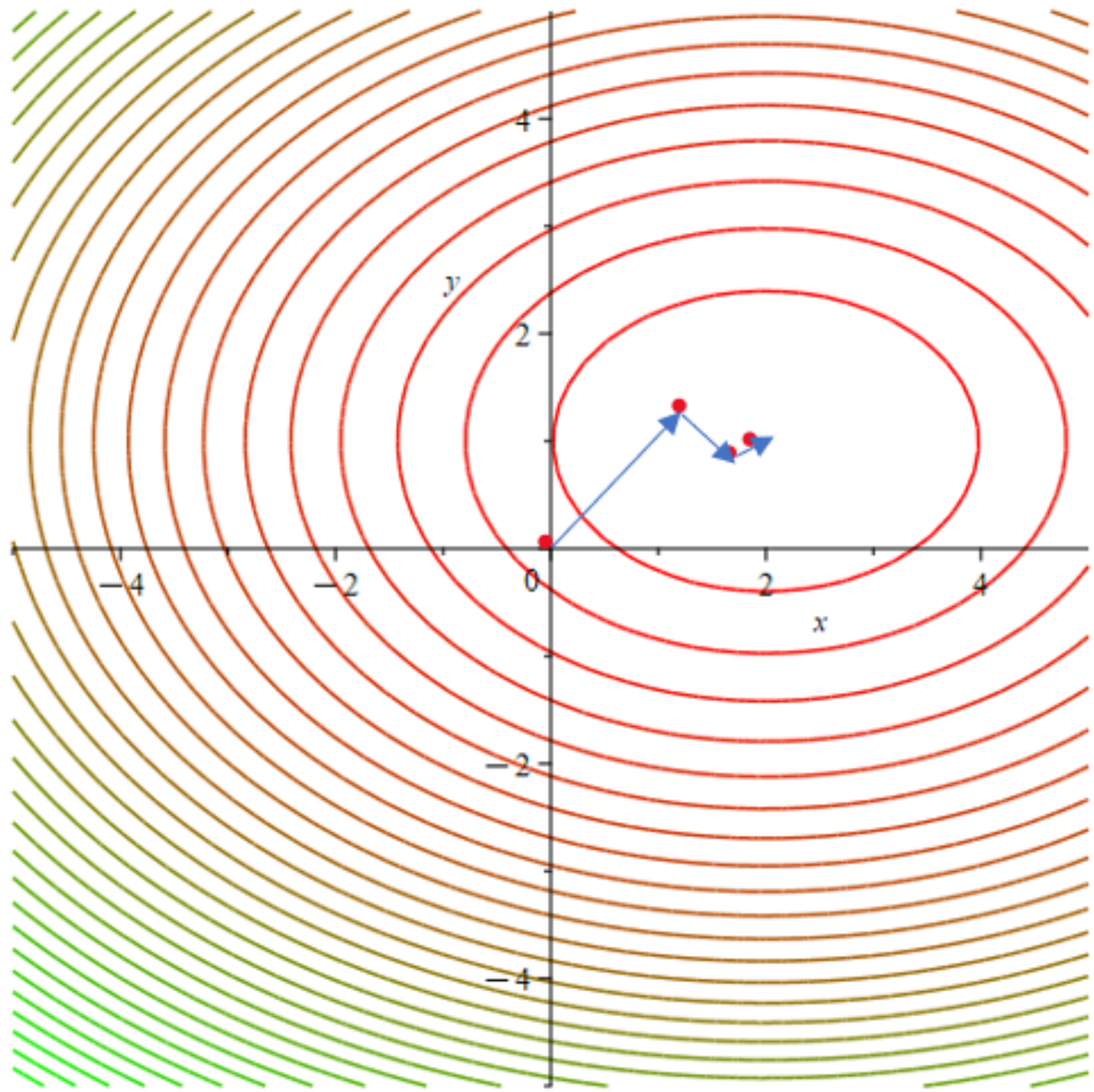
# Gradient Descent Method

9.  $x^3 = x^2 + \alpha_2 g^2 = \left(\frac{52}{27}; \frac{28}{27}\right)^T$ . We check the fulfillment of the iteration termination criterion  $\|Ax^3 + b\| = \frac{\sqrt{32}}{3^3}$ . The iteration termination criterion is not met, this norm does not tend to zero, which means we continue the solution and move on to point 1 of the algorithm.

