Constrained optimization

The Lagrange function

Numerical methods for constrained optimization

Creating of numerical methods for solving optimization problems with constraints is the more difficult problem than the building methods for unconstrained optimization.
Efficient algorithms manage to build only for special classes of conditional problems which include the problem of linear, quadratic and convex programming.
As before, methods for constrained minimization can be divided into zero, first and second order, depending from the order of derivatives used.

Furthermore, they can be divided into direct methods, which are constructed in the iteration space direct variables, and dual methods, in which the transition to the solution of the dual problem, and then - again a direct problem.
In many methods for constrained optimization at each step solve some auxiliary problems, for example, unconstrained optimization, the linear programming problem, the quadratic programming problem.
Consider important in the ideological and computational aspects of the method of constrained optimization.

The constrained optimization problem can be formulated as follows: find

$$\inf_{x \in X} f(x), \quad x \in X \subset \mathbb{R}^n. \tag{1}$$

In the derivation of the necessary conditions for an extremum in the presence of constraints large role played by the conditions imposed on the constraints ensure the existence of a nonempty the set X. They are called regularity conditions.

When solving problems of form (1) besides choosing the direction you must also ensure that the to remain inside the set or aspire to it, if the stationary point located at the boundary.

Thus the choice of the step-size is the extremely important task.

Classification of constrained optimization problems

View of the objective function and the set X in problem (1) affects the title of nonlinear programming problems.

There are problems of linear, quadratic, convex programming, and others.

Let the set X is given in the form

$$X = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} h_i(x) \leq 0, & i \in I = 1, \dots, k \\ t_j(x) = 0, & j \in J = 1, \dots, m \end{array} \right\}, \qquad (2)$$

If the objective function f is linear and h_i, $i \in I$, t_j, $j \in J$, are affine, then problem (1) is called the problem of linear programming.

In each linear programming problem variables are searched subject to these values satisfy of a certain system of linear equations or inequalities and for these values of the linear objective function is minimal or maximal.

In this case, the set X is called a polyhedral. One of the universal methods of solving this problem is the simplex method, which is used when the problem has a canonical form: find

$$\inf_{x \in X} \langle c, x \rangle,$$

$$X = \left\{ x = (x_1, \dots, x_n) \in \mathbb{R}^n \;\middle|\; \begin{matrix} Ax = b, \\ x \geq 0. \end{matrix} \right\},$$

where A is a matrix of order $(m \times n)$, $x \in \mathbb{R}^n$, $b \in \mathbb{R}^m$.

The inequality $x \geq 0$ means that all coordinates of the vector x are non-negative, i.e, $x_i \geq 0, \;\; i = 1, \ldots, n$.
Thus all constraints of the system consist of equations and inequalities expressing nonnegative variables. The objective function must be minimized.
However, in most applications (such as economic) often in the system of constraints initially arer included not only the linear equations but also inequalities.
It can be shown that any general linear programming problem can be reduced to the canonical form by introducing new (they are called complementary) variables.

The quadratic programming problem can be wrote as follows:

$$\inf_{x \in X} \frac{1}{2}\langle Cx, x\rangle + \langle d, x\rangle,$$

$$X = \left\{ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \;\middle|\; Ax \leq b \right\},$$

where C is an $(n \times n)$ real symmetric matrix, A is an $(m \times n)$ matrix,

$$x \in \mathbb{R}^n,\; b \in \mathbb{R}^m,\; d \in \mathbb{R}^n.$$

If C is a positive definite matrix, then the problem of quadratic programming is a problem of convex programming and in this case there exists an unique minimizer.

The problems of linear and quadratic programming are often arise in practice and at the same time they are used as auxiliaries in the numerical solution of problems of a more complex kind. Methods for their solution are well studied and can be applied in practice.
If the objective function f is convex and in the set X of form (2) the functions $h_i, \in I$, are convex, $t_j, j \in J$ are affine then such optimizations problem is called the problem of convex programming
.
In the general case we speak about a nonlinear programming problem. In such problems it is usually assumed the smoothness of the functions appearing in them.
Namely these problems will be in focus.

Necessary and sufficient conditions for a minimum of continuously differentiable functions on closed sets

Theorem 1.

Let f be continuously differentiable function on $\mathbb{R}^n$, a nonempty set X be convex and $x^* \in X$. In order to the point $x^* \in X$ is a minimizer of f on $\mathbb{R}^n$ necessary that the inequality

$$\langle f'(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X. \tag{3}$$

holds.

