Control
and Cybernetics
VOL. 9 (1980) No. 4

Stochastic feasible direction methods for nonsmooth stochastic optimization problems

by

ANDRZEJ RUSZCZYŃSKI

Instytut Automatyki
Politechnika Warszawska

In the paper stochastic feasible direction methods for nonsmooth and stochastic optimization problems are considered. A point-to-set map model of the algorithm is given and convergence is proved under general assumptions on mappings describing the method. Finally, the general theory is used to develop stochastic nonsmoth functions oriented analogues of some classical feasible direction algorithms.

1. Introduction

In this paper we discuss stochastic feasible direction algorithms for the solution of the problem

(P) $$\min_{x \in X} F(x)$$

$$X = \{x \in R^n : g_i(x) \leqslant 0 \quad \text{for} \quad i = 1, \ldots, m\},$$

where $g_1, \ldots, g_m$ and F are real-valued functions defined on R^n. It is assumed througout this paper that the set X is compact, the function F is Lipschitz continuous on an open set $\mathscr{X}$ containing X and the functions $g_1, \ldots, g_m$ are continuously differentiable. In particular, we are interested in the case, when F is a nonsmooth function, possibly defined by

$$F(x) = Ef(x, \theta),$$

where $f: R^n \times \Theta \dashrightarrow R^1$ and $(\Theta, \mathscr{B}, P)$ is a probability space. The random event θ is used here to represent all stochastic factors, i.e. $f(x, \theta_k)$ denotes the noise-corrupted value of $F(x)$, connected with an event θ_k.

Problems of this kind arise in various fields of technology and management, such as stochastic optimal control, hierarchical control, long-term planning, etc. (see e.g. [5], [16], [19]). We mention here two typical examples.

The stochastic minimax problem

Let $\varphi_0: R^n \times R^l \times \Theta \dashrightarrow R^1$, $Y \subset R^l$ and let F in (P) be defined by

$$F(x) = E \max_{y \in Y} \varphi_0(x, y, \theta).$$

The two-stage stochastic programming problem

Let $\varphi_j: R^n \times R^l \times \Theta \to R^1$, $j=0, 1, ..., k$, and let $Y(x, \theta)=\{y \in R^l: \varphi_j(x, y, \theta) \leqslant 0$, $=1, ..., k\}$. We define F in (P) by

$$F(x)=E \min_{y \in Y(x, \theta)} \varphi_0(x, y, \theta).$$

In both cases F is in general a nonsmooth function, even if the functions $\varphi_j(., y, \theta)$ are differentiable.

When dealing with problems under consideration, we meet two basic difficulties. The first one is the nonsmoothness of the objective function. Even in the deterministic case it causes serious difficulties for both theory and computations (see e.g. [3], [4], [13], [14]). The second difficulty is connected with the stochastic nature of the problem. It can hardly be assumed that it is possible to compute the value of F at a given x; we can only observe noise-corrupted values $f(x, \theta^i)$, where θ^i is a draw of θ. This feature makes it necessary to use for the solution of (P) stochastic approximation type algorithms (see e.g. [5], [6], [8], [11], [18]).

The principal objective of this paper is to develop a unified framework for the construction of stochastic feasible direction algorithms for (P). These algorithms construct stochastic vectors, which correspond in a certain sense to generalized gradients of the objective function. The vectors are used in direction-finding subproblems which produce random search directions. We note here that some algorithms from this class were suggested in [1], [12] for smooth problems, and in [8] an algorithm for nonsmooth problems has been proposed. In this paper we shall give an abstract point-to-set map model of the algorithm, similar to those employed in [9], [7], [22] for deterministic algorithms. For this model we shall formulate general assumptions, under which the algorithm converges with probability one to the set of stationary points of (P). It will be shown that the general framework makes it possible to construct stochastic analogues of many classical feasible direction algorithms. In this sense the present paper extends the approach developed in [20] to the class of nonsmooth problems. The problem of the nonsmoothness of the objective function will be overcome by the generation of a sequence $\{F_k\}$ of smooth functions convergent uniformly to F. The functions F_k, constructed by means of integral transformations of F, are used in the general model only and they do not appear in definite algorithms. This idea was employed for nonsmooth optimization in [8].

