Lecture 4

Numerical methods for solving optimization problems

Unconditional nonlinear optimization of a function of one variable

Let the function $f(x)$ be a numerical function of a real variable x. Then the problem of one-dimensional optimization of the function $f(x)$ on the interval $[a,b]$ has the form:

$$f(x) \to \min_{x \in [a,b]}$$

The one-dimensional optimization problem is solved by localizing the minimum points and further searching for the optimal solution on the resulting segment. Let's consider various methods for solving the problem formulated above.

Brute force method

Brute force method is a simple and direct method to solve problems, which is often based on the description of the problem and the design concept definition.

Let a one-extremal function $f(x)$ be given on the interval $[a, b]$. It is necessary to find any of its minimum points on a given segment with a given accuracy $\varepsilon > 0$.

Let us divide the segment $[a, b]$ into n equal parts so that:

$$x_j = a + jh \ (j = 0,1, \ldots, n)$$

where $h = \frac{b-a}{n} -$ step and $n -$ integer, $n \geq \frac{b-a}{\varepsilon}$. That is, the length of each subinterval does not exceed ε.

Brute force method

Among all points $x_j, (j = 0,1, \dots, n)$ select a point x_m:

$$f(x_m) = \min_{0 \le j \le n} f(x_j)$$

We get the solution $x^* = x_m, f(x^*) = f(x_m)$.

Brute force method

Brute force algorithm:

1. We introduce the boundaries of the segment a, b and the accuracy ε.
2. Calculate $n = \left[\frac{b-a}{\varepsilon}\right] + 1;\ h = \frac{b-a}{n}.\ x_m = a$ and $min = f(a)$.
3. For each value $i, i = 1, \dots, n$:
    1. Calculate $x = a + i \cdot h$
    2. Calculate $y = f(x)$
    3. Compare y and $\min$. If $y \leq \min$, then $\min = y$, $x_m = x$ and move on to the next iteration. If this condition is not met, then we set $i = n$ and finish the calculations.

Brute force method

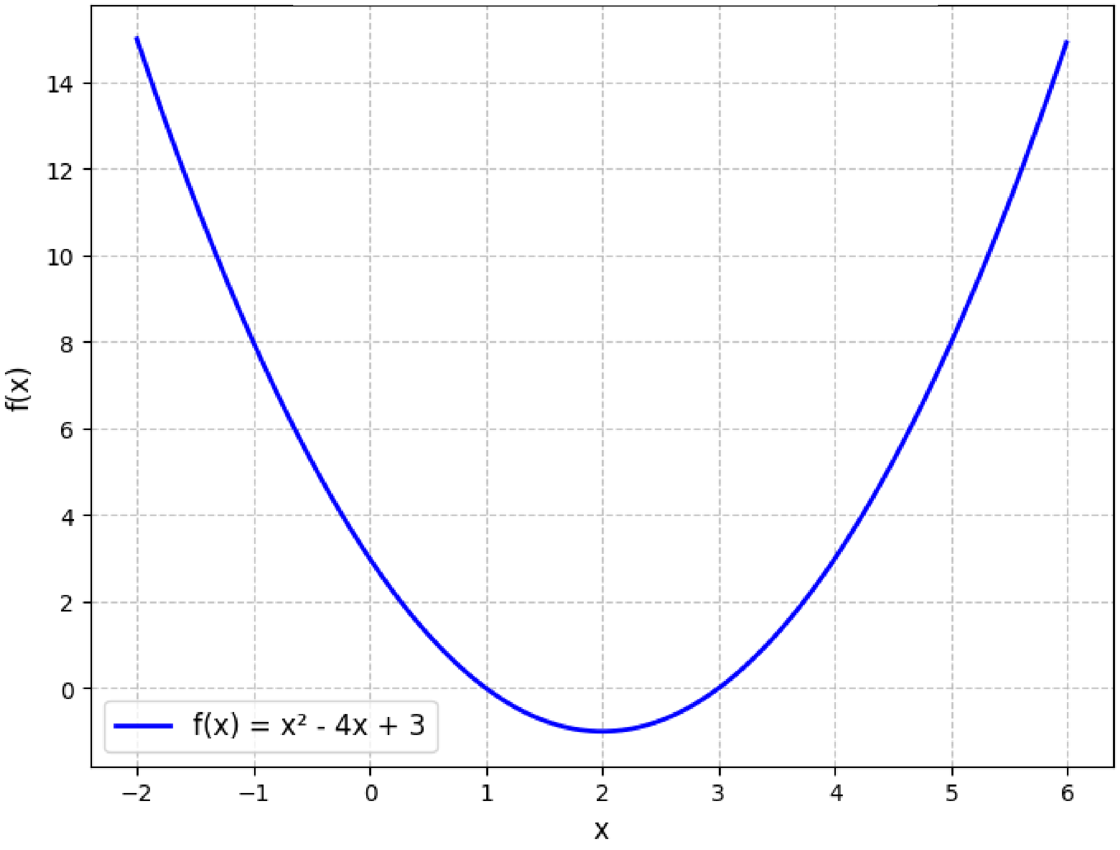
4. As a result, the minimum value of x is in the variable x_m, the minimum value of the function $f(x)$ is in the variable $\min$.

Example. Find the minimal value of the function $f(x) = x^2 - 4x + 3$ on the segment $[0,1]$ with accuracy $\varepsilon = 0{,}1$.