We continue solving until  $\|Ax^k + b\|$  is close to zero. Note that  $\|Ax^k + b\| = \frac{\sqrt{32}}{3^k} \rightarrow 0, k \rightarrow \infty$ .

During the three steps of this method, the following descent trajectory was obtained:

$$\left\{ (0; 0); \left(\frac{4}{3}; \frac{4}{3}\right); \left(\frac{16}{9}; \frac{8}{9}\right); \left(\frac{52}{27}; \frac{28}{27}\right); \dots \right\}.$$



# Coordinate descent method

**Numerical methods for solving optimization problems that allow not to use a gradient.**

Let us have some approximation  $x^k$  and it is necessary to solve a problem of the form  $f(x) \rightarrow \min$ . The cycle with number  $k+1$  has the form:

1. Descent along coordinate  $x_1$ . To do this, we fix all other coordinates  $x_2 = x_2^k, x_3 = x_3^k, \dots, x_n = x_n^k$  and solve the problem of minimizing a function of one variable  $f(x_1, x_2^k, x_3^k, \dots, x_n^k) \rightarrow \min_{x_1}$ . For example,  $x_1^{k+1}$  can be found as  $x_1^{k+1} = x_1^k - \alpha_1^{k+1} \frac{\partial f}{\partial x_1}$ , where  $\alpha_1^{k+1}$  is a certain step. That is, at this step there is a shift from point  $x_1^k$  to point  $x_1^{k+1}$  towards decreasing function value.

# Coordinate descent method

2. Descent along coordinate  $x_2$ . To do this, we fix all other coordinates  $x_1 = x_1^{k+1}$ ,  $x_3 = x_3^k, \dots, x_n = x_n^k$  and solve the problem of minimizing a function of one variable  $f(x_1^{k+1}, x_2, x_3^k, \dots, x_n^k) \rightarrow \min_{x_2}$ . For example,  $x_2^{k+1}$  can be found as  $x_2^{k+1} = x_2^k - \alpha_2^{k+1} \frac{\partial f}{\partial x_2}$ , where  $\alpha_2^{k+1}$  is a certain step. That is, at this step there is a shift from point  $x_2^k$  to point  $x_2^{k+1}$  towards decreasing function value.

# Coordinate descent method

3. We descend in the same way along the remaining coordinates

4. At the  $n$ -th step, we find  $x_n^{k+1}$  by solving the problem  $f(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \dots, x_n^k) \rightarrow \min_{x_n}$ .

As a result, we obtain a sequence of points  $(x_1^{k+1}, x_2^k, x_3^k, \dots, x_n^k), (x_1^{k+1}, x_2^{k+1}, x_3^k, \dots, x_n^k), \dots, (x_1^{k+1}, x_2^{k+1}, x_3^k, \dots, x_n^{k+1})$ .