In § 2 we review briefly important properties of Lipschitz continuous functions and we remind necessary optimality conditions for (P). In § 3 we describe the general structure of the algorithms under consideration and we make all relevant assumptions. §§ 4 and 5 are devoted to the convergence analysis. In § 6 we show practical methods for the construction of random vectors corresponding to the generalized gradients of the objective function (in the sense of the assumptions from § 3). Finally, in § 7 we discuss some definite direction-finding subproblems taken from deterministic nonlinear programming and we prove (on the base of the general theory from the previous sections) the convergence of the resulting stochastic feasible direction algorithms. It should be stressed that the algorithms considered here are adapted to off-line computations wite random effects simulated in the computer.

In the paper we use $\|\cdot\|$ to denote the Euclidean norm in R^n and $\|\cdot\|_*$ denotes an arbitrary norm in R^n. We denote by $U_\delta(x)$ the δ-neighborhood of x, i.e. $U_\delta(x)=\{y\in R^n: \|y-x\|\leqslant\delta\}$. If $V\subset R^n$ then $U_\delta(V)=\{y\in R^n: \inf\|y-v\|\leqslant\delta,\ v\in V\}$. For a closed convex set $Z\subset R^n$ we denote by Π_Z the orthogonal projection on Z, i.e. $y=\Pi_Z(x)$ if $y\in Z$ and $\|y-x\|=\min\|_z-x\|$, $z\in Z$. If $X\subset R^n$ then $\mathscr{P}(X)$ denotes the set of all subsets of X. We denote by $(\Omega, \mathscr{F}, P)$ a probability space and we use ω to denote a single element of Ω (a sample point). The event ω corresponds to a single run of the algorithm, and so Ω is the space of all possible sequences of simulation results necessary to geneiate one path $\{x^k\}$. We use $E\xi$ to denote the mathematical expectation of a random variable $\xi: \Omega\to R^n$ and $E\{\xi/\mathscr{F}_t\}$ denotes the conditional expectation with respect to the σ-field $\mathscr{F}_t\subset\mathscr{F}$. We use abbreviation "wp 1" for "with probability one".

2. Preliminaries.

In this section we shall recall briefly some important properties of Lipschitz continuous functions. Let $\mathscr{X}\subset R^n$ be open and let $F: R^n\to R^1$ be *Lipschitz* on $\mathscr{X}$, i.e. there exists a constant L such that

$$|F(x_1)-F(x_2)|\leqslant L\|x_1-x_2\| \quad \text{for all} \quad x_1, x_2\in\mathscr{X}. \tag{2.1}$$

Let $x\in\mathscr{X}$, $d\in R^n$ and let

$$F^0(x; d)=\overline{\lim_{\substack{h\to 0\\ \tau\downarrow 0}}}\,[F(x+h+\tau d)-F(x+h)]/\tau. \tag{2.2}$$

The *generalized gradient* of F at x is defined by

$$\partial F(x)=\{v\in R^n: \langle v, d\rangle\leqslant F^0(x; d) \quad \text{for all} \quad d\in R^n\}. \tag{2.3}$$

The following proposition collects together important properties of F, F^0 and ∂F from [3].

PROPOSITION 1.

(a) $\partial F(x)$ is nonempty, convex and compact.
(b) $F^0(x; d)=\max_{v\in\partial F(x)}\langle v, d\rangle$.
(c) F is differentiable almost everywhere in $\mathscr{X}$ and $\partial F(x)$ is the convex hull of all the points v of the form

$$v=\lim_{i\to\infty}\nabla F(x^i)$$

where $\{\mathscr{X}^i\}\to x$ and F has a gradient ∇F at each $\mathscr{X}^i\in\mathscr{X}$.
(d) ∂F is bounded on bounded subsets of $\mathscr{X}$ and the mapping $x\to\partial F(x)$ is upper-semicontinuous (closed) on $\mathscr{X}$.

