А.Ш. Кажикенова (доцент), Д.Б. Алибиев (доцент),
Р.А. Оразбекова (ст.преподаватель), А.С. Смаилова (преподаватель),
М.Ж. Жақсылық (М-21-1)
Карагандинский университет им. Е.А. Букетова

Numerical methods of solving problems on unconditional extremum

Method of gradient descent. The method of steepest descent. Statement of the task of minimizing the function n variables $J(u) = J(u_1, u_2, ..., u_n)$ on the set U on the set the space E^n does not differ from the statement in the one-dimensional case. If $U = E^n$ then it's said of the *unconditional minimization* of the function $J(u)$.

To solve the problems of unconditional minimization of the function $J(u)$ widely used approximate methods, which are based on the calculation of first-order derivatives of the function $J(x)$. Such methods are usually called *gradient.* Consider one of them - the *method of gradient descent*.

Let $J(u)$ - convex differentiable in the whole space E^n function, it is required to find its minimum pointa u^*.

We describe the method of gradient descent. The method assumes the choice of the initial approximation - some point $u^{(0)} \in E^n$. General rules for selecting a point $u^{(0)}$ in the gradient descent method is not present. In those cases when from geometrical, physical and other considerations there is information about the location of the minimum point, then the initial approximation $u^{(0)}$ try to choose closer to this area. We will assume that some initial point $u^{(0)}$ already choosen. Then the method of gradient descent consists in constructing a sequence $\{u^{(k)}\}$ by rule

$$u^{(k+1)} = u^{(k)} - \alpha^{(k)} J'(u^{(k)}),\ \alpha^{(k)} > 0,\ k = 0,1,2,.... \qquad (1)$$

Number $\alpha^{(k)}$ из (1) called the *step length* or *the gradient descent method step.*

If $J'(u^{(k)}) \neq 0$, $\alpha^{(k)} > 0$ can be chosen so that $J(u^{(k+1)}) < J(u^{(k)})$.

If $J'(u^{(k)}) = 0$, то $u^{(k)}$ - stationary point. In this case, process (1) ceases, and if necessary, additional investigation of the behavior of the function in the

Created by FolderMill | www.foldermill.com

neighborhood of point $u^{(k)}$ to determine whether the point $u^{(k)}$ minimum of function $J(u)$ achieved or not achieved. In particular, if $J(u)$ - convex function, then, according to the optimality criterion, a minimum is always attained at a stationary point. There are different ways of choosing $\alpha^{(k)}$ in the method (1). Depending on the selection method $\alpha^{(k)}$ It is possible to obtain various variants of the method of gradient descent. We indicate the most frequently used methods of choice $\alpha^{(k)}$.

1) Method with step division. In practice, a sufficiently small $\alpha^{(k)} > 0$ often chosen so that the monotonicity condition is satisfied $J(u^{(k+1)}) < J(u^{(k)})$, $k = 0,1,2,.....$ If this condition is violated, the step $\alpha^{(k)}$ crushed until the monotony is restored. Then proceed to calculate the next iteration.

2) The method of steepest descent. on the ray

$$\{u \in E^n : u = u^{(k)} - \alpha J'(u^{(k)}),\ \alpha \geq 0\},$$

directed along the antiprodient, we introduce the function of one variable

$$f_k(\alpha) = J(u^{(k)} - \alpha J'(u^{(k)})),\ \alpha > 0$$

and define $\alpha^{(k)}$ from the conditions

$$f_k(\alpha^{(k)}) = \inf_{\alpha > 0} f_k(\alpha) = f_k^*,\ \alpha^{(k)} > 0. \qquad (2)$$

Method (1), (2) is called the *method of steepest descent*. For determining $\alpha^{(k)}$ at each step a one-dimensional minimization problem (2) is solved, for which it is possible to use the methods we considered earlier. Let us give an example when the quantity $\alpha^{(k)}$, defined by condition (2), exists and can be written out in explicit form.

Example. Let there be given a quadratic function

$$J(u) = \frac{1}{2} < Au,\ u > - < b,\ u >,\ u \in E^n, \qquad (3)$$

where A – symmetric non-negative definite matrix of $n \times n$ order; b, u – vectors from E^n, $< b,\ u >$ – scalar product of E^n.

$J(u)$ - convex function, and its derivatives are calculated by formulas

$$J'(u) = Au - b,\ J''(u) = A.$$

Therefore, method (1) in this case will look like this:

Created by FolderMill | www.foldermill.com

$$u^{(k+1)} = u^{(k)} - \alpha^{(k)}(Au^{(k)} - b), \alpha^{(k)} > 0, k = 0,1,2,....$$

Thus, the method of gradient descent for function (3) is a well-known iterative method for solving a system of linear algebraic equations $Au = b$.