Let's determine the number of partition segments $n \geq \frac{b-a}{\varepsilon}$, $n \geq 10$. Let's divide the segment $[0,1]$ into 10 equal parts. $h = \frac{b-a}{n} = 0{,}1$.

Let $x_m = 0$ and $min = f(0) = 3$.

Brute force method



Brute force method

Let $i = 1$, then $x = 0 + 1 \cdot 0{,}1 = 0{,}1$; $f(x) = 0{,}01 - 0{,}4 + 3 = 2{,}61$. $2{,}61 < min \Rightarrow min = 2{,}61;\ x_m = 0{,}1$.

$i = 2$, then $x = 0{,}2$; $f(x) = 0{,}04 - 0{,}8 + 3 = 2{,}24$. $2{,}24 < min \Rightarrow min = 2{,}24;\ x_m = 0{,}2$.

$i = 3$, then $x = 0{,}3$; $f(x) = 0{,}09 - 1{,}2 + 3 = 1{,}89$. $1{,}89 < min \Rightarrow min = 1{,}89;\ x_m = 0{,}3$.

$i = 4$, then $x = 0{,}4$; $f(x) = 0{,}16 - 1{,}6 + 3 = 1{,}56$. $1{,}56 < \mathrm{min} \Rightarrow min = 1{,}56;\ x_m = 0{,}4$.

$i = 5$, then $x = 0{,}5$; $f(x) = 0{,}25 - 2 + 3 = 1{,}25$. $1{,}25 < \mathrm{min} \Rightarrow min = 1{,}25;\ x_m = 0{,}5$.

$i = 6$, then $x = 0{,}6$; $f(x) = 0{,}36 - 2{,}4 + 3 = 0{,}96$. $0{,}96 < \mathrm{min} \Rightarrow min = 0{,}96;\ x_m = 0{,}6$.

$i = 7$, then $x = 0{,}7$; $f(x) = 0{,}49 - 2{,}8 + 3 = 0{,}69$. $0{,}69 < min \Rightarrow min = 0{,}69;\ x_m = 0{,}7$.

$i = 8$, then $x = 0{,}8$; $f(x) = 0{,}64 - 3{,}2 + 3 = 0{,}44$. $0{,}44 < min \Rightarrow min = 0{,}44;\ x_m = 0{,}8$.

Brute force method

$i = 9$, then $x = 0{,}9$; $f(x) = 0{,}81 - 3{,}6 + 3 = 0{,}21$. $0{,}21 < min \Rightarrow min = 0{,}21;\ x_m = 0{,}9$.

$i = 10$, then $x = 1$; $f(x) = 1 - 4 + 3 = 0$. $0 < min \Rightarrow \boldsymbol{min = 0};\ \boldsymbol{x_m = 1}$.

Dichotomy method

The dichotomy method constructs a sequence of nested sub-segments, each of which contains at least one of the minimum points.

The dichotomy method is based on dividing the current segment $[a, b]$ containing the extremum into two equal parts. Then, one of these two parts is selected, containing the minimum. The extremum is localized by comparing two values of the optimality criterion at points located from the middle of the segment by $\delta = \varepsilon/2$, where ε is the solution error described above.

Let us consider the segment $[a, b]$ and divide it in half by the point c. The minimum point can be located either on the left or on the right side of the segment $[a, b]$.

Let's choose two points $x_1 = \frac{a+b}{2} - \frac{\delta}{2}, x_2 = \frac{a+b}{2} + \frac{\delta}{2}$.

Dichotomy method

Thus, the point x^* will fall either in the segment $[a, x_2]$ or in the segment $[x_1, b]$. If $f(x_1) < f(x_2)$, then $x^* \in [a, x_2]$ cannot fall in the segment $[x_2, b]$. If $f(x_1) > f(x_2)$, then $x^* \in [x_1, b]$. If $f(x_1) = f(x_2)$, then то $x^* \in [x_1, x_2]$.

After the first step, the problem remains the same, but the uncertainty segment has become smaller.

Dichotomy method

Algorithm of the dichotomy method.

1. We introduce the segment boundaries a, b, the accuracy ε, and calculate the parameter $\delta = \varepsilon/2$.

2. Let $x_1 = \frac{a+b-\delta}{2}$; $x_2 = \frac{a+b+\delta}{2}$. If $f(x_1) < f(x_2)$, we assume $b = x_2$, otherwise we assume $a = x_1$.

3. We check the condition for ending the iterations: $|b - a| < \varepsilon$. If the condition for ending the iterations is met, then the problem is solved, we assume $x^* = a$. If not, we go to step 2.

Dichotomy method

Example. Find the minimal value of the function $f(x) = x^2 - 4x + 3$ on the segment $[0,2]$ with accuracy $\varepsilon = 0,1$.

1. The boundaries of the segment $a = 0, b = 2$, accuracy $\varepsilon = 0.1$, parameter $\delta = \varepsilon/2 = 0.05$.

2. Let $x_1 = \frac{0+2-0,05}{2} = 0,975;$ $x_2 = \frac{0+2+0,05}{2} = 1,025.$ $f(x_1) = f(0,975) = 0.050625;$ $f(x_2) = f(1,025) = -0.049375.$ $f(x_1) > f(x_2)$, therefore, we assume $a = x_1 = 0,975$. We check the condition for ending the iterations: $|b - a| = |2 - 0.975| = 1.025 > \varepsilon$. The condition for ending the iterations was not met, so we continue the solution and narrow the segment to [0.975; 2] and move on to step 2 of the algorithm.