Let us now consider the problem (P) with g_i $(i=1, \ldots, m)$ differentiable and F Lipschitz on an open set $\mathscr{X}$ containing X. Proposition 2 to follow is a simple consequence of general theorems from [4], [14].

PROPOSITION 2. If $\bar{x}$ is optimal for (P) then there exist $v \in \partial F(\bar{x})$ and $u_i \geqslant 0$ ($i=0, 1, \ldots, m$) such that

$$u_0 v + \sum_{i=1}^{m} u_i \nabla g_i(\bar{x}) = 0$$

$$u_i g_i(\bar{x}) = 0, \quad i = 1, 2, \ldots, m,$$

$$\sum_{i=1}^{m} u_i = 1.$$

A point x which satisfies the above necessary conditions of optimality will be called *stationary*.

3. General feasible direction algorithm

In this section we give a general point-to-set map description of stochastic feasible direction algorithms for the solution of (P) and we formulate sufficient conditions for the convergence with probability one to an abstract solution set X^*.

Let $Z \subset R^n$ be a closed convex set. Let the sequences $\{x^k\} \subset X$, $\{d^k\}$, $\{z^k\} \subset Z$ of R^n-valued random variables be defined by the relations

$$x^{k+1} = x^k + \tau_k d^k, \quad x^0 \in X, \tag{3.1}$$

$$d^k \in D_k(x^k, z^k), \tag{3.2}$$

$$z^{k+1} = \Pi_Z\big(z^k + \rho_k(\xi^k - z^k)\big), \quad z^0 \in Z, \tag{3.3}$$

where the D_k 's are some point-to-set maps, i.e. $D_k: X \times Z \to \mathscr{P}(R^n)$ for all $k \geqslant 0$. Let $\mathscr{F}_k$ be the minimal σ-algebra which measures $((x^0, d^0, z^0), \ldots, (x^k, d^k, z^k))$. We assume that random variables $\tau_k \geqslant 0$, $\rho_k \geqslant 0$ are $\mathscr{F}_k$-measurable and τ_k is chosen so as to satisfy

$$x^k + \tau_k d^k \in X \tag{3.4}$$

for all $k \geqslant 0$. The R^n-valued random variables ξ^k satisfy the relations

$$E\{\xi^k / \mathscr{F}_k\} = \nabla F_k(x^k) + b^k, \tag{3.5}$$

where $F_k: \mathscr{X} \to R^1$ is a continuously differentiable function. We assume that for all $x, y \in \mathscr{X}$ and all $k \geqslant 0$ the functions F_k satisfy the inequalities

$$|F_k(x) - F(x)| \leqslant \alpha_k, \tag{3.6}$$

$$|F_k(x) - F_{k+1}(x)| \leqslant \beta_k, \tag{3.7}$$

$$\|\nabla F_k(x) - \nabla F_k(y)\| \leqslant \lambda_k \| x - y\|, \tag{3.8}$$

$$\|\nabla F_k(x) - \nabla F_{k+1}(x)\| \leqslant \mu_k, \tag{3.9}$$

where the variables α_k, β_k, λ_k, μ_k are $\mathscr{F}_k$—measurable. Let $X^* \subset X$ be a "solution set"; note that in mathematical programming X^* is not necessarily identical with the set of actual solutions of (P), e.g. X^* may be the set of all stationary points of (P). We assume that the set $F(X^*)$ is at most countabte. We make the following assumptions.

(A1) $\max_{x \in X} \min_{z \in Z} \|\nabla F_k(x) - z\| \to 0$ wp 1.