P r o o f. Let $x^* \in X$ be a solution of problem (1). Assume that inequality (3) is not satisfied, then there exists a point $\tilde{x} \in X$ and a number $a > 0$, for which

$$\langle f'(x^*), \tilde{x} - x^* \rangle = -a < 0.$$

Denote $g = \tilde{x} - x^*$. As $\tilde{x}$, $x^* \in X$ and the set X is convex then

$$x^* + \alpha g = x^* + \alpha(\tilde{x} - x^*) = \alpha\tilde{x} + (1 - \alpha)x^* \in X \quad \forall \alpha \in [0, 1].$$

Moreover

$$f(x^* + \alpha g) = f(x^*) + \alpha\langle f'(x^*), g\rangle + o(\alpha) = f(x^*) - \alpha a + o(\alpha).$$

As $\dfrac{o(\alpha)}{\alpha} \xrightarrow[\alpha\to 0]{} 0$, then there exists $\tilde{\alpha} > 0$, that

$$\frac{o(\alpha)}{\alpha} < \frac{a}{2} \quad \forall\alpha \in (0, \tilde{\alpha}].$$

From here we have

$$o(\alpha) < \frac{a}{2}\alpha \le \frac{a}{2}\tilde{\alpha} \quad \forall\alpha \in (0, \tilde{\alpha}],$$

and

$$f(x^* + \tilde{\alpha}g) \le f(x^*) - \tilde{\alpha}a + \frac{a}{2}\tilde{\alpha} = f(x^*) - \frac{a}{2}\tilde{\alpha} < f(x^*).$$

This contradiction proves the theorem.

Note that in this theorem it is proved a necessary condition for the local extremum in optimization problem (1).

It shows that the gradient of f at $x^* \in X$ amounts to not obtuse angle with any vector $x - x^*,\ x \in X$.

A point $x^* \in X$ in which inequality (3) holds is called a stationary point for f on X.

Corollary.

Let $x^* \in \text{int}X$. Then condition (3) is equivalent to the condition

$$f'(x^*) = 0_n.$$

Regularity conditions for the set X.

Necessary conditions for an extremum in nonlinear programming problems are correct only when the additional conditions are imposed on the set under which the optimization problem is considered. Let the set X has the form:

$$X = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} h_i(x) \leq 0, & i \in I = 1, \dots, k \\ t_j(x) = 0, & j \in J = 1, \dots, m \end{array} \right\}, \qquad (4)$$

where $h_i,\ i \in I,\ t_j,\ j \in J$, are continuously differentiable functions on $\mathbb{R}^n$.

Classical optimization problem is called the problem of minimizing a smooth function f subject to $x \in X$, where

$$X = \left\{ x \in \mathbb{R}^n \middle| \ t_j(x) = 0 \ , \quad j \in J = 1, \dots, m \right\}, \qquad (5)$$

where t_j, $j \in 1, \dots, m$, is continuously differentiable functions on $\mathbb{R}^n$.

Naturally, the study of this problem makes sense only in the case when the set X is not empty. The equations

$$t_j(x) = 0, \quad j \in J,$$

are called the equations of connection.

It can be shown that the regularity condition for set (5) is performed if gradients $t_j'(x)$, $j \in 1, \dots, m$ at the point x are linearly independent.

In problems with mixed constraints, i.e., when the set X is written in the form

$$X = \left\{ x \in \mathbb{R}^n \,\middle|\, \begin{array}{ll} h_i(x) \le 0, & i \in I = 1, \ldots, k \\ t_j(x) = 0, & j \in J = 1, \ldots, m \end{array} \right\}, \qquad (6)$$

where $h_i,\ i \in I,\ t_j,\ j \in J$, are continuously differentiable functions on $\mathbb{R}^n$ the condition of linearly independent of gradients

$$h_i'(x),\ i \in I(x),\ I(x) = \{ i \in I \mid h_i(x) = 0 \},\ t_j'(x),\ j \in J,$$

at $x \in X$ is sufficient for the existence in the neighborhood of x of the nontrivial set X written in form (5).

For a given set the regularity condition of Mangasarian - Fromovitz at a point $x \in X$ is the next:

at a point x gradients $t'_j(x), j \in J$, are linearly independent and there exists a vector $g \in \mathbb{R}^n$ that

$$\langle h'_i(x), g \rangle < 0 \quad \forall i \in I(x).$$

In convex programming as a condition of regularity widely used the Slater condition. Let the set X be given in the form

$$X = \left\{ x \in \mathbb{R}^n \;\middle|\; h_i(x) \le 0 \;, \quad i \in I \right\},$$

where $h_i, i \in I$, are finite convex functions on $\mathbb{R}^n$.
It is said that for a set X the Slater condition is performed if there exists a point $\overline{x} \in X$ for which an inequality $h_i(\overline{x}) < 0$ holds for each $i \in I$. This condition guarantees the existence of a non-empty interior of X.

