Finite dimensional optimization methods
Lecture 1

2021

Introduction to Optimization

Currently, optimization is used in almost every field of science and technology:

1. in operations research: optimization of technical and economic systems, transport tasks, management, etc;
2. in numerical analysis: approximation, solution of linear and nonlinear problems etc;
3. in automation: optimal control systems, robots etc;
4. in engineering: managing the size and optimize the structure of the optimal planning of complex technical systems, such as information systems;
5. in combinatorial optimization, many basic algorithms: : in traffic problems, in integer and Boolean programming and in many others, using necessary optimality conditions, the concept of duality.

Formulation of optimization problem

Consider the general formulation of the minimization problem.
Let $X \subset \Omega \subset \mathbb{R}^n$ be a closed set and a function f be defined on an open set Ω ($f : \Omega \to \mathbb{R}$).
Required to find a point $x^* \in X$, for which the inequality

$$f(x^*) \leq f(x) \quad \forall x \in X, \tag{1}$$

holds.
Of course, it may be such x^* does not exist.
A point $x^* \in X$ satisfying condition (1) is called a global minimizer of f in X.

Condition (1) can be rewritten as

$$\inf_{x \in X} f(x). \tag{2}$$

A function f is called an objective function.
A set X is called an admissible set.
A point $x^* \in X$ is called a solution of (1).
If $X = \mathbb{R}^n$, then problem (2) is called an unconstrained minimization problem and the point x^* is called a gobal minimizer of f over $\mathbb{R}^n$.

If at x^*

$$f(x^*) \leq f(x) \quad \forall x \in U(x^*) \subset X, \tag{3}$$

is satisfied (where $U(x^*) \subset \mathbb{R}^n$ is a neighborhood of x^*) then x^* is called a local minimizer of f in X.

If in (1) and (3) a strict inequality holds at $x \neq x^*$, then we say that x^* is called a strict global or local minimizer.

X can be defined through the system of equalities and inequalities in the form:

$$X = \{x \in \Omega \subset \mathbb{R}^n \mid \varphi_i(x) \leq 0, i = 1, .., m, \psi_j(x) = 0, \ j = 1, .., s\} .$$

Functions φ_i, ψ_j are called constraints of problem (1).

Any point $x \in X$, satisfying these constraints is called an admissible solution of (1).

It should be noted that sometimes a constraint equation can be written in the form of two inequalities.

Classification of optimization problems

Let us also assume that all functions are continuous unless mentioned otherwise.
The classification of optimization problems can be carried out on several characteristics, depending on the type of f and X:

1. static, dynamic (eg, task management);
2. unconstrained and constrained optimization;
3. with continuous and discrete variables (partially - integer, integer Boolean variables);
   For example, if X is a discrete set: discrete optimization
   $$\min(x + x^2 - x^3) \quad \text{over} \quad N$$
4. one and multi-objective optimization;
5. linear and non-linear optimization;
6. one-dimensional and multidimensional, and multidimensional problems can be small and large dimension;
7. with convex and nonconvex objective functions.

In most cases, optimization problem (1) cannot be solved on the basis of necessary and sufficient conditions of optimality or using a geometric interpretation of the problem, and you have to solve it numerically using computer technologies. Moreover, the most effective methods are methods specifically designed to address for a particular class of optimization problems, as they allow to account of its specificity fuller.

Any numerical method has two phases.
The first phase of any numerical method for solving the optimization problem is based on the exact or approximate calculation of values of the objective function f, values of the functions defining the feasible set, as well as their derivatives. Then we have to construct an approximation for the solution of the problem.

Depending on the type of objective functions algorithms are divided into:

1. zero-order - they use only information about the values of the objective function;
2. the first order - they use information also about values of the first derivatives;
3. a second order - using, in addition, information about the second derivatives.

Methods of minimization can be divided into finite- and infinite-step methods. Finite-(or end) methods are called methods which theoretically guarantee the finding a solution of the problem in a finite number of steps.

These include, for example, Newton's method, methods of conjugate directions for minimization of a convex quadratic function with a positive definite matrix, the simplex method solving the problem of linear programming.

In practical implementation, of course, there exist computational errors.