(A2) There exists a constant C such that $\|d^k\| \leqslant C$ and $E\{\|\xi^k\|^2 / \mathscr{F}_k\} \leqslant C$ wp 1 for all $k \geqslant 0$.

(A3) $\sum_{k=0}^{\infty} E\{\rho_k^2 + \rho_k \|b^k\|\} < \infty$.

(A4) $\sum_{k=0}^{\infty} \rho_k = \infty$ wp 1.

(A5) $(\lambda_k \tau_k + \mu_k)/\rho_k \to 0$ wp 1.

(A6) $\sum_{k=0}^{\infty} E\tau_k^2 < \infty$.

(A7) $\sum_{i=0}^{\infty} \tau_k = \infty$ wp 1.

(A8) $\beta_k / \tau_k \to 0$ and $\alpha_k \to 0$ wp 1.

(A9) If $x' \notin X^*$, $x' \in X$ then we can find $j \geqslant 0$, $\delta > 0$, $\gamma > 0$ such that for all $x \in U_\delta(x') \cap X$, $k \geqslant j$, $z \in U_\delta(\nabla F_k(x))$, $d \in D_k(x, z)$ there is $\langle \nabla F_k(x), d \rangle \leqslant -\gamma$.

Under the above assumptions we shall prove that for P-almost all $\omega \in \Omega$ all accumulation points of the path $\{x^k(\omega)\}$ belong to X^*.

Before proceeding to the convergence analysis let us make some comments. The general algorithm (3.1)–(3.3) may be considered as a two-level method. The auxiliary procedure (3.3) generates the sequence $\{z^k\}$ by means of averaging of ξ^i $(i < k)$. The basic algorithm (3.1)–(3.2) uses z^k to produce a feasible direction d^k and makes a feasible step in the direction d^k. We shall see in § 5 that (3.2) may be constructed as in classical feasible directions methods of nonlinear programming with z^k treated as a gradient of the objective function.

The filter (3.3) makes it possible to prove convergence without special conditions imposed on line search. This is due to the stabilizing effect of the operation of averaging.

4. Convergence of the auxiliary averaging procedure

In this section we shall prove that under the conditions of § 3 $z^k - \nabla F_k(x^k) \to 0$ wp 1. To do this we shall use the following result [5].

LEMMA 1. *Let $(\Omega, \mathscr{F}, P)$ be a probability space and let $\{\mathscr{F}_k\}$ be an increasing sequence of σ-fields contained in $\mathscr{F}$. Let $\{\eta^k\}$, $\{z^k\}$ be sequences of $\mathscr{F}_k$-measurable R^n-valued random variables satisfying the relations*

$$z^{k+1} = \Pi_Z(z^k + \rho_k(\xi^k - z^k)),\ z^0 \in Z,$$

$$E\{\xi^k/\mathscr{F}_k\} = \eta^k + b^k,$$

where $\rho_k \geqslant 0$ and b^k are $\mathscr{F}_k$-measurable and the set $Z \subset R^n$ is convex and closed. Next, let

(a) *all accumulation points of the sequence $\{\eta^k\}$ belong to Z* wp 1.

(b) *there exists a constant C such that $E\{\|\xi^k\|^2/\mathscr{F}_k\} \leqslant C$* wp 1 *for all $k \geqslant 0$,*

(c) $\sum_{k=0}^{\infty} E\{\rho_k^2 + \rho_k\|b^k\|\} < \infty$,

(d) $\sum_{k=0}^{\infty} \rho_k = \infty$ wp 1 and

(e) $\|\eta^{k+1} - \eta^k\|/\rho_k \to 0$ wp 1.

Then $z^k - \eta^k \to 0$ wp 1.

Proof of this lemma may be found in [5] (Ch. 2, thm 4.1).

As an immediate consequence of the above general result we obtain the following property of the auxiliary algorithm (3.3).