In the problem with mixed constraints, i.e., when the set X is written in the form

$$X = \left\{x \in \mathbb{R}^n \;\middle|\; h_i(x) \leq 0,\ i \in I = 1, \dots, k,\ t_j(x) = 0,\ j \in J = 1, \dots, m\right.$$

where $h_i,\ i \in I$, are finite convex functions, $t_j,\ j \in J$, are affine functions on $\mathbb{R}^n$, the Slater condition is written as follows: there exists a point $\overline{x}$, for which

a) $h_i(\overline{x}) < 0,\ i \in I$;

б) gradients $t_j'(\overline{x}),\quad j \in J$, are linear independent.

Classical optimization problem

Necessary conditions for a local minimum problems with constraints are studied for a long time. One of the main methods is proposed by Lagrange. He proposed a method for solving a classical optimization problem.

Introduce the Lagrange function

$$L(x, \lambda) = \lambda_0 f(x) + \sum_{j=1}^{m} \lambda_j t_j(x),$$

where $x \in \mathbb{R}^n, \ \lambda = (\lambda_0, \lambda_1, \ldots, \lambda_m) \in \mathbb{R}^{m+1}$.

Finding the conditional extremum with using auxiliary functions and forms the basis of the method of Lagrange multipliers.

The coordinates of the vectors λ, i.e, numbers λ_j, $j \in J$, are called Lagrange multipliers or dual variables.

Theorem 1.

Let x^* be a local minimizer of f under constraints $t_j(x) = 0$, $j \in J$. Then there exists a nonzero vector $\lambda^* \in \mathbb{R}^{m+1}$, for which

$$\lambda_0^* f'(x^*) + \sum_{j=1}^{m} \lambda_j^* t_j'(x^*) = 0_n, \tag{7}$$

i.e., vectors f', t_j', $j \in J$, at x^* are linearly dependent.
The last equation can be written as a system

$$\lambda_0^* \frac{\partial f(x^*)}{\partial x_i} + \lambda_1^* \frac{\partial t_1(x^*)}{\partial x_i} + \ldots \lambda_m^* \frac{\partial t_m(x^*)}{\partial x_i} = 0, \quad i = 1, \ldots, n. \tag{8}$$

If the gradients $t_j'(x^*)$, $j \in J$, are linearly independent (the condition of regularity is performed) then $\lambda_0^* \neq 0$.

P r o o f. (By contradiction). Let x^* be a solution of the classical optimization problem but a nonzero vector λ^* does not exist. In this case gradients

$$f'(x^*),\ t_j'(x^*),\ j \in J,$$

are linear independent. Form the Jacobi matrix of system (8)

$$J(x^*) = \begin{pmatrix} \dfrac{\partial f(x^*)}{\partial x_1}, \dots, \dfrac{\partial f(x^*)}{\partial x_n} \\ \dfrac{\partial t_1(x^*)}{\partial x_1}, \dots, \dfrac{\partial t_1(x^*)}{\partial x_n} \\ \dots \\ \dfrac{\partial t_m(x^*)}{\partial x_1}, \dots, \dfrac{\partial t_m(x^*)}{\partial x_n} \end{pmatrix}.$$

The rank of $J(x^*)$ is equal to $m+1$. Suppose, for definiteness, that the first $m+1$ columns of this matrix form the nonzero minor of order $(m+1)$, i.e.,

$$\begin{vmatrix} \dfrac{\partial f(x^*)}{\partial x_1}, \dots, \dfrac{\partial f(x^*)}{\partial x_{m+1}} \\ \dfrac{\partial t_1(x^*)}{\partial x_1}, \dots, \dfrac{\partial t_1(x^*)}{\partial x_{m+1}} \\ \dots \\ \dfrac{\partial t_m(x^*)}{\partial x_1}, \dots, \dfrac{\partial t_m(x^*)}{\partial x_{m+1}} \end{vmatrix} \neq 0.$$

Let

$$y = (x_1, \dots, x_{m+1}) \in \mathbb{R}^{m+1},\ z = (x_{m+1}, \dots, x_n) \in \mathbb{R}^{n-m-1},$$

$$x = (y, z) \in \mathbb{R}^{m+1} \times \mathbb{R}^{n-m-1},\ x^* = (y^*, z^*) \in \mathbb{R}^{m+1} \times \mathbb{R}^{n-m-1}.$$

Denote $f(x^*) = f^*$ and consider the system of function

$$f(x) - f(x^*) - s, t_1(x), t_2(x), \dots, t_m(x)$$

at $x = x^*, s = 0$.