In practice, often there are ofen optimization problems that have many extremes.
There is no universal answer to this question. The simplest method is when the search is conducted several times, starting from different starting points.
If results have different answers, then the values of the objective function are compared and chosen the smallest values.
Calculations stop if several new steps do not change the results obtained earlier.

Some theorems of Mathematical Analysis and Algebra.

Formulate some known results of mathematical analysis, which will be used in the future.
Before solving an optimization problem, it is necessary to prove a theorem of the existence of solutions.
Of course, the problem of the existence of a minimizer of a continuous function in a compact set always has at least one solution.

Theorem 1 (Weierstrass).

A continuous function on a compact set attains its minimum and maximum values.

In the case when $X \in \mathbb{R}^n$ is not compact for the existence of minimizers, it is necessary to impose additional conditions.

Theorem 2.

Let f be continuous on $\mathbb{R}^n$ and for some $\alpha \in \mathbb{R}$ the set

$$\mathcal{L}_\alpha = \{x \in \mathbb{R}^n \mid f(x) \leq \alpha\}$$

is nonempty and bounded. Then the function f attains its minimum on $\mathbb{R}^n$.

The set $\mathcal{L}_\alpha$ is called the Lebesgue set.

Proof. Since f is continuous, then set $\mathcal{L}_\alpha$ is closed. By the Weierstrass theorem, the function f reaches its minimum over the set $\mathcal{L}_\alpha$ at x^*

$$x^* = \arg\min_{x \in \mathcal{L}_\alpha} f(x).$$

It is obvious that x^* is also a minimizer of f over $\mathbb{R}^n$.

Consider some lemmas which will be used to prove convergence theorems.

Lemma 1.

Let f be a twice continuously differentiable function on $\mathbb{R}^n$, $x^*, g \in \mathbb{R}^n$ and $x = x^* + g$. Then

$$f(x) = f(x^*) + \int_0^1 \langle f'(x^* + tg), g \rangle dt$$

and

$$f'(x) = f'(x^*) + \int_0^1 f''(x^* + tg)\, g\, dt.$$

Lemma 2.

Let f be a twice continuously differentiable function in a neighborhood of $x^* \in \mathbb{R}^n$. Then for any $g \in \mathbb{R}^n$ for which its norm is sufficiently small, we have the Taylor formula

$$f(x^* + g) = f(x^*) + \langle f'(x^*), g \rangle + \frac{1}{2}\langle f''(x^*)g, g \rangle + o(\|g\|^2),$$

where

$$\frac{o(\|g\|^2)}{\|g\|^2} \xrightarrow[||g|| \to 0]{} 0.$$

Here by $f'(x^* + g)$ denotes the gradient of f at $x^* + g$,
by $f''(x^*)$ denotes the Hessian matrix (the matrix of second order partial derivatives) of the function f at x^*,
by $\langle \cdot, \cdot \rangle$ denotes the canonical inner product

$$\langle x, y \rangle = \sum_{i=1}^{n} x_i \, y_i$$

.

Consider a symmetric square matrix A of order $n \times n$.
A matrix A is called nondegenerate if its determinant is nonzero.
In this case, there exists an inverse matrix A^{-1}.
A matrix A is called positive definite ($A > 0$) if

$$\langle Ag, g \rangle > 0 \quad \forall g \in \mathbb{R}^n, \ ||g|| \neq 0.$$

A matrix A is called nonnegative definite or positive semidefinite ($A \geq 0$) if

$$\langle Ag, g \rangle \geq 0 \quad \forall g \in \mathbb{R}^n.$$

Similarly, we define the concept of negative and nonpositive definite matrix.

Lemma 3 (The Sylvester criterion)

A matrix A is positive definite if and only if all principal diagonal minors of the matrix A are positive.
A matrix A is negative definite if and only if the signs of the principal diagonal minors of the matrix A alternate and moreover $\Delta_1 = a_{11} < 0$.

The equation

$$|A - \lambda E| = 0 \tag{4}$$

is called the characteristic equation, where E is the identity matrix.
It is known from the fundamental theorem of algebra that equation (4) has n roots.
The roots of equation (4) are called eigenvalues of A.