5. Check the loop termination condition:

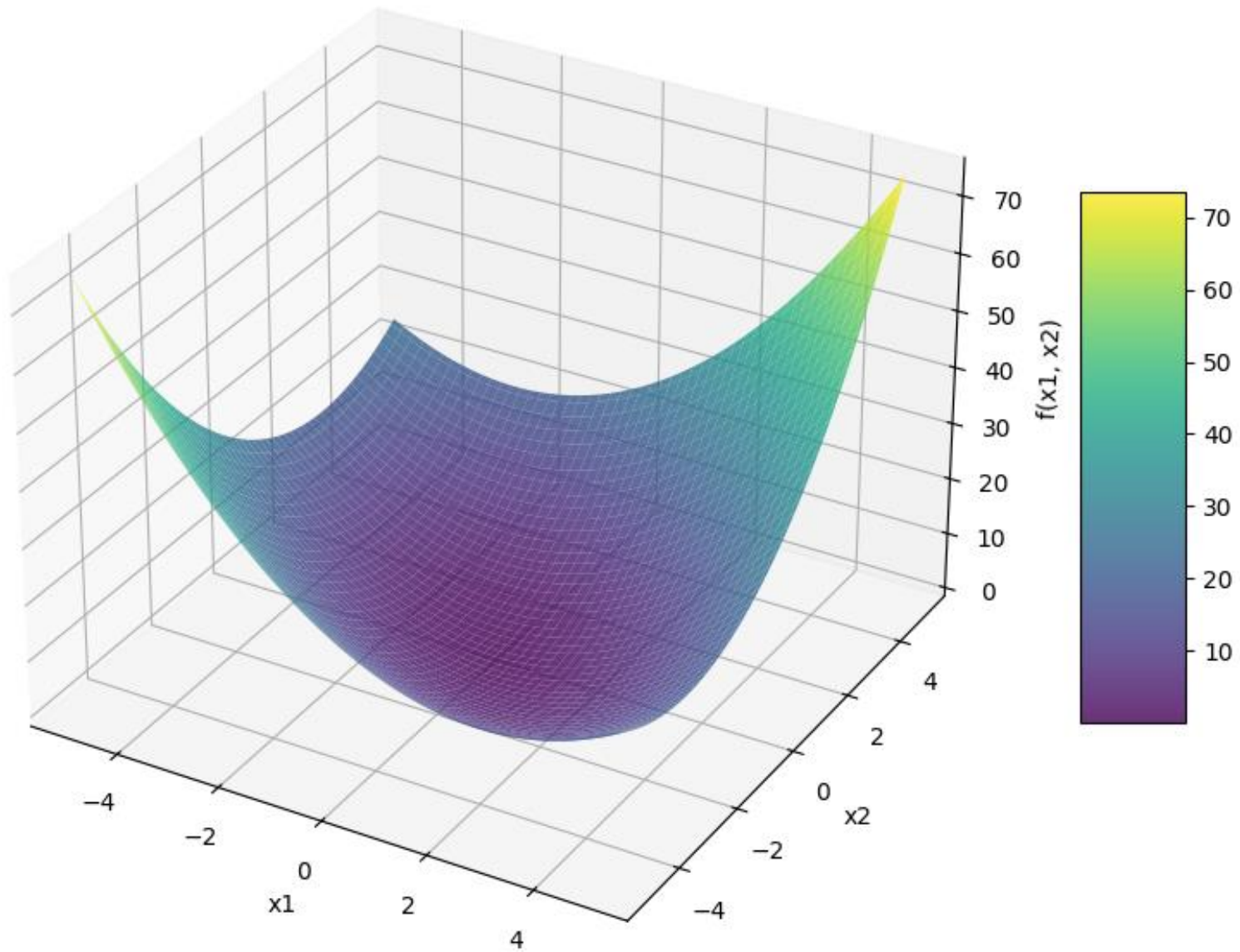
$$|f(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \dots, x_n^{k+1}) - f(x_1^k, x_2^k, x_3^k, \dots, x_n^k)| < \varepsilon$$

