AN ORTHOGONAL DESCENT ALGORITHM TO FIND THE
ZERO OF A CONVEX FUNCTION, UNSOLVABILITY TEST,
AND RATE OF CONVERGENCE

M. B. Shchepakin

The paper is the first in a series of articles on the orthogonal descent method. The unsolvability test is considered and the rate of convergence is determined for the algorithm that finds a point with a prescribed value of a convex function.

The main idea of the orthogonal descent method was proposed in 1977 [1]. The method takes its name from the fact that under certain conditions the successive steps are made in orthogonal directions, which are formed by orthogonalization of the antigradient (antisubgradient [2]) of the objective function of the mathematical programming problem [3]. The orthogonal descent method is a Fejér process, whose main properties are the following:

\[ x_{k+1} = x_k + h_k, \quad \| y^* - x_{k+1} \| < \| y^* - x_k \| \quad \forall y \in M, \lim_{k \to \infty} \| x^* - x_k \| = 0, x^* \in M. \]  (*)

The vectors \( h_k \) are computed by the rules of the specific algorithm with the properties (*), \( M \) is the solution set of the mathematical programming problem, and \( \{ x_k \} \) is the sequence of points generated by the algorithm.

The study of methods with these properties was stimulated by the work of Polyak [4] on subgradient minimization of smooth convex functions and also by the work of Gurin, Polyak, and Raik [5] on the projection method for finding the common point of a system of convex sets.

The orthogonal descent method has proved to be quite universal: it can find a point where a convex, not necessarily differentiable, function takes a prescribed value; find the common point of a system of convex sets; solve an unconstrained optimization problem and problems with constraints on the variables; and solve sequential optimization problems. It is also applicable to linear, nonlinear, smooth, and nonsmooth problems, to finding local extrema of multiextremum problems, and also as a convergence-accelerating procedure in nonconvex optimization algorithms that satisfy the properties (*).

The orthogonal descent method is theoretically sound; its properties have been studied and rate of convergence bounds have been obtained [6]. Numerical results and tests have established the reliability and efficiency of orthogonal descent algorithms [7]. The method is effective for solving ill-conditioned problems. However, ill-conditioning often is not inherent to the problem: it is merely the result of inappropriate reduction of the original problem to a form convenient for the application of some method. Standard operations, such as feasible set regularization, application of penalty and barrier functions, smoothing of nondifferentiable functions, and the use of Lagrange multipliers may adversely affect the conditioning of the problem, producing a negative impact on solution accuracy, computing time, and stability of the numerical solutions. Moreover, the solution procedure of a particular problem often expands into a special investigation that requires fairly extensive knowledge of mathematical programming. Another advantage of the orthogonal descent method is therefore its ability to solve any convex programming problem in its original form, without requiring any transformation to a special form suitable for the application of the method. The user has to perform a minimum preparatory work, which is actually needed when solving the problem by any method. This preparatory work includes constructing a program to evaluate the optimand function and its gradient (subgradient) at a point.

Let the function \( \varphi(x) \) be given in \( R^n \). Denote by \( \partial \varphi(x) \) the gradient of this function if \( \varphi(x) \) is differentiable; if \( \varphi(x) \) is nondifferentiable, then \( \partial \varphi(x) \) denotes its subgradient [2]. By \( M \) we denote the set of points where \( \varphi(x) = \varphi^* \).
LEMMA 1. Let $M \neq \emptyset$ and $\forall x: \varphi(x) > \varphi^*$,

$$\sigma(x) = - \frac{\partial \varphi(x) (\varphi(x) - \varphi^*)}{\|\partial \varphi(x)\|^2};$$

(1)
denote $y = x + v(x)$. If

$$(x^* - y, y - x) \geq 0 \ \forall x \in M,$$

(2)
then $\varphi(x)$ is a convex function.

Proof. Rewrite inequality (2) using formula (1):

$$\left( (x^* - x + \frac{\partial \varphi(x)(\varphi(x) - \varphi^*)}{\|\partial \varphi(x)\|^2}, (x - \frac{\partial \varphi(x)(\varphi(x) - \varphi^*)}{\|\partial \varphi(x)\|^2} - x) \right) \geq 0,$$

$$- (x^* - x, \partial \varphi(x))(\varphi(x) - \varphi^*) - (\varphi(x) - \varphi^*)^2 \geq 0,$$

(3)

Inequality (3) implies that $\varphi(x)$ is a convex function [3]. We can also show that (3) implies inequality (2).

Let us now consider the subgradient method of finding a point where $\varphi(x) \leq \varphi^*$ that follows from property (2). For simplicity we denote $f(x) = \varphi(x) - \varphi^*$ and consider the following equivalent problem: find a point where

$$f(x) \leq 0.$$

(4)

Denote by $M$ the solution set of the problem (4). We know from convex analysis that $M$ is a convex set. From the definition of convexity (3) we see that if $M \neq \emptyset$, then

$$\forall x^* \in M \ni x \notin M: (-\partial f(x), x^* - x) > 0.$$

(5)
Denote

$$v_k = - \frac{\partial f(x_k) f(x_k)}{\|\partial f(x_k)\|^2},$$

$$x_{k+1} = x_k + v_k,$$

(6)
(7)
then the sequence $\{x_k\}$ has the following properties:

$$(x^* - x_k, x_{k+1} - x_k) > 0 \ \forall x^* \in M,$$

(8)
i.e., the vector $v_k$ always makes an acute angle with any point of the set $M$,

$$(x^* - x_{k+1}, x_{k+1} - x_k) \geq 0 \ \forall x^* \in M$$

(9)
i.e., the point $x_{k+1}$ and the vector $v_k$ span the half-space $(x - x_{k+1}, v_k) \geq 0$ that contains the set $M$ (Fig. 1).

Polyak [3, p. 131] has proved a theorem, which can be stated in the following form for the problem (4) and the method (7).

THEOREM 1. If $M \neq \emptyset$, then the sequence of points $\{x_k\}$ satisfies the inequality

$$\|x^* - x_{k+1}\| < \|x^* - x_k\|$$

$$\forall x^* \in M,$$

(10)
$x_k \rightarrow x^* \in M$, the rate of convergence for an arbitrary convex function $f(x)$ has the bound

$$\lim_{k \rightarrow \infty} k^{1/2} f(x_k) = 0,$$

(11)
and for functions with a sharp minimum, i.e.,

$$\forall x^* \in M: f(x) \geq \lambda \|x^* - x\|,$$

$$\lambda > 0,$$