$\alpha^{(k)}$ is determined by the formula

$$\alpha^{(k)} = \frac{|Au^{(k)} - b|^2}{< A(Au^{(k)} - b), Au^{(k)} - b >} > 0. \qquad (4)$$

Thus, the method of steepest descent for a quadratic function (3) consists in constructing a sequence of points $\{u^{(k)}\}$ by formulas (1), (4).

3) <u>The method with the Lipschitz constant.</u> Let the function $J(u) \in C^{1,1}(E^n)$, i.e. function $J(u) \in C^1(E^n)$ and gradient $J'(u)$ satisfies the Lipschitz condition:

$$|J'(u) - J'(v)| \le L|u - v|$$

for any points $u, v \in E^n$, and a constant $L > 0$ is called the *Lipschitz constant.*

If the constant L is known, then in (1) as $\alpha^{(k)}$ any number satisfying conditions

$$0 < \delta_0 \le \alpha^{(k)} \le \frac{2}{L + 2\delta},$$

where δ_0, δ - positive numbers that are parameters of the method.

In particular, when $\delta = \frac{L}{2}$, $\delta_0 = \frac{1}{L}$ we obtain method (1) with a constant step $\alpha^{(k)} = \frac{1}{L}$. Hence it is clear that if L - a large value or obtained using rough estimates, then step $\alpha^{(k)}$ will be small, and so the method will slowly converge to the minimum point of the given function.

As a condition for gradient descent, the gradient is usually close to zero $J'(u^{(k)})$, i.e. execution the inequalities

$$\left|\frac{\partial J(u^{(k)})}{\partial u_i}\right| \le \varepsilon, \ i = 1,2,...,n$$

or

$$\| J'(u^{(k)}) \| = \sqrt{\sum_{i=1}^{n} \left[\frac{\partial J(u^{(k)})}{\partial u_i}\right]} \le \varepsilon, \qquad (5)$$

Created by FolderMill | www.foldermill.com

where ε - given sufficiently small number. If at some step k one of these inequalities is satisfied, then it is assumed that the approximate equalities $u^* \approx u^{(k)}$, $J^* \approx J(u^{(k)})$. Then we say that the minimum point of a given function is found with an accuracy ε.

The conjugate direction method. Consider the task

$$f(x) \to \inf; \quad x \in U \equiv E^n,$$

where function $f(x) \in C^1(E^n)$.

The conjugate direction method consists in constructing successive approximations $x^{(k)}$ to the minimum of the function $f(x)$ in the following way:

$$x^{(k+1)} = x^{(k)} - \alpha_k p^{(k)}, \; k = 0,1,..., \; x^{(0)} \in E^n, \tag{6}$$

where $x^{(0)}$ - preselected initial approximation, step α_k is chosen from condition

$$\Phi_k(\alpha_k) = \min_{\alpha>0} \Phi_k(\alpha),$$

where $\Phi_k(\alpha) = f[x^{(k)} - \alpha f'(x^{(k)})]$.

That is, step α_k is chosen as in the method of steepest descent. And the direction of descent - $p^{(k)}$ is determined by the formula

$$p^{(k)} = f'(x^{(k)}) + \beta_k p^{(k-1)}, \; k = 0,1,..., \; p^{(0)} = f'(x^{(0)}),$$

where

$$\beta_k = \frac{\| f'(x^{(k)}) \|^2}{\| f'(x^{(k-1)}) \|^2} = \frac{\sum_{i=1}^{n} \left(\frac{\partial f(x^{(k)})}{\partial x_i} \right)^2}{\sum_{i=1}^{n} \left(\frac{\partial f(x^{(k-1)})}{\partial x_i} \right)^2}. \tag{7}$$

Thus, the method of conjugate directions differs from the method of steepest descent only by choosing the direction of decreasing the function at each step ($-p^{(k)}$ instead of $-f'(x^{(k)})$).

Note that $p^{(k)}$ из (7) is determined not only by the antigradient $-f'(x^{(k)})$, but also by the direction of descent $-p^{(k-1)}$ in the previous step. This allows more fully

Created by FolderMill | www.foldermill.com

than in gradient methods to take into account the features of the function when constructing successive approximations (6) to its minimum point.

The criterion for achieving a given accuracy of calculations in the method of conjugate directions is usually the inequalities

$$\left|\frac{\partial f(x^{(k)})}{\partial x_i}\right| \le \varepsilon, \; i = 1,2,\ldots,n$$

or

$$\| f'(x^{(k)}) \| = \sqrt{\sum_{i=1}^{n}\left[\frac{\partial f(x^{(k)})}{\partial x_i}\right]} \le \varepsilon$$

Often to reduce the impact of accumulating errors in computing every N iterations (6) assume $\beta_{mN} = 0, \; m = 0,1,\ldots,$ i.e. update method (N - algorithm parameter).

To minimize the convex quadratic function in E^n requires no more than n iterations of the conjugate direction method.

Created by FolderMill | www.foldermill.com