Lemma 4.

Eigenvalues of a symmetric matrix A are real.
If the matrix A is positive definite, then the next inequalities

$$\lambda_{min}||g||^2 \leq \langle Ag, g \rangle \leq \lambda_{max}||g||^2 \quad \forall g \in \mathbb{R}^n,$$

hold, where
$\lambda_{min} > 0$ *is a minimum eigenvalue of A,*
λ_{max} *is a maximum eigenvalue of A.*
If the matrix A is positive definite, all its eigenvalues are positive.

A function f is called sublinear or semi-additive if the inequality

$$f(x+y) \leq f(x) + f(y) \quad \forall x, y \in \mathbb{R}^n.$$

holds.

A function f is called positive homogeneous if the following equality

$$f(\alpha x) = \alpha f(x) \quad \forall x \in \mathbb{R}^n, \ \forall \alpha > 0.$$

holds.

A function f is called homogeneous if the following equality

$$f(\alpha x) = |\alpha| f(x) \quad \forall x \in \mathbb{R}^n, \ \forall \alpha \in \mathbb{R}.$$

holds.

A homogeneous and sublinear function $p : \mathbb{R}^n \to \mathbb{R}$ is called a seminorm.
If a seminorm $p : \mathbb{R}^n \to \mathbb{R}$ has the following property

$$1)\ p(x) = 0 \Rightarrow x = 0_n \quad \forall x \in \mathbb{R}^n,$$

then it is called a norm.
From this it follows that

$$p(x) \geq 0 \quad \forall x \in \mathbb{R}^n.$$

Here are examples of norms that are used in $\mathbb{R}^n$. Let $x = (x_1, x_2, \dots, n) \in \mathbb{R}^n$.

$$1)\ ||x||_1 = \sum_{i=1}^{n} |x_i|;$$

$$2)\ ||x||_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2};$$

$$3)\ ||x||_\infty = \max_{i=1,\dots,n} |x_i|;$$

$$4)\ ||x||_q = \sqrt[q]{\sum_{i=1}^{n} |x_i|^q}, \quad q \geq 1.$$

The norm $||*||_2$ is called the Euclidean norm.
The latest norm is called the $q-$ Holder norm.
Note that the norms with indices 1 and ∞ are nonsmooth.

The concept of a unit circle (the set of all vectors of norm 1) is different in different norms:
for the 1-norm, the unit circle in $\mathbb{R}^2$ is a square,
for the 2-norm (the Euclidean norm), it is the well-known unit circle, while for the infinity norm it is a different square.
Due to the definition of the norm, the unit circle must be convex and centrally symmetric (therefore, for example, the unit ball may be a rectangle but cannot be a triangle).

Examples of optimization problems

Let us

$$Ax = b, \tag{5}$$

be a system of linear equations, where A is a matrix of order $m \times n$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

If system (5) is not compatible, then there is a problem of the minimization of the norm of the difference of right and left parts of the system. Required to find

$$\min_{x \in \mathbb{R}^n} ||Ax - b||.$$

The norm can be taken various. For example, if we take the Euclidean norm, we then have the problem of the best quadratic approximation

$$\sqrt{\sum_{i=1}^{m}\left(\sum_{j=1}^{m} a_{ij}x_i - b_j\right)^2} \to \min,$$

or

$$\sum_{i=1}^{m}\left(\sum_{j=1}^{m} a_{ij}x_i - b_j\right)^2 \to \min, \tag{6}$$

It may be that the problem (6) is not the only solution.

Example 1.

Let us

$$\begin{cases} x_1 + x_2 = 1, \\ x_1 + x_2 = 2. \end{cases}$$

be a system of linear equations. Here

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \in \mathbb{R}^2, \quad x = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2.$$

It is obvious that this system does not have solutions. Construct the quadratic functional

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

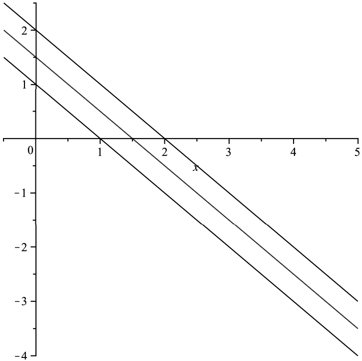
Find the gradient of f and equate it to zero. Then

$$f^{'}(x) = \begin{pmatrix} 4x_1 + 4x_2 - 6 \\ 4x_1 + 4x_2 - 6 \end{pmatrix}.$$

Thus, all points on the line

$$L = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 + 2x_2 - 3 = 0\}$$

are minimizers of f on $\mathbb{R}^2$.
On fig. 1 you can see a graph of these lines.