Since the Jacobian of this system is non-zero, then by using the theorem on implicit functions we have that the system

$$\begin{cases} f(x) - f(x^*) - s = 0, \\ t_1(x) = 0, \\ \dots \\ t_m(x) = 0, \end{cases}$$

has the solution for all s, $|s| \leq s_0$, where s_0 is sufficiently small.

Therefore there exists a vector-function

$$x(s) = (x_1(s), \ldots, x_n(s)),$$

which is defined and differentiable for all s, $|s| \leq |s_0|$, and

$$x(0) = 0, f(x(s)) = f(x^*) + s, t_j(x(s)) = 0 \quad \forall j \in J.$$

From this we have

$$f(x(s)) = f(x^*) + s > f(x^*) > f(x^*) - s = f(x(-s))$$

under $0 < s \leq s_0$.

This contradiction proves the theorem.

Sometimes this theorem is called the Lagrange multiplier rule.

A point $x^* \in X$, satisfying condition (7) is called a stationary point of f on the set X.

The Lagrange multiplier rule gives a necessary condition of the local minimum and therefore allows us to search stationary points.

An algorithm for solving the classical problem

To find the stationary point, it is necessary:

Step 1. Create a generalized Lagrange function

$$L(x,\lambda) = \lambda_0 f(x) + \sum_{j=1}^{m} \lambda_j t_j(x).$$

Step 2. Write the necessary optimality conditions of the first order, i.e., form the system r from $n+m$ equations

$$\begin{cases} \lambda_0 f'(x) + \sum\limits_{j=1}^{m} \lambda_j t_j'(x) = 0_n. \\ \lambda_j t_j(x) = 0,\ j \in J. \end{cases}, \tag{9}$$

(Equate to zero the partial derivatives of the Lagrangian $L(x, \lambda)$ by $x_i, \ i = 1, \dots, n$, and $\lambda_j, \ j \in J$).

Step 3. Solve the system for two cases:

1) $\lambda_0^* = 0$;

2) $\lambda_0^* \neq 0$ (in this case divide the first equation of the system on λ_0^* and replace $\frac{\lambda_j^*}{\lambda_0^*}$ on λ_j).

If the resulting system has a solution (x^*, λ^*), then x^* may be a conditional extremum i.e. the solution of the original problem.

Step 4. To establish the presence or absence of a conditional extremum at each stationary point x^*, you need to examine the sign of the second differential of the Lagrangian.

1) write the expression for the second differential of the classical Lagrangian function at the point (x^*, λ^*),

$$d^2L(x^*, \lambda^*) = \sum_{i,j}^{n} \frac{\partial^2 L(x^*, \lambda^*)}{\partial x_i \partial x_j} dx_i dx_j;$$

2) write the system, satisfying the equations of differentiated communication

$$\frac{\partial t_j(x^*)}{\partial x_1}dx_1 + \cdots + \frac{\partial t_j(x^*)}{\partial x_n}dx_n = 0, \quad j \in J;$$

3) from the previous system express any m differentials dx_i through the remaining $(n-m)$ and substitute them in $d^2L(x^*, \lambda^*)$;
4) if $d^2L(x^*, \lambda^*) > 0$ for the nonzero dx, then x^* is a point of conditional local minimum. If $d^2L(x^*, \lambda^*) < 0$ for the nonzero dx, then x^* is a point of conditional local maximum.

Step 5. Calculate the value of the function at the points of conditional extremum.

In solving practical problems in many cases the existence of the conditional extremum at a stationary point is determined by the existing task.

Remark.

Sometimes the following Lagrange function

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j t_j(x),$$

where $x \in \mathbb{R}^n, \ \lambda = (\lambda_1, \dots, \lambda_m) \in \mathbb{R}^m$. is considered.

Find the conditional extremum

$$\inf_{x \in X} x_1^2 + x_2^2,$$

$$X \in \{x = (x_1, x_2) \in X | \; t_1(x) = x_1 + x_2 - 2 = 0\}.$$

Check the condition of regularity. Since $t'(x_1, x_2) = (1, 1)$, then the condition of linear independence is done, so we can consider the classical Lagrange function.

1. Form the Lagrange function

$$L(x, \lambda_1) = x_1^2 + x_2^2 + \lambda_1(x_1 + x_2 - 2).$$

2. Write the necessary conditions of the extremum of the first order

$$\frac{\partial L(x, \lambda_1)}{\partial x_1} = 2x_1 + \lambda_1 = 0, \quad \frac{\partial L(x, \lambda_1)}{\partial x_2} = 2x_2 + \lambda_1 = 0,$$