# Coordinate descent method

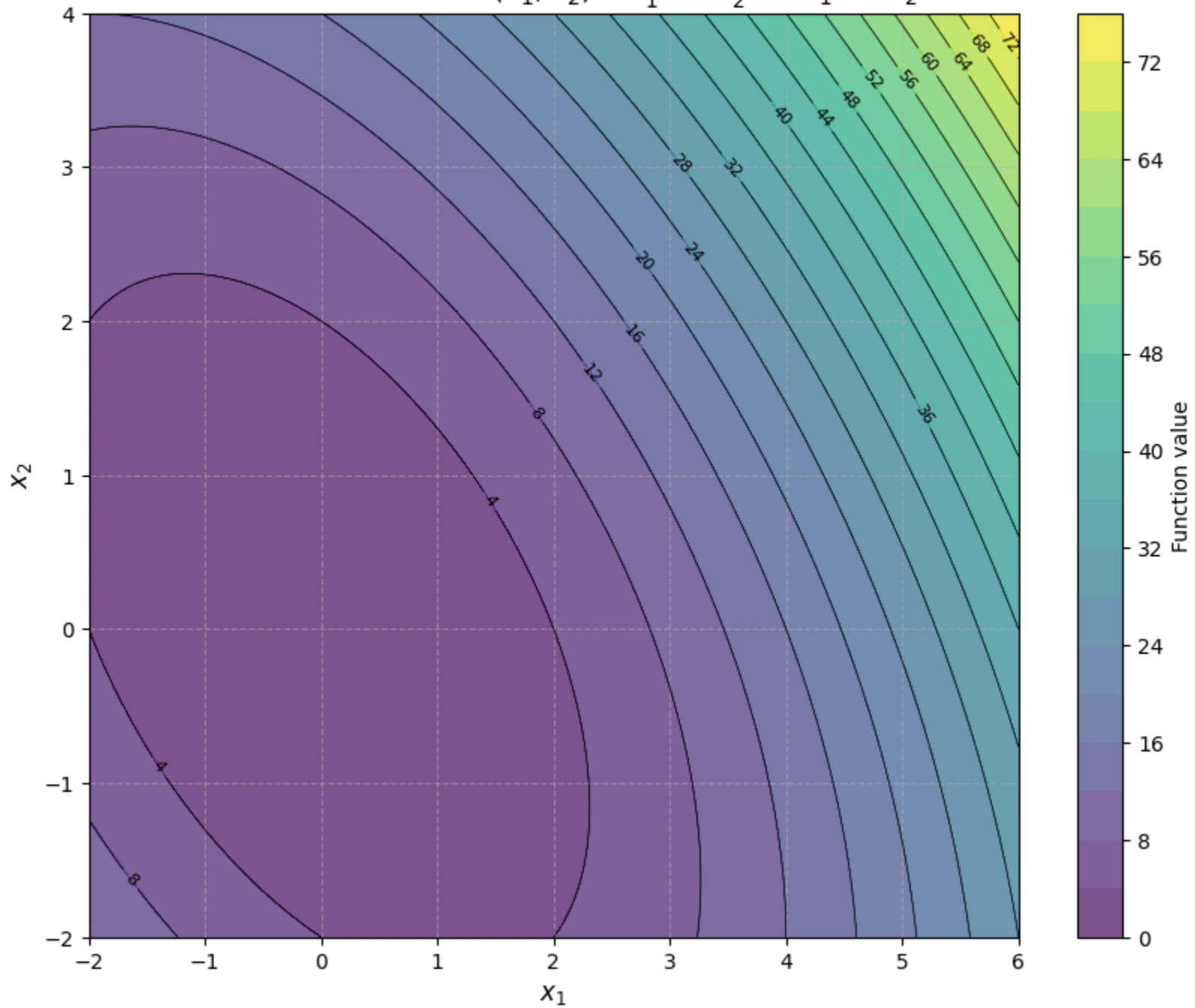


---

**Example.** Find the minimal value of the function  $f(x_1, x_2) = x_1^2 + x_1x_2 + x_2^2$ .



Function level lines  $f(x_1, x_2) = x_1^2 + 2x_2^2 - 4x_1 - 4x_2$



# Coordinate descent method

1. Choose an initial approximation, for example  $x^0(1; 1)$ . Then  $f(x^0) = 3$ . We fix the coordinate  $x_2$  ( $x_2 = 1$ ) and proceed to solving the following optimization problem:

$$f(x_1) = \frac{3}{4}(x_1 + 1)^2 + \frac{1}{4}(x_1 - 1)^2 \rightarrow \min$$

$$f(x_1) = \frac{3}{4}(x_1^2 + 2x_1 + 1) + \frac{1}{4}(x_1^2 - 2x_1 + 1) = x_1^2 + x_1 + 1 \rightarrow \min$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + 1 = 0 \Rightarrow x_1 = -\frac{1}{2}$$