2
1
0
1
2
3
4
5
x
-1
-2
-3
-4

Example 2.

Let

i	1	2	3	4	5
x_i	0	1	2	3	4
y_i	2	1	3	2	5

be a table of numbers.
It is necessary to find constants a and b of a line $y = ax + b$, so the function

$$f(a,b) = \sum_{i=1}^{5} (y_i - (ax_i + b))^2 = (2-b)^2 + (1-a-b)^2 +$$

$$+(3-2a-b)^2 + (2-3a-b)^2 + (5-4a-b)^2$$

reaches a minimum value on $\mathbb{R}^2$.

Find the gradient of f:

$$\frac{\partial f(a,b)}{\partial a} = 60a + 20b - 66,$$

$$\frac{\partial f(a,b)}{\partial b} = 20a + 10b - 26.$$

Solve the system of equations

$$\begin{cases} 30a + 10b = 33, \\ 10a + 5b = 13. \end{cases}$$

Numbers $a^* = 0.7,\ b^* = 1.2$ are the solutions of this system of linear equations. Thus, the sum of squares of the difference between the specified points and line $y = 0.7x + 1.2$ is the smallest, with $f(a^*, b^*) = 4.3$ and

$$| y_1 - a^* x_1 - b^* | = 0.8,$$

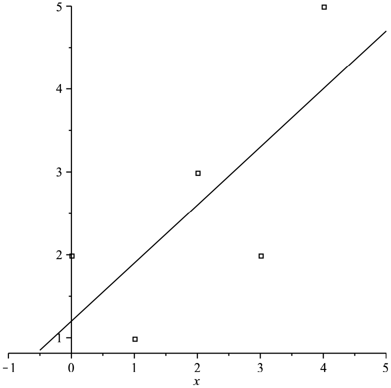
$$| y_2 - a^* x_2 - b^* | = 0.9,$$

$$| y_3 - a^* x_3 - b^* | = 0.4,$$

$$| y_4 - a^* x_4 - b^* | = 1.3,$$

$$| y_5 - a^* x_5 - b^* | = 1.$$

For this problem the minimizer (a^*, b^*) is unique.
On Fig. 2 you can see a graph of this linear function and given points.


5
4
3
2
1
−1
0
1
2
3
4
5
x

Convex sets

A set $X \subset \mathbb{R}^n$ is called convex, if for all $x_1 \in X$ and $x_2 \in X$ the next formula

$$\lambda x_1 + (1-\lambda)x_2 \in X \quad \forall \lambda \in [0,1] \subset \mathbb{R},$$

holds.
A convex set is a set that contains all the points of any line segment joining two points of the set.

We assume that the empty set $\emptyset$ is convex by definition.
The sum of two convex sets $X_1, X_2 \subset \mathbb{R}^n$ is called the set of

$$X = X_1 + X_2 = \{x_1 + x_2 \mid x_1 \in X_1, \quad x_2 \in X_2\}.$$

Sometimes the set $X = X_1 + X_2$ is called *the algebraic sum of two convex sets* X_1 *and* X_2 or *the Minkowski sum.*
By writing $X_1 - X_2$, we will understand the set $X_1 + (-X_2)$.
Let $a \in \mathbb{R}^n$. The set of $X + a$ is called a *translator* set $X \subset \mathbb{R}^n$.

If the set is the set $X \subset \mathbb{R}^n$ is convex, then its every scalar multiple $\alpha X, \alpha \in \mathbb{R}$,

$$\alpha X = \{y \in \mathbb{R}^n \mid y = \alpha x,\ x \in X\}.$$

The set of αX for $\alpha > 0$ is the image set X with stretching or squeezing the space $\mathbb{R}^n$ in α times relative to the origin.
Internal points of X we denote by int (X). The closure of X will be denoted by cl (X), by using bd (X), we denote a set of boundary points of X.