Dichotomy method

3. We set $x_1 = \frac{0,975+2-0,05}{2} = 1,4625;\ x_2 = \frac{0,975+2+0,05}{2} = 1,5125.$

$f(x_1) = f(1,4625) \approx -0.711;\ f(x_2) = f(1,5125) \approx -0.762.$

We got that $f(x_1) > f(x_2)$, therefore, we set $a = x_1 = 1,4625$. We check the condition for ending the iterations: $|b-a| = |2-1,4625| = 0,5375 > \varepsilon$. The condition for ending the iterations was not met, which means we continue the solution and narrow the segment to $[1.4625;\ 2]$ and go to step 2 of the algorithm.

Dichotomy method

4. We set $x_1 = \frac{1,4625+2-0,05}{2} \approx 1,706;\ x_2 = \frac{1,4625+2+0,05}{2} \approx 1,756.$

$f(x_1) = f(1,706) \approx -0.914;\ f(x_2) = f(1,756) \approx -0.941.$

We got that $f(x_1) > f(x_2)$, therefore, we set $a = x_1 = 1,70625$. We check the condition for ending the iterations: $|b - a| = |2 - 1,706| = 0,293 > \varepsilon$. The condition for ending the iterations was not met, which means we continue the solution and narrow the segment to $[1.706;\ 2]$ and go to step 2 of the algorithm.

Dichotomy method

5. We set $x_1 = \frac{1,706+2-0,05}{2} \approx 1,828;\ x_2 = \frac{1,706+2+0,05}{2} \approx 1,878.$

$f(x_1) = f(1,828) \approx -0.97;\ f(x_2) = f(1,878) \approx -0.985$

We got that $f(x_1) > f(x_2)$, therefore, we set $a = x_1 = 1,828$. We check the condition for ending the iterations: $|b - a| = |2 - 1,828| = 0,172 > \varepsilon$. The condition for ending the iterations was not met, which means we continue the solution and narrow the segment to $[1,828;\ 2]$ and go to step 2 of the algorithm.

Dichotomy method

6. We set $x_1 = \frac{1{,}828+2-0{,}05}{2} \approx 1{,}889;\ x_2 = \frac{1{,}828+2+0{,}05}{2} = 1{,}939$

$$f(x_1) = f(1{,}889) \approx -0.988;\ f(x_2) = f(1{,}939) \approx -0.996$$

We got that $f(x_1) > f(x_2)$, therefore, we set $a = x_1 = 1{,}889$. We check the condition for ending the iterations: $|b - a| = |2 - 1{,}889| = 0{,}111 > \varepsilon$. The condition for ending the iterations was not met, which means we continue the solution and narrow the segment to $[1{,}889;\ 2]$ and go to step 2 of the algorithm.

Dichotomy method

7. We set $x_1 = \frac{1,889+2-0,05}{2} \approx 1,919;\ x_2 = \frac{1,889+2+0,05}{2} \approx 1,969$

$f(x_1) = f(1,919) \approx -0.994;\ f(x_2) = f(1,969) \approx -0.999$

We got that $f(x_1) > f(x_2)$, therefore, we set $a = x_1 = 1,919$. We check the condition for ending the iterations: $|b - a| = |2 - 1,919| = 0,08 < \varepsilon$. The condition for ending the iterations was met, so we set $x^* = a \approx 1,919$ and complete the solution. Thus, we obtained the minimum point $\boldsymbol{x_{min}} \approx \mathbf{1,919}$ and the smallest value of the function on the considered segment is $-\mathbf{0.993}$.

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 α 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, t he 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 $F_k(\alpha) = f(x^k - \alpha_k f'(x^k))$, α_k we find from the problem of one-dimensional minimization $F_k(\alpha) \to \min_{\alpha \geq 0}$.