LEMMA 2. *Let (A1)–(A5) hold. Then $z^k - \nabla F_k(x^k) \to 0$ wp 1.*

Proof. We shall show that the sequences $\{\nabla F_k(x^k)\}$ and $\{z^k\}$ satisfy the assumptions of Lemma 1. Assumptions (b), (c), (d) are identical with (A2), (A3), (A4). Let us verify (a).

Let v^∞ be any accumulation point of the sequence $\{\nabla F_k(x^k)\}$. Then there exists a set of indices $\mathscr{K}$ such that

$$v^\infty = \lim_{\substack{k\to\infty \\ k \in \mathscr{K}}} \nabla F_k(x^k).$$

It follows from (A1)[1]) that we can find a sequence $\{v^k\} \subset Z$ such that $v^k - \nabla F_k(x^k) \to 0$. Therefore

$$v^\infty = \lim_{\substack{k\to\infty \\ k \in \mathscr{K}}} v^k \in Z,$$

which proves (a).

Let us now verify (e). If follows from (3.8), (3.9), (3.1) that

$$\|\nabla F_{k+1}(x^{k+1}) - \nabla F_k(x^k)\| \leqslant \|\nabla F_{k+1}(x^{k+1}) - \nabla F_k(x^{k+1})\| + \\ + \|\nabla F_k(x^{k+1}) - \nabla F_k(x^k)\| \leqslant \mu_k + \lambda_k \tau_k \|d^k\|.$$

[1]) The assumption (A1) is used only here. We shall see in §6 (formula (6.3)) that (A1) holds if $\bigcup_{x \in X} \partial F(x) \subset Z$, which simplifies the question of the choice of the set Z.

Since d^k is bounded then it follows from (A5) that

$\|\nabla F_{k+1}(x^{k+1})-\nabla F_k(x^k)\|/\rho_k \to 0$ wp 1, which completes the proof. ■

5. Convergence of the basic algorithm

In this section we establish convergence wp 1 of the sequence $\{x^k\}$ to the set X^*. The analysis will be based on the following result [15].

THEOREM 1. *Let $X^* \subset R^n$. Let $\{x^k\}$ be a bounded sequence in R^n, which satisfies the following conditions:*

(a) *if a subsequence $\{x^k\}_{k\in\mathscr{K}}$ converges to $x' \in X^*$ then $\|x^{k+1}-x^k\| \to 0$ for $k \in \mathscr{K}$.*

(b) *if a subsequence $\{x^k\}_{k\in\mathscr{K}}$ converges to $x' \notin X^*$ then there exists $\varepsilon_0>0$ such that for all $\varepsilon \in (0, \varepsilon_0]$ and all $k_0 \in \mathscr{K}$ the index*

$$s(k_0, \varepsilon)=\min\{k>k_0 : \|x^k-x^{k_0}\|>\varepsilon\}$$

is finite;

(c) *there exists a continuous function $W(x)$, attaining on X^* an at most countable set of values, such that if $\{x^k\}_{k\in\mathscr{K}} \to x' \notin X^*$ then there exists $\varepsilon_1>0$ such that for all $\varepsilon \in (0, \varepsilon_1]$*

$$\overline{\lim_{k_0\in\mathscr{K}}}\, W(x^{s(k_0,\varepsilon)}) < \lim_{k_0\in\mathscr{K}} W(x^{k_0})$$

where $s(k_0, \varepsilon)$ is defined as in (b).

Then the sequence $\{W(x^k)\}$ converges and all accumulation on points of the sequence $\{x^k\}$ belong to X^*.

In what follows we shall show that for P-almost all $\omega \in \Omega$ the paths $\{x^k(\omega)\}$ of the process $\{x^k\}$ satisfy the asumptions of Theorem 1.