We note some elementary properties of convex sets.

Lemma 5. *1. Let I be an arbitrary set of indices i, a set $X_i \subset \mathbb{R}^n$ is convex for each index $i \in I$. Then*

$$X = \bigcap_{i \in I} X_i$$

is convex as well.

2. Let sets $X_1 \subset \mathbb{R}^n$ and $X_2 \subset \mathbb{R}^n$ be convex, then the algebraic sum of these sets is also convex.

Convex functions

Convex functions play an important role in optimization theory. For them, the questions of existence and uniqueness can be solved quite easily.
The function $f : \mathbb{R}^n \to \mathbb{R}$ is called convex if

$$f(\lambda_1 x_1 + \lambda_2 x_2) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) \tag{7}$$

$$\forall\ x_1, x_2 \in \mathbb{R}^n,\ \ \lambda_1, \lambda_2 \in [0,1],\ \ \lambda_1 + \lambda_2 = 1.$$

Geometrically convexity means that the graph of the function f on an interval $[x_1, x_2]$ connecting points x_1 and x_2 lies not above a straight line connecting points $(x_1, f(x_1)) \in \mathbb{R}^{n+1}$ and $(x_2, f(x_2)) \in \mathbb{R}^{n+1}$

From the definition of convex functions it is easy to establish the Jensen inequality:

$$f\left(\sum_{i=1}^{k}\lambda_i x_i\right) \leq \sum_{i=1}^{k}\lambda_i f(x_i)$$

$$\forall x_i \in \mathbb{R}^n,\ \lambda_i \geq 0,\ i = 1, \ldots, k,\ \sum_{i=1}^{k}\lambda_i = 1.$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called strictly convex if

$$f(\lambda_1 x_1 + \lambda_2 x_2) < \lambda_1 f(x_1) + \lambda_2 f(x_2)$$

$$\forall\ x_1, x_2 \in \mathbb{R}^n,\ x_1 \neq x_2,\ \lambda_1, \lambda_2 \in (0,1),\ \lambda_1 + \lambda_2 = 1.$$

A function $f : \mathbb{R}^n \to \mathbb{R}$ is called strongly convex with a constant of strongly convex $m > 0$, if

$$f\left(\lambda_1 x_1 + \lambda_2 x_2\right) \leq \lambda_1 f(x_1) + \lambda_2 f(x_2) - m\lambda_1\lambda_2 ||x_1 - x_2||^2$$

$$\forall\ x_1, x_2 \in \mathbb{R}^n,\ x_1 \neq x_2,\ \lambda_1, \lambda_2 \in (0,1),\ \lambda_1 + \lambda_2 = 1.$$

We give some properties of convex functions.

1. A function

$$f(x) = \sum_{i=1}^{m} \lambda_i f_i(x)$$

is convex, if functions $f_i,\ i \in 1, \dots, m$, *are convex and numbers* λ_i *is non-negative.*

2. A function

$$f(x) = \sup_{i \in I} f_i(x)$$

is convex if functions f_i, *are convex and* $i \in I$, I *is an arbitrary set of indices.*

3. A set $\mathcal{L} = \{x \in \mathbb{R}^n \;\big|\; f(x) \le a\},\ a \in \mathbb{R}$, *is convex.*

4. If a convex function is finite at every point of the space $\mathbb{R}^n$, *then it is continuous on* $\mathbb{R}^n$.

Lemma 6. *The convexity of a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is equivalent to the following inequality*

$$f(x_2) \geq f(x_1) + \langle f'(x_1), x_2 - x_1 \rangle \quad \forall\, x_1, x_2 \in \mathbb{R}^n, \tag{8}$$

the strictly convexity is equivalent to the following inequality

$$f(x_2) > f(x_1) + \langle f'(x_1), x_2 - x_1 \rangle \quad \forall\, x_1, x_2 \in \mathbb{R}^n,\ x_1 \neq x_2, \tag{9}$$

the strongly convexity is equivalent to the following inequality

$$f(x_2) \geq f(x_1) + \langle f'(x_1), x_2 - x_1 \rangle + m||x_1 - x_2||^2 \quad \forall\, x_1, x_2 \in \mathbb{R}^n. \tag{10}$$