Gradient Descent Method

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

$$f(x) = \frac{1}{2}x^T A x + b^T x + c$$

1. Calculate the gradient descent direction $g^k = -(Ax^k + b)$.

2. Determine the step α_k from the condition $F_k(\alpha) \to \min\limits_{\alpha \geq 0}$:

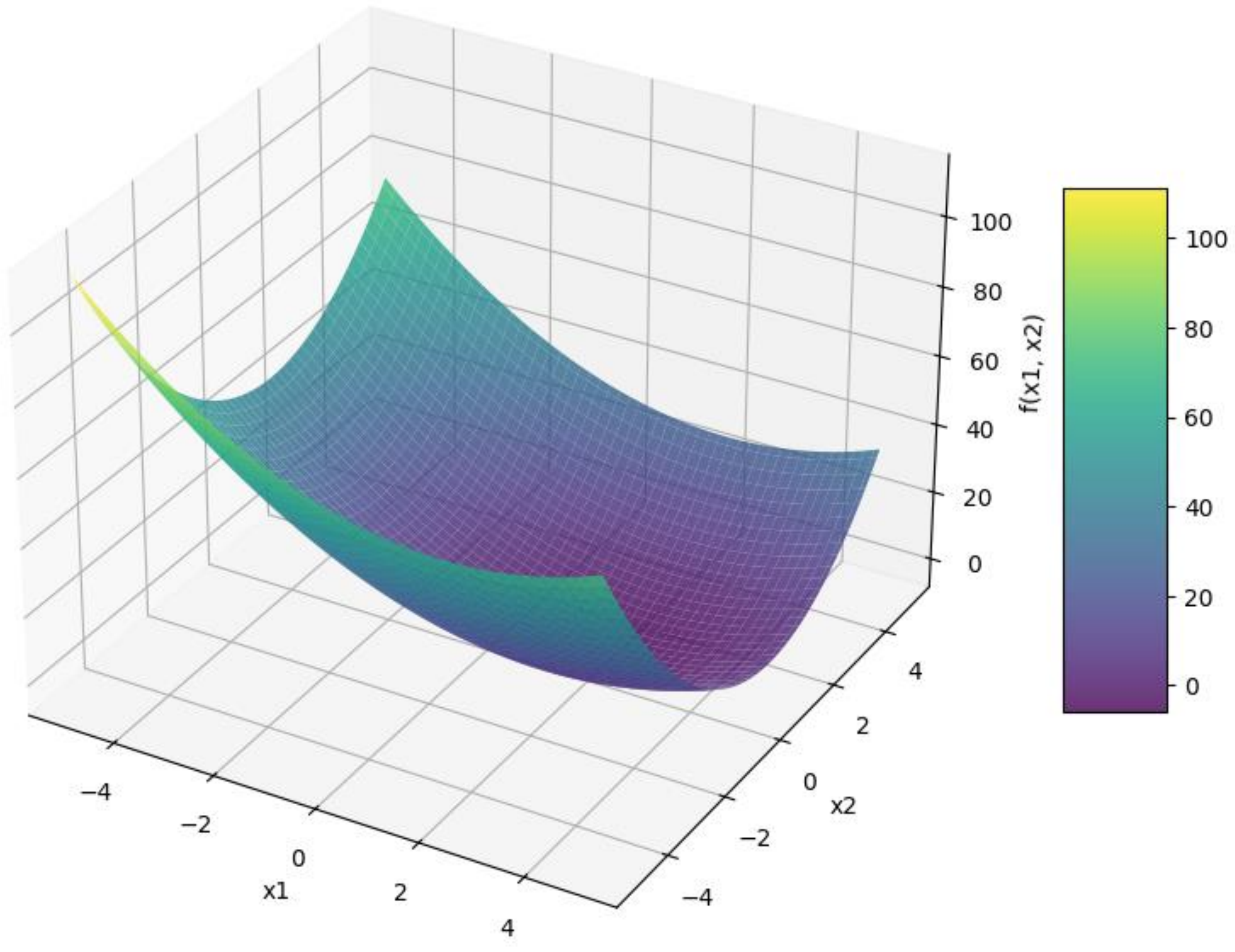
$$F_k(\alpha) = f\left(x^k + \alpha g^k\right) = f\left(x^k\right) - \alpha\left\|g^k\right\|^2 + \frac{1}{2}\alpha^2(Ag^k, g^k)$$

$$F_k'(\alpha) = -\|g^k\|^2 + \alpha(Ag^k, g^k) = 0 \qquad \alpha_k = \frac{\left\|g^k\right\|^2}{(Ag^k, g^k)} > 0$$

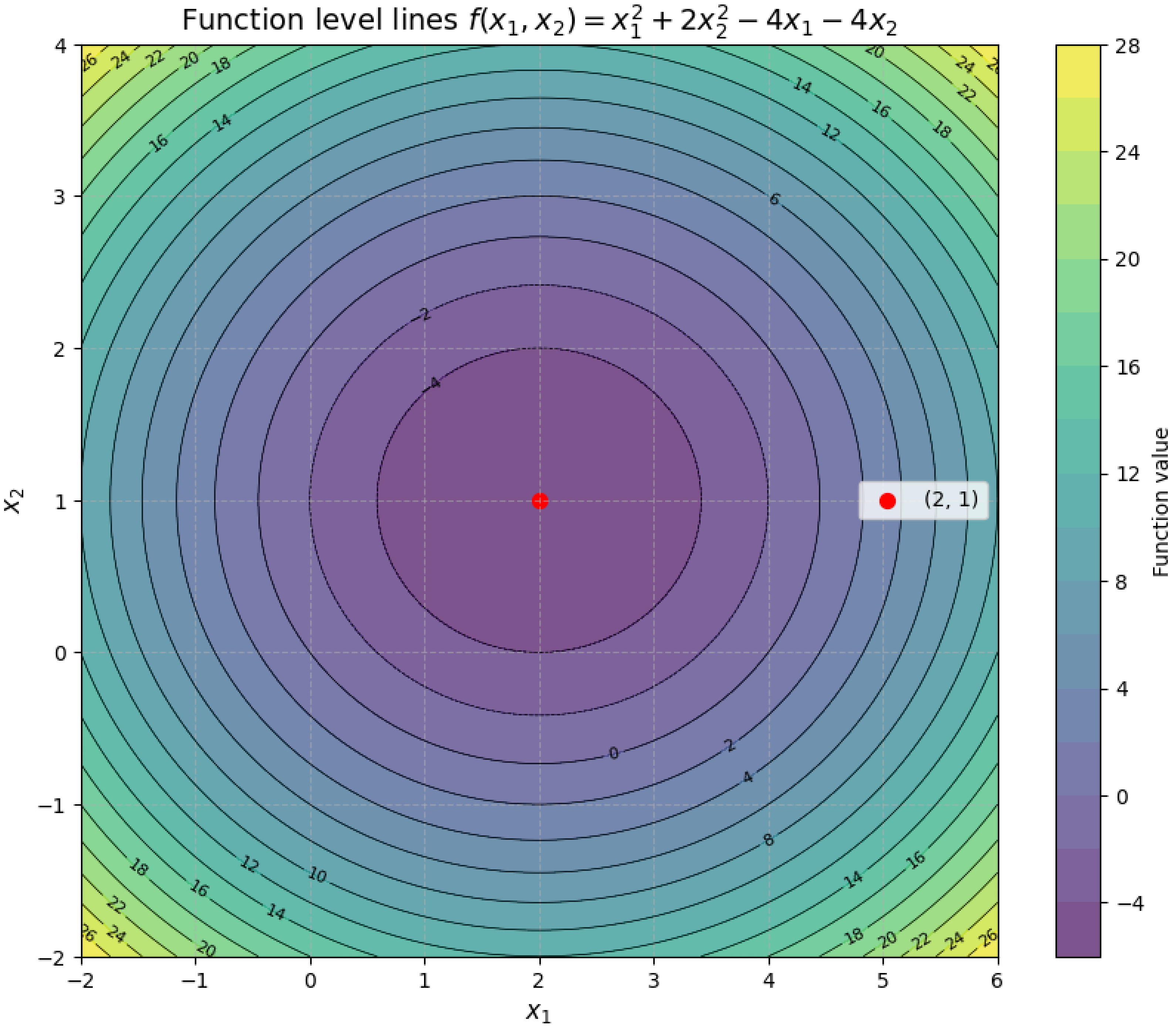
3. We proceed to the next iteration $x^{k+1} = x^k + \alpha_k g^k$. We check the fulfillment of the iteration termination criterion. If the criterion is fulfilled, then we assume $x^* \approx x^{k+1}$, otherwise we proceed to point 1. As a result, we obtained a sequence $\{x^k\}$ of decreasing function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

Gradient Descent Method

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



−4
−2
0
2
4
x1
−4
−2
0
2
4
x2
0
20
40
60
80
100
f(x1, x2)
0
20
40
60
80
100


Function level lines $f(x_1, x_2) = x_1^2 + 2x_2^2 - 4x_1 - 4x_2$
(2, 1)
x_1
x_2
Function value

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

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

6. $x^2 = x^1 + \alpha_1 g^1 = \left(\frac{4}{3}; \frac{4}{3}\right) - (\frac{4}{3}; -\frac{4}{3})^T = \left(\frac{16}{9}; \frac{8}{9}\right)^{\mathrm{T}}$. 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. $g^2 = -(Ax^2 + b) = (\frac{4}{9}; \frac{4}{9})^T$

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

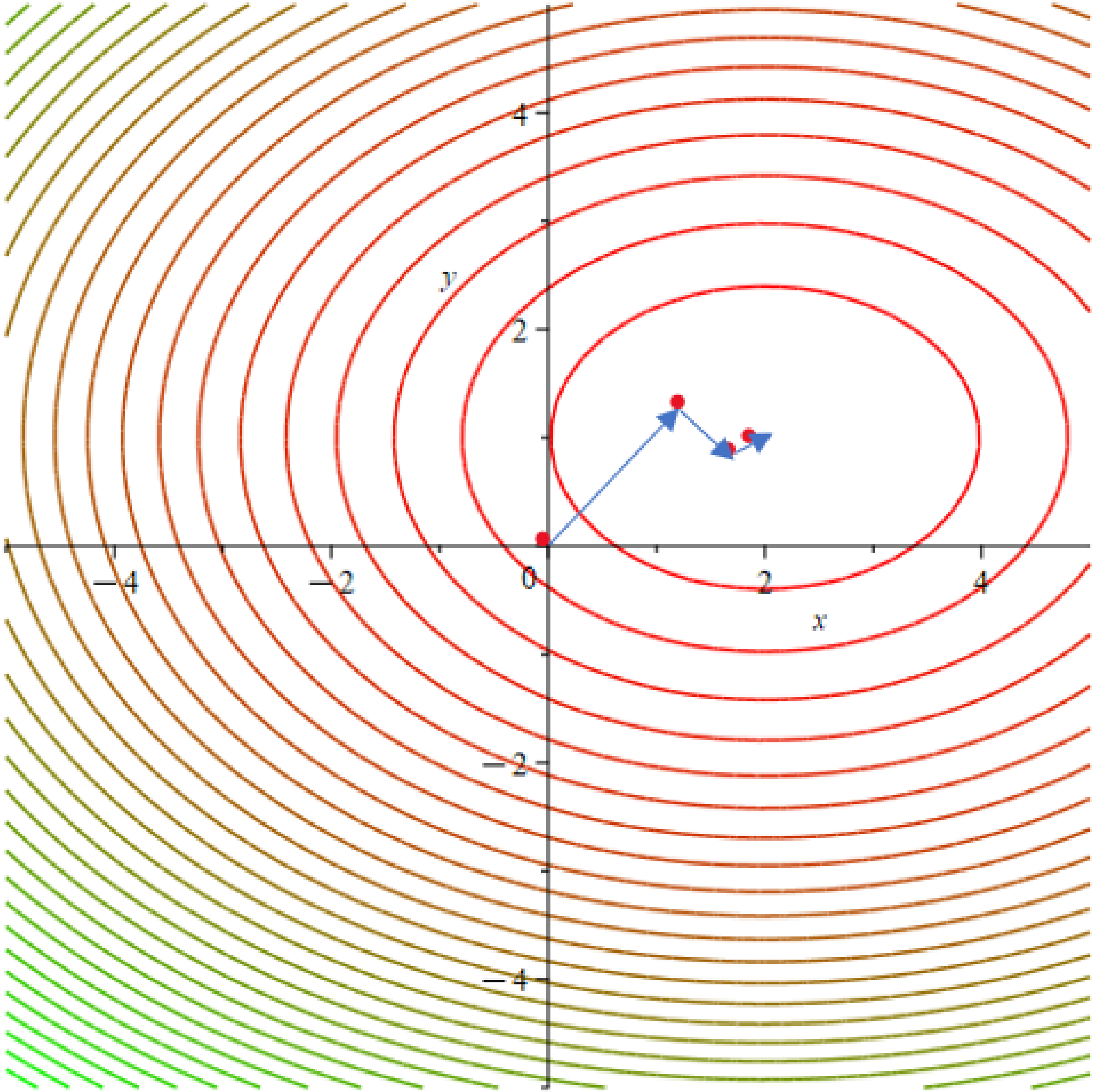
Gradient Descent Method

9. $x^3 = x^2 + \alpha_2 g^2 = \left(\frac{52}{27}; \frac{28}{27}\right)^{\mathrm{T}}$. 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 $\left\|Ax^k + b\right\|$ is close to zero. Note that $\left\|Ax^k + b\right\| = \frac{\sqrt{32}}{3^k} \to 0, k \to \infty$.

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

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



4
2
y
0
−4
−2
2
4
x
−2
−4

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 $f(x) \to min$. The cycle with number k+1 has the form:

1. Descent along coordinate x_1. To do this, we fix all other coordinates $x_2 = x_2^k, x_3 = x_3^k, \ldots, x_n = x_n^k$ and solve the problem of minimizing a function of one variable $f\left(x_1, x_2^k, x_3^k, \ldots, x_n^k\right) \to \min_{x_1}$. 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 α_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 $f\left(x_1^{k+1}, x_2, x_3^k, \dots, x_n^k\right) \to \min\limits_{x_2}$. 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 $f\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \dots, x_n^k\right) \to \min\limits_{x_n}$.

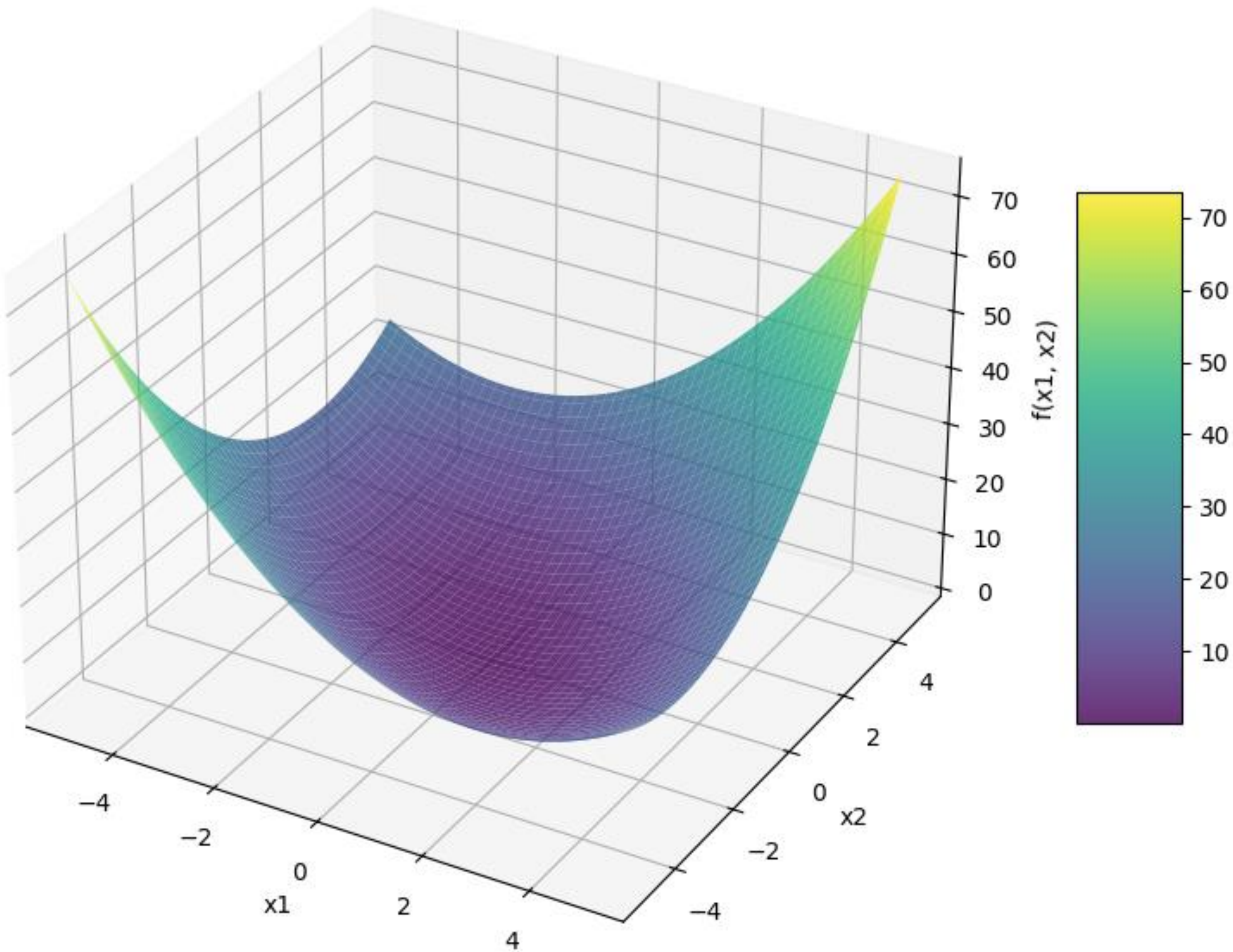
As a result, we obtain a sequence of points $\left(x_1^{k+1}, x_2^k, x_3^k, \dots, x_n^k\right), \left(x_1^{k+1}, x_2^{k+1}, x_3^k, \dots, x_n^k\right), \dots,$ $\left(x_1^{k+1}, x_2^{k+1}, x_3^k, \dots, x_n^{k+1}\right)$.