We got that  $x_1 = -\frac{1}{2}$  - the minimum point.  $f(x_1) = \frac{3}{4}$ . Checking the loop termination condition.  $\varepsilon = 0.05$ .  $|f(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \dots, x_n^{k+1}) - f(x_1^k, x_2^k, x_3^k, \dots, x_n^k)| = \left| \frac{3}{4} - 3 \right| > \varepsilon$ .

# Coordinate descent method

2. Let us fix the coordinate  $x_1 = -\frac{1}{2}$  and proceed to solving the following optimization problem:

$$f(x_2) = \frac{3}{4} \left(x_2 - \frac{1}{2}\right)^2 + \frac{1}{4} \left(x_2 + \frac{1}{2}\right)^2 \rightarrow \min$$

$$\frac{\partial f}{\partial x_2} = 2x_2 - \frac{1}{2} = 0 \Rightarrow x_2 = \frac{1}{4}$$

We got that  $x_2 = \frac{1}{4}$  - the minimum point.  $f(x_2) = \frac{3}{16}$ . We check the condition for ending the cycle.  $\left|\frac{3}{16} - \frac{3}{4}\right| > \varepsilon$ . We proceed to the next step.

# Coordinate descent method

3. Let us fix the coordinate  $x_2 = \frac{1}{4}$  and proceed to solving the following optimization problem:

$$f(x_1) = \frac{3}{4} \left( x_1 + \frac{1}{4} \right)^2 + \frac{1}{4} \left( x_1 - \frac{1}{4} \right)^2 \rightarrow \min$$

$$\frac{\partial f}{\partial x_1} = 2x_1 + \frac{1}{4} = 0 \Rightarrow x_1 = -\frac{1}{8}$$

We got that  $x_1 = -\frac{1}{8}$  - the minimum point.  $f(x_1) = \frac{3}{64}$ . We check the condition for ending the cycle.  $\left| \frac{3}{64} - \frac{3}{16} \right| > \varepsilon$ . We proceed to the next step.

# Coordinate descent method

4. Let's fix the coordinate  $x_1 = -\frac{1}{8}$  and move on to solving the following optimization problem:

$$f(x_2) = \frac{3}{4} \left(x_2 - \frac{1}{8}\right)^2 + \frac{1}{4} \left(x_2 + \frac{1}{8}\right)^2 \rightarrow \min$$

$$\frac{\partial f}{\partial x_2} = 2x_2 - \frac{1}{8} = 0 \Rightarrow x_2 = \frac{1}{16}$$

We got that  $x_2 = \frac{1}{16}$  - the minimum point.  $f(x_2) = \frac{3}{256}$ . We check the cycle termination condition.  $\left| \frac{3}{256} - \frac{3}{64} \right| < \varepsilon$ . The cycle termination condition is met. This means that the optimization problem is solved and the optimal solution  $x^* \left(-\frac{1}{8}; \frac{1}{16}\right)$ ,  $f(x^*) = \frac{3}{256}$  is obtained.

# Coordinate descent method

In the course of solving the problem using the coordinate descent method, the following descent trajectory was obtained:  $\left\{ (1; 1); \left(-\frac{1}{2}; 1\right); \left(-\frac{1}{2}; \frac{1}{4}\right); \left(-\frac{1}{8}; \frac{1}{4}\right); \left(-\frac{1}{8}; \frac{1}{16}\right) \right\}$ .