Proof. Let us prove inequality (10). Let $x_2 = x_1 + g$. As f is the strong convex function, then for every $\alpha \in (0;1)$ we have the following

$$f(x_2) = f(x_1 + \alpha g) = f(\alpha x_2 + (1-\alpha)x_1) \leq \alpha f(x_2) +$$

$$+(1-\alpha)f(x_1) - m\alpha(1-\alpha)\ ||g||^2.$$

From this it follows

$$\frac{f(x_1 + \alpha g) - f(x_1)}{\alpha} + m(1-\alpha)||x_2 - x_1||^2 \leq f(x_2) - f(x_1).$$

Taking the limit when $\alpha \to +0$ we get

$$f(x_1 + \alpha g) - f(x_1) \geq f'(x_1, g) = \langle f'(x_0), x_1 - x_2 \rangle + m||x_2 - x_1||^2.$$

Inequality (10) is proved.

Similarly we prove (8), (9).

Corollary. *From (8) – (10) it follows*

a) for differentiable convex functions, the following inequality

$$\langle f'(x_1) - f'(x_2), x_1 - x_2 \rangle \geq 0 \quad \forall\, x_1, x_2 \in \mathbb{R}^n$$

is satisfied;

b) for differentiable strictly convex functions the following inequality

$$\langle f'(x_1) - f'(x_2), x_1 - x_2 \rangle > 0 \quad \forall\, x_1, x_2 \in \mathbb{R}^n,\ x_1 \neq x_2$$

is satisfied;

c) for differentiable strong convex functions the following inequality

$$\langle f'(x_1) - f'(x_2), x_1 - x_2 \rangle \geq 2m||x_1 - x_2||^2 \quad \forall\, x_1, x_2 \in \mathbb{R}^n. \quad (11)$$

is satisfied.

Proof. Let us prove inequality (11). For any points $x_1, x_2 \in \mathbb{R}^n$ we have

$$f(x_2) - f(x_1) \geq \langle f'(x_1), x_2 - x_1 \rangle + m||x_2 - x_1||^2,$$

$$f(x_1) - f(x_2) \geq \langle f'(x_2), x_1 - x_2 \rangle + m||x_1 - x_2||^2.$$

Summing these inequalities, we obtain (11).

Lemma 7. *If a differentiable function f is strongly convex with a constant m, then it attains its minimum at a single point on $\mathbb{R}^n$.*
Proof. First, we prove that the minimum of strongly convex functions is attained. Fix an arbitrary point $x \in \mathbb{R}^n$. Denote by $r = \dfrac{1}{m}||f'(x)||$ and consider the closed ball with center at x and radius r

$$S(x, r) = \{y \in \mathbb{R}^n \mid ||y - x|| \leq r\}.$$

If a point $y \notin S(x, r)$, then from the definition of strongly convexity and the Cauchy-Schwarz inequality we have following

$$f(y) \geq f(x) - ||f'(x)|| \cdot ||x - y|| + m||x - y||^2 =$$

$$= f(x) - mr||x - y|| + m||x - y||^2 =$$

$$= f(x) + m||x - y||(||x - y|| - r) > f(x).$$

Thus, at any point outside the ball $S(x, r)$ the value of the function f is greater than its value in the center of the ball.
Hence, the global minimum of the function f is attained at some point of the ball $S(x, r)$.
Let us prove the uniqueness of the minimizer of the function f on $\mathbb{R}^n$.
Assume the contrary, and suppose that there are two minimizers $x^*, \ y^* \in \mathbb{R}^n, x^* \neq y^*, \ f'(x^*) = f'(y^*)$.