5. Check the loop termination condition:

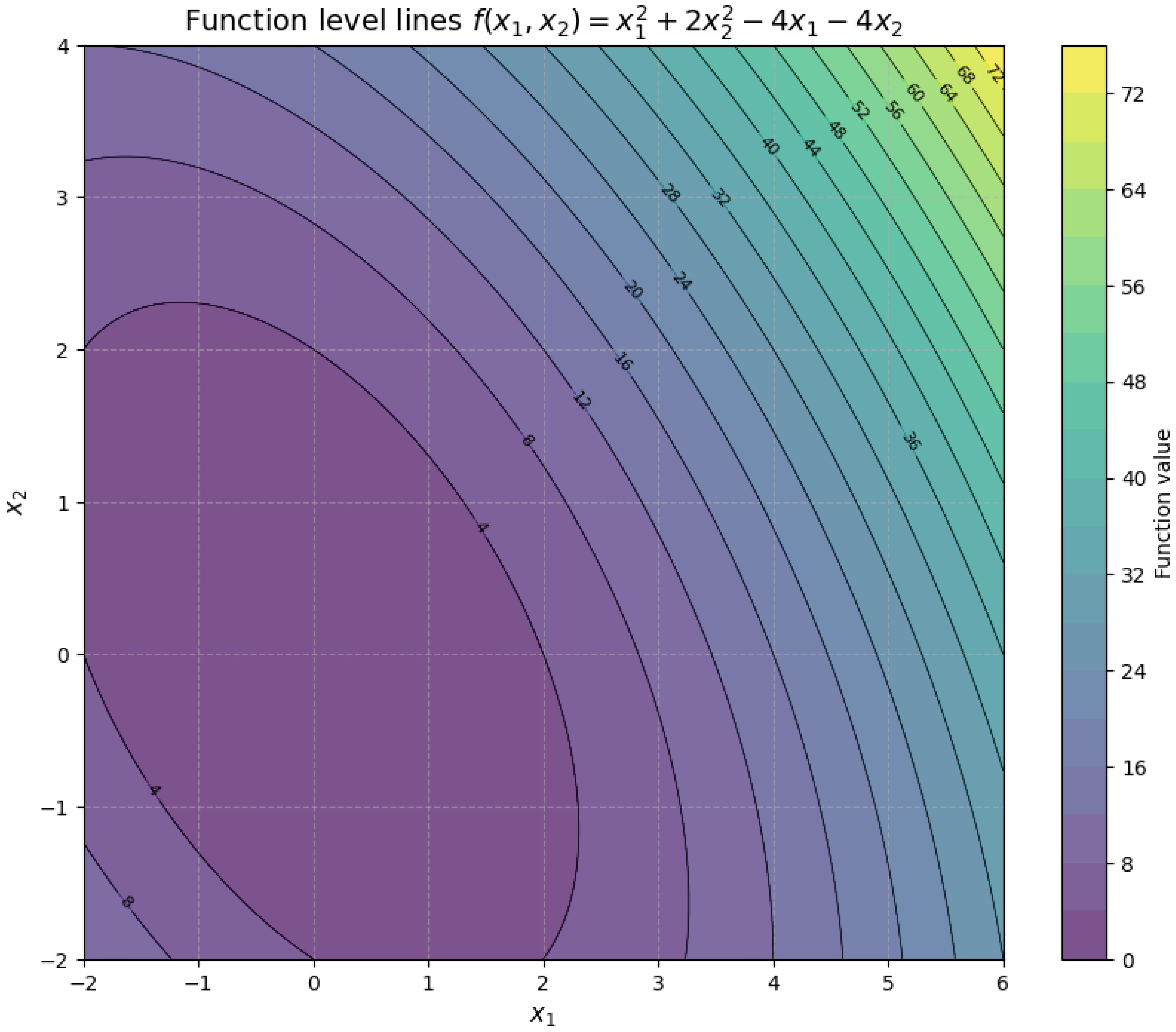
$$\left|f\left(x_1^{k+1}, x_2^{k+1}, x_3^{k+1}, \dots, x_n^{k+1}\right) - f\left(x_1^k, x_2^k, x_3^k, \dots, x_n^k\right)\right| < \varepsilon$$

Coordinate descent method

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



x1
x2
f(x1, x2)
−4
−2
0
2
4
−4
−2
0
2
4
0
10
20
30
40
50
60
70
10
20
30
40
50
60
70



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

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:

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

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

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

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

Coordinate descent method

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

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

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

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:

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

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

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:

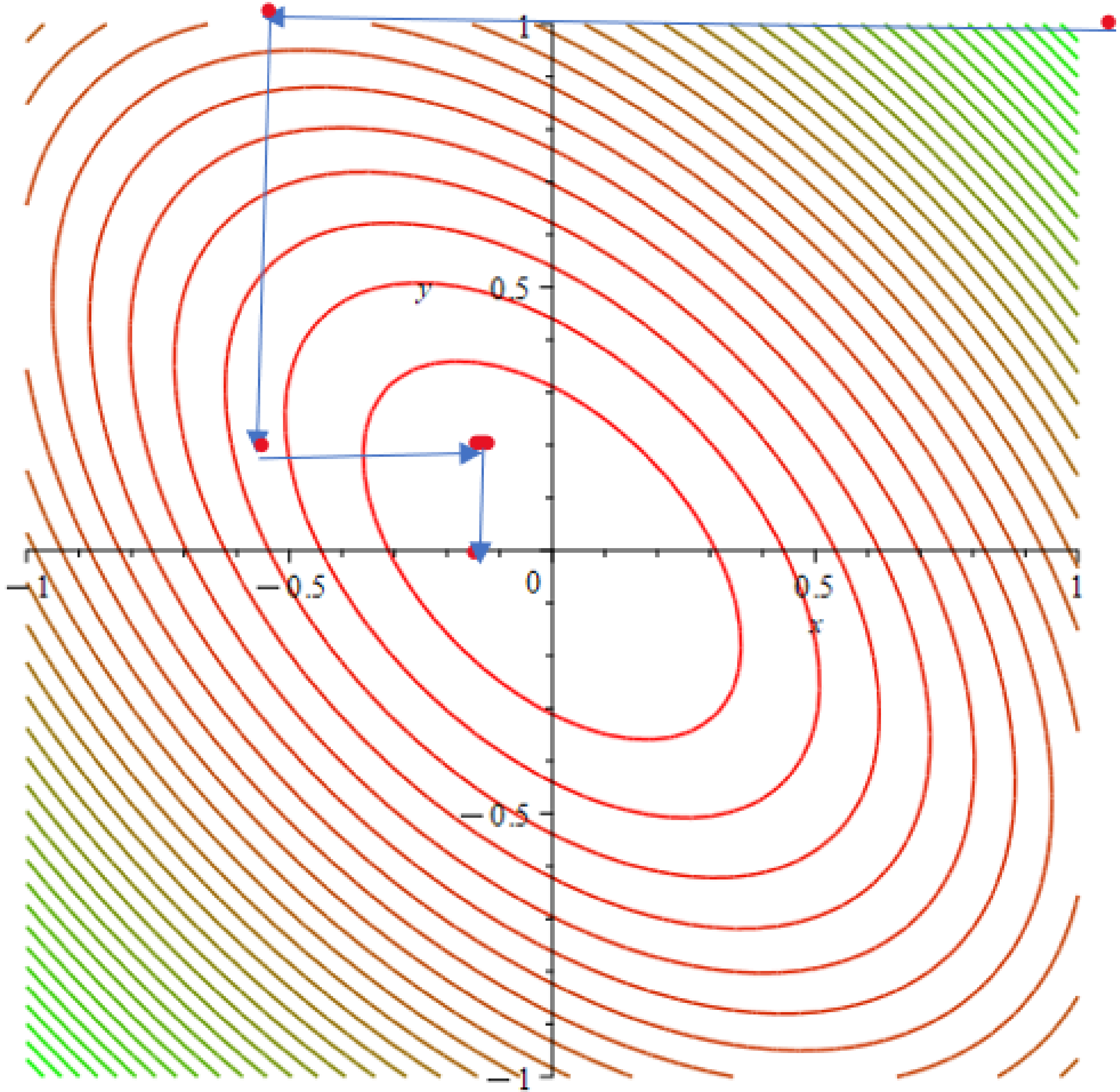
$$f(x_2) = \frac{3}{4}\left(x_2 - \frac{1}{8}\right)^2 + \frac{1}{4}\left(x_2 + \frac{1}{8}\right)^2 \rightarrow min$$

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

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: $\left\{(1;1);\left(-\frac{1}{2};1\right);\left(-\frac{1}{2};\frac{1}{4}\right);\left(-\frac{1}{8};\frac{1}{4}\right);\left(-\frac{1}{8};\frac{1}{16}\right)\right\}$.



1
0.5
y
−1
−0.5
0
0.5
1
x
−0.5
−1