From here we have

$$x_1 = -\frac{\lambda_1}{2}, \; x_2 = -\frac{\lambda_1}{2},$$

$$t_1(x) = x_1 + x_2 - 2 = 0.$$

3. Solve this system. Get a stationary point

$$x^* = (1,1),\ \lambda_1^* = -2.$$

4. Check sufficient conditions

$$a) \quad d^2L(x^*, \lambda_1^*) = 2dx_1^2 + 2dx_2^2,$$

as

$$\frac{\partial^2 L(x,\lambda_1)}{\partial x_1^2} = \frac{\partial^2 L(x,\lambda_1)}{\partial x_2^2} = 2, \quad \frac{\partial^2 L(x,\lambda_1)}{\partial x_1 \partial x_2} = 0,$$

$$b)\ dt_1(x) = dx_1 + dx_2 = 0.$$

c) Express the differential of dx_1 from dx_2: $dx_1 = -dx_2$ and substitute in d^2L.

d) As $d^2L(x^*, \lambda_1^*) = 4dx_2^2 > 0$ under $dx_2 \neq 0$, then at the point $x^* = (1, 1)$ we have a local minimizer.

Find

$$\inf_{x \in X} x_1$$

$$X = \{x = (x_1, x_2) \in \mathbb{R}^2 \mid t_1(x) = x_2^2 - x_1^3 = 0\}.$$

Check the condition of regularity $t_1'(x) = (-3x_1^2, 2x_2) = 0$ at the point $x^* = (0, 0)$, then the condition of linear independence is not done.

Therefore, we will use the generalized Lagrange function.

1. Form the generalized Lagrange function

$$L(x, \lambda_0, \lambda_1) = \lambda_0 x_1 + \lambda_1 (x_2^2 - x_1^3).$$

2. Write the necessary conditions of the extremum of the first order

$$\frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_1} = \lambda_0 - 3\lambda_1 x_1^2 = 0, \quad \frac{\partial L(x, \lambda_0, \lambda_1)}{\partial x_2} = 2\lambda_1 x_2 = 0,$$

$$t_1(x) = x_2^2 - x_1^3.$$

3. Solve the system for the two cases.

Case 1. $\lambda_0 = 0$. Then $\lambda_1 \neq 0$, since all the multipliers can not simultaneously be equal to zero. Here $x^* = (0, 0)$.

Case 2. $\lambda_0 \neq 0$. Divide the equation system on λ_0 and replace $\frac{\lambda_1}{\lambda_0}$ for λ_1. Then

$$1 - 3\lambda_1 x_1^2 = 0;\ 2\lambda_1 x_2 = 0,\ x_2^2 - x_1^3 = 0.$$

If $\lambda_1 = 0$, then the system is inconsistent. If $x_2 = 0$ then $x_1 = 0$ and then the system is also inconsistent.

The application of the classical Lagrange function does not work.

4. So $\lambda_0 = 0$, then sufficient conditions are not checked. The point $x^* = (0, 0)$ is a point of local irregular and the global minimum. Value of the objective function $f(x^*) = 0$

The Karush-Kuhn-Tucker conditions

Necessary conditions for a local minimum of constrained optimizations problem are studied for a long time. The Karush-Kuhn-Tucker (KKT) theorem is a generalization of the Lagrange theorem for optimization problems with constraints in the form of equalities and inequalities. In the literature sometimes it can be seen these theorems without mentioning Karush.

Karush in his unpublished master's thesis of 1939г found the necessary conditions in the general case, when conditions might contain equations and inequalities.

Independently there of to the same conclusion reached by Kuhn and Tucker.

They generalized the method of Lagrange multipliers (for using it in the construction of the criteria optimality for problems with equality constraints) to a generalг problem of nonlinear programming with constraints in the form equalities and inequalities.

Consider a minimization problem

$$\inf_{x \in X} f(x), \quad x \in X \subset \mathbb{R}^n, \tag{10}$$

where

$$X = \left\{ x \in \mathbb{R}^n \middle| \begin{array}{l} h_i(x) \leq 0, \; i \in I = 1, \ldots, k, \\ t_j(x) = 0, \; j \in J = 1, \ldots, m, \end{array} \right\},$$

and functions $f, \; h_i, \; i \in I, \; t_j, \; j \in J$, are continuously differentiable.

Define the generalized Lagrangian (the generalized Lagrange function)

$$L(x, \lambda_0, \lambda, \mu) = \lambda_0 f(x) + \sum_{i=1}^{k} \lambda_i h_i + \sum_{j=1}^{m} \mu_i t_j(x),$$

where $x \in \mathbb{R}^n$, $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^k$, $\mu = (\mu_1, \ldots, \mu_m) \in \mathbb{R}^m$. Numbers $\lambda_0, \lambda_1, \ldots, \lambda_k, \mu_1, \ldots, \mu_m$ are called Lagrange multipliers.

Denote

$$I(x) = \left\{ i \in I \;\middle|\; h_i(x) = 0 \right\}.$$

Inequality constraints with indices $i \in I(x)$ are called active at the point x, and the rest constraints are called inactive.

If it is possible to find the constraints that are inactive at a stationary point before solving the problem, then these constraints can be deleted from the model and thereby reduce its size.

Theorem 1.

Let x^* be a point of local minimum of problem (10). Then there exists a number $\lambda_0 \geq 0$ and numbers λ^*, μ^*, not all zero such $\lambda_i^* \geq 0$ if $i \in I$ and following conditions hold:

r - the condition for stationarity of the generalized Lagrange function for x:

$$\lambda_0^* f'(x^*) + \sum_{i=1}^{I(x^*)} \lambda_i^* h_i'(x^*) + \sum_{j=1}^{m} \mu_j^* t_j'(x^*) = 0_n; \qquad (11)$$

- condition of primal feasibility:

$$h_i(x^*) \leq 0 \quad \forall\ i \in I; \;\; t_j(x^*) = 0 \quad \forall\ j \in J; \qquad (12)$$

- condition of complementary slackness:

$$\lambda_i^* h_i(x^*) = 0 \quad \forall\ i \in I. \qquad (13)$$

A point $x^* \in X$, satisfying (11) is called a stationary point of problem (10).

A regularity condition in problem (10) is called any additional assumption under which in Theorem 1 $\lambda^* > 0$. It holds, for example, if the gradients of equality constraints and active inequality constraints at the point x^* are linearly independent.

Sometimes consider the classical Lagrange function.

Thus, the problem of Karush - Kuhn - Tucker is to find the vectors x^*, λ^*, μ^* satisfying the following conditions:

$$\begin{cases} f'(x^*) + \sum\limits_{i=1}^{k} \lambda_i^* h_i'(x^*) + \sum\limits_{j=1}^{m} \mu_j^* t_j'(x^*) = 0_n; \\ h_i(x^*) \leq 0 \quad \forall\, i \in I; \\ t_j(x^*) = 0 \quad \forall\, j \in J; \\ \lambda_i^* h_i(x^*) = 0 \quad \forall\, i \in I; \\ \lambda_i^* \geq 0 \quad \forall\, i \in I. \end{cases} \tag{14}$$

Theorem 2. Let x^* be a stationary point of (10) and (x^*, λ^*, μ^*) is solution of (14). Then the second differential of classical Lagrange function calculated at this point is non-negative

$$d^2L(x^*, \lambda^*, \mu^*) \geq 0,$$

for all $dx \in \mathbb{R}^n$ such that

$$dt_j(x^*) = 0, \;\; j \in J,$$

$$dh_i(x^*) = 0, \;\; \lambda_i^* > 0, \;\; i \in I,$$

$$dh_i(x^*) \leq 0, \;\; \lambda_i^* = 0, \;\; i \in I.$$

The algorithm of solving the problem (14)

Step 1. Create the generalized Lagrange function

$$L(x, \lambda, \mu) = f(x) + \sum_{i=1}^{k} \lambda_i h_i + \sum_{j=1}^{m} \mu_i t_j(x);$$

Step 2. write the necessary conditions for a minimum of the first order

$$\frac{\partial L(x^*, \lambda_0^*, \lambda^*, \mu^*)}{\partial x_i} = 0, \;\; i = 1, \dots, n;$$

$$h_i(x^*) \leq 0 \quad \forall\; i \in I; \;\; t_j(x^*) = 0 \quad \forall\; j \in J;$$

$$\lambda_i^* h_i(x^*) = 0 \quad \forall\; i \in I;$$

Step 3. solve the system for two cases: $\lambda_0^* = 0$; и $\lambda_0^* \neq 0$.
After solving of this system find a stationary point x^*;

Step 4. For found stationary points sufficient optimality conditions are checked.

To check the sufficient conditions it is necessary:

- determine the number of constraints - equalities and active inequality constraints;

- write the expression for the second differential of the classical Lagrangian point (x^*, λ^*, μ^*):

$$d^2 L(x^*, \lambda^*, \mu^*) = \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 L(x^*, \lambda^*, \mu^*)}{\partial x_i \partial x_j} dx_i dx_j;$$

- write conditions, imposed on the first differentials of equality constraints and active inequality constraints at the point x^*

$$dt_j(x^*) = \sum_{r=1}^{n} \frac{\partial t_j(x^*)}{\partial x_r} dx_r = 0, \; j \in J; \tag{15}$$

$$dh_i(x^*) = \sum_{r=1}^{n} \frac{\partial h_i(x^*)}{\partial x_r} dx_r = 0, \; \lambda_i > 0, \; i \in I(x^*); \tag{16}$$

$$dh_i(x^*) = \sum_{r=1}^{n} \frac{\partial h_i(x^*)}{\partial x_r} dx_r \leq 0, \; \lambda_i = 0, \; i \in I(x^*);$$

If the conditions are not satisfied, it is necessary to conduct additional studies.

- investigate the sign of the second differential of the Lagrangian for nonzero dx, satisfying (15) and (16).

If $d^2L(x^*, \lambda^*, \mu^*) > 0$ at x^* then x^* is a point of conditional local minimum. and

If $d^2L(x^*, \lambda^*, \mu^*) < 0$ at x^* then x^* is a point of conditional local maximum.

If the conditions are not satisfied, it is necessary to conduct additional studies.

Step 5. Calculate the value of the objective function at the extremum points.

Example 1. Consider the next constraints optimization problem: find

$$\min_{x \in X} f(x),$$

where

$$f(x) = (x_1 + 2)^2 + (x_2 - 2)^2, \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

$$X = \left\{ x \in \mathbb{R}^2 \;\middle|\; \begin{array}{l} h_1(x) = 2 - x_1 \le 0, \\ h_2(x) = -x_2 \le 0, \\ h_3(x) = x_1 + x_2 - 4 \le 0. \end{array} \right\}$$

Note that this problem is a problem of quadratic programming. The objective function is strongly convex, inequality constraints are affine. In our case, the set X is a triangle.
We have

$$L(x,\lambda_0,\lambda)=\lambda_0((x_1+2)^2+(x_2-2)^2)+\lambda_1(2-x_1)+$$

$$+\lambda_2(-x_2)+\lambda_3(x_1+x_2-4),$$

$$f'(x)=(2(x_1+2),2(x_2-2)),\ \ h_1'(x)=(-1,0),$$

$$h_2'(x)=(0,-1),\ \ h_3'(x)=(1,1).$$

Write the system (14) for this problem

$$
\begin{cases}
2\lambda_0(x_1+2)+\lambda_1\cdot(-1)+\lambda_2\cdot 0+\lambda_3\cdot 1=0,\\
2\lambda_0(x_2-2)+\lambda_1\cdot 0+\lambda_2\cdot(-1)+\lambda_3\cdot 1=0,\\
\lambda_1(2-x_1)=0.\\
\lambda_2(-x_2)=0,\\
\lambda_3(x_2+x_1-4)=0,\\
\lambda_1\geq 0,\\
\lambda_2\geq 0,\\
\lambda_3\geq 0.
\end{cases}
\tag{17}
$$

To find the stationary point is necessary to solve the system (17).

Solve the system for two cases:

1) Let $\lambda_0 = 0$, then

$$-\lambda_1 + \lambda_3 = 0, \ -\lambda_2 + \lambda_3 = 0.$$

From here we have

$$\lambda_1 = \lambda_2 = \lambda_3.$$

Then we can assume that

$$\lambda_1 = \lambda_2 = \lambda_3 = 1.$$

Under these assumptions we have

$$2 - x_1 = 0, \ x_2 = 0, \ x_1 + x_2 - 4 = 0.$$

This system of equations has no solution.

2) Let $\lambda_0 \neq 0$, then we can assume that $\lambda_0 = 1$.

Transform system (17). We have

$$\begin{cases} 2x_1 + 4 - \lambda_1 + \lambda_3 = 0, \\ 2x_2 - 4 - \lambda_2 + \lambda_3 = 0, \\ 2\lambda_1 - \lambda_1 x_1 = 0. \\ -\lambda_2 x_2 = 0, \\ \lambda_3(x_2 + x_1 - 4) = 0, \\ \lambda_1 \geq 0, \\ \lambda_2 \geq 0, \\ \lambda_3 \geq 0. \end{cases} \tag{18}$$

It can be seen that the point $x^* = (2, 2)$ is a stationary point of this problem. In this case,

$$f'(x^*) = (8, 0), \ \lambda^* = (8, 0, 0).$$

In addition, at x^* the condition (3) hold. In fact,

$$\langle f'(x^*), x - x^* \rangle = 8(x_1 - 2) + 0 \cdot (x_2 - 2) \geq 0 \quad \forall x \in X.$$

Example 2. Maximize

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

subject to the constraint

$$x_1^2 + x_2^2 = 1.$$

The feasible set is the unit circle, and the level sets of f are diagonal lines (with slope -1), so we can see graphically that the maximum occurs at $\left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$, and that the minimum occurs at $\left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right)$.

Using the method of Lagrange multipliers, we have

$$g(x) - c = x_1^2 + x_2^2 - 1,$$

hence

$$L(x, \lambda) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 1).$$

Setting the gradient

$$L'(x, \lambda) = 0$$

yields the system of equations

$$\frac{\partial L}{\partial x_1} = 1 + 2\lambda x = 0,$$
$$\frac{\partial L}{\partial x_2} = 1 + 2\lambda y = 0,$$
$$\frac{\partial L}{\partial \lambda} = x_1^2 + x_2^2 - 1 = 0,$$

where the last equation is the original constraint.

The first two equations yield

$$x_1 = x_2 = -\frac{1}{2\lambda}, \qquad \lambda \neq 0.$$

Substituting into the last equation yields

$$\frac{1}{4\lambda^2} + \frac{1}{4\lambda^2} = 1,$$

so

$$\lambda = \mp\frac{1}{\sqrt{2}},$$

which implies that the stationary points are

$$x^* = \left(\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right), \quad x^{**} = \left(-\frac{\sqrt{2}}{2}, -\frac{\sqrt{2}}{2}\right).$$

Evaluating the objective function f at these points yields

$$f(x^*) = \sqrt{2} \text{ and } f(x^{**}) = -\sqrt{2},$$

thus the maximum is $\sqrt{2}$, which is attained at x^*, and the minimum is $-\sqrt{2}$, which is attained at x^{**}.

Consider the convex programming problem

$$\inf_{x \in X} f(x), \quad x \in X \subset \mathbb{R}^n. \tag{19}$$

where

$$X = \{x \in \mathbb{R}^n \mid h_i(x) \leq 0,\ i \in I = \{1, \dots, k\}\}.$$

functions f and $h_i,\ i \in I$, are convex continuous differentiable on $\mathbb{R}^n$.

Suppose that the Slater condition is satisfied for the set X:
- there exists a point $\hat{x} \in X$ for which the valid inequality

$$h_i(\hat{x}) < 0 \quad \forall\, i \in I.$$

The Slater condition can be replaced by the condition:
- there exist points $\hat{x}_i,\ i \in I$ for which the valid inequality

$$h_i(\hat{x}_i) < 0 \quad \forall\, i \in I.$$

Here the set $R(x)$ is the set of active constraints at the point x:

$$R(x) = \{i \in I \mid h_i(x) = 0\}.$$

Theorem 3. *If the point $x^* \in X$ is the local minimum of problem (19), the gradients of the active constraints $h_i'(x^*),\ i \in R(x^*)$, are linearly independent at the point x^*, then there are numbers $\lambda_i \geq 0,\ i \in I$, for which*

$$-f'(x^*) = \sum_{i=1}^{k} \lambda_i h_i'(x^*),$$

$$\lambda_i h_i(x^*) = 0, \quad i \in I.$$

This condition expresses that at the point of optimum the gradient of the objective function is represented as a linear combination of the gradients of the constraints and all the coefficients of this linear combination non-positive.

Proof.
Let $x^* \in X$ be a minimizer of f on X. From Theorem 1 it follows that there exists a number λ_0^* and nonnegative numbers $\lambda_1^*, \ldots, \lambda_k^* \geq 0$, for which the equality

$$\lambda_0^* f'(x^*) + \sum_{i=1}^{k} \lambda_i^* h_i'(x^*) = 0$$

is true.

Prove that under our assumptions on the set X the multiplier λ_0^* does not equal zero. Suppose to the contrary. Let $\lambda_0^* = 0$. Then

$$\sum_{i \in R(x^*)} \lambda_i^* h_i'(x^*) = 0.$$

But the last equality contradicts the fact that the vectors $h_i'(x^*),\ i \in R(x^*)$, are linearly independent at the point x^*. Therefore, $\lambda_0^* \neq 0$.

Prove that $\lambda_0^* > 0$. From the properties of convex functions have

$$0 > h_i(\hat{x}) \geq h_i(x^*) + \langle h_i'(x^*), \hat{x} - x^* \rangle = \langle h_i'(x^*), \hat{x} - x^* \rangle \quad \forall\, i \in R(x^*).$$

From here we have

$$\langle h_i'(x^*), \hat{x} - x^* \rangle < 0 \quad \forall\, i \in R(x^*).$$

Multiplying each inequality on the corresponding the non-negative multiplier $\lambda_i,\ i \in R(x^*)$, and summing them, we receive inequalities

$$0 > \left\langle \sum_{i \in R(x^*)} h_i'(x^*), \hat{x} - x^* \right\rangle = \langle -\lambda_0 f'(x^*), \hat{x} - x^* \rangle =$$

$$= -\lambda_0 \langle f'(x^*), \hat{x} - x^* \rangle.$$

As x^* is a minimizer of f on X, then

$$\langle f'(x^*), x - x^* \rangle \geq 0 \quad \forall x \in X.$$

Hence $\lambda_0 > 0$.

The theorem is proved.