Then, according to (11), we have

$$0 = \langle f'(x^*) - f'(y^*), x^* - y^* \rangle \geq 2m||x^* - y^*||^2 > 0.$$

The obtained contradiction proves our statement. The lemma is proved.
If a strongly convex function f is differentiable, then for minimizer x^* of f on $\mathbb{R}^n$ the following inequalities

$$f(x) \geq f(x^*) + m||x - x^*||^2 \quad \forall x \in \mathbb{R}^n,$$

$$\langle f'(x), x - x^* \rangle \geq 2m||x - x^*||^2 \quad \forall\, x \in \mathbb{R}^n,$$

$$||f'(x)|| \geq 2m||x - x^*|| \quad \forall\, x \in \mathbb{R}^n.$$

hold.

Lemma 8. If a differentiable function f is strongly convex (with constant $m > 0$) and x^* is a global minimizer, then for any $x \in \mathbb{R}^n$ the following inequality

$$||f'(x)||^2 \geq 4m(f(x) - f(x^*)). \tag{12}$$

holds.

Proof. As

$$\left|\left|\frac{f'(x)}{2\sqrt{m}} + \sqrt{m}(x^* - x)\right|\right|^2 \geq 0,$$

and

$$f(x) - f(x^*) + \langle f'(x), x^* - x\rangle + m||x^* - x||^2 \leq 0,$$

then

$$\frac{||f'(x)||^2}{4m} + \langle f'(x), x^* - x\rangle + m||x^* - x||^2 \geq 0 \geq f(x) - f(x^*) + \langle f'(x), x^* - x\rangle$$

From this we have (12). The Lemma is proved.

From (12) we imply the upper estimate of a possible decrease of the function f as a result of its minimization.
In many methods, it is necessary to minimize a function along a direction. Let $f : \mathbb{R}^n \to \mathbb{R}$ be a convex function. fix $x, g \in \mathbb{R}^n$. Define a function

$$h_{x,g}(\alpha) = f(x + \alpha g),\ \alpha \in \mathbb{R}.$$

Lemma 9. *In order for the function $f : \mathbb{R}^n \to \mathbb{R}$ is convex, necessary and sufficient that the function $h_{x,g} : \mathbb{R} \to \mathbb{R}$ is also convex for all x, g.*

Proof.

Let f be a convex function. Choose arbitrary points $\alpha_1, \alpha_2 \in \mathbb{R}$ and $\lambda_1, \lambda_2 \in [0,1]$, $\lambda_1 + \lambda_2 = 1$. Fix $x, g \in \mathbb{R}^n$. Then, using the convexity of the function f, we have

$$h_{x,g}(\lambda_1\alpha_1 + \lambda_2\alpha_2) =$$

$$= f(x + (\lambda_1\alpha_1 + \lambda_2\alpha_2)g) = f(\lambda_1(x + \alpha_1 g) + \lambda_2(x + \alpha_2 g)) \leq$$

$$\leq \lambda_1 f(x + \alpha_1 g) + \lambda_2 f(x + \alpha_2 g) = \lambda_1 h_{x,g}(\alpha_1) + \lambda_2 h_{x,g}(\alpha_2).$$

Therefore, the function $h_{x,g}$ is convex on $\mathbb{R}$.

Now let the function $h_{x,g}$ is convex on $\mathbb{R}$. Let us prove the convexity of f.
Choose $\lambda_1, \lambda_2 \in [0,1],\ \lambda_1 + \lambda_2 = 1$ and points $x_1, x_2 \in \mathbb{R}^n$.
We have

$$f(\lambda_1 x_1 + \lambda_2 x_2) = f(x_1 + \lambda_2(x_2 - x_1)) =$$

$$= h_{x_1, x_2 - x_1}(\lambda_2) = h_{x_1, x_2 - x_1}(\lambda_1 \cdot 0 + \lambda_2 \cdot 1) \leq$$

$$\leq \lambda_1 h_{x_1, x_2 - x_1}(0) + \lambda_2 h_{x_1, x_2 - x_1}(1) =$$

$$= \lambda_1 f(x_1 + 0 \cdot (x_2 - x_1)) + \lambda_2 f(x_1 + 1 \cdot (x_2 - x_1)) =$$

$$= \lambda_1 f(x_1) + \lambda_2 f(x_2).$$

The lemma is proved.