Let $\{x^k(\omega)\}$ be a path of the process $\{x^k\}$. Let $\varepsilon>0$ $k_0 \geqslant 0$. We introduce the denotations

$$I(k_0, \varepsilon, \omega)=\{k \geqslant k_0 : x^i(\omega) \in U_\varepsilon(x^{k_0}(\omega)) \text{ for all } k_0 \leqslant i \leqslant k\},$$

$$s(k_0, \varepsilon, \omega)=\sup I(k_0, \varepsilon, \omega)+1.$$

LEMMA 3. *Let* (A1)–(A9) *hold. Let $\{x^k(\omega)\}_{k\in\mathscr{K}}$ be a subsequence convergent to $x'(\omega) \notin X^*$. Then wp* 1 *there exist $\varepsilon_0>0$, $\gamma>0$ and $k_{\min} \geqslant 0$* such *that if $k_0 \in \mathscr{K}$, $k_0 \geqslant k_{\min}$, $\varepsilon \in (0, \varepsilon_0]$ then for all $k \in I(k_0, \varepsilon, \omega)$ we have*

$$F(x^k(\omega)) \leqslant F(x^{k_0}(\omega))-\gamma \sum_{i=k_0}^{k-1} \tau_i(\omega)+r(k_0), \tag{5.1}$$

where $\lim_{k_0\to\infty} r(k_0)=0$.

Proof. For all $\varepsilon>0$, $k_0 \in \mathscr{K}$ and all $k \in I(k_0, \varepsilon, \omega)$ it follows from (3.7), (3.8) that

$$F_{k+1}(x^{k+1})-F_k(x^k) \leqslant [F_{k+1}(x^{k+1})-F_k(x^{k+1})]+F_k(x^{k+1})+$$
$$-F_k(x^k) \leqslant \beta_k+\tau_k\langle \nabla F_k(x^k), d^k\rangle+\lambda_k \tau_k^2 \| d^k\|^2,$$

where $d^k \in D_k(x^k, z^k)$. Combining (3.6) and the above inequality we obtain

$$F(x^k)-F(x^{k_0}) \leqslant \alpha_k + \alpha_{k_0} + \sum_{i=k_0}^{k-1} [\tau_i \langle \nabla F_i(x^i), d^i \rangle + \beta_i + \lambda_i \tau_i^2 \|d^i\|^2].$$

It follows from (A9) that we can find $j \geqslant 0$, $\delta > 0$, $\gamma > 0$ such that $\langle \nabla F_i(x^i), d^i \rangle \leqslant -2\gamma$ if $i \geqslant j$, $x^i \in U_\delta(x') \cap X$, $z^i \in U_\delta(\nabla F_i(x^i))$.

By virtue of Lemma 2, $z^i - \nabla F_i(x^i) \to 0$ wp 1. Next, $\|x^i - x'\| \leqslant \|x^i - x^{k_0}\| + \|x^{k_0} - x'\| \leqslant \varepsilon + \|x^0 - x'\|$, and $\|x^{k_0} - x'\| \to 0$ as $k_0 \to \infty$, $k_0 \in \mathscr{K}$. Therefore wp 1 we can find $\varepsilon_0(\omega) > 0$ and $j_1(\omega)$ such that $z^i(\omega) \in U_\delta(\nabla F_i(x^i(\omega)))$, $x^i(\omega) \in U_\delta(x'(\omega))$ for all $k_0 \geqslant j_1$, $k_0 \in \mathscr{K}$, $\varepsilon \in (0, \varepsilon_0]$, $i \in I(k_0, \varepsilon, \omega)$. Then

$$F(x(\omega)) - F(x^{k_0}(\omega)) \leqslant \alpha_k(\omega) - \alpha_{k_0}(\omega) +$$
$$+ \sum_{i=k_0}^{k-1} \left(-2\gamma + \frac{\beta_i(\omega)}{\tau_i(\omega)} + \lambda_i(\omega)\, \tau_i(\omega) \|d^i(\omega)\|^2 \right) \tau_i(\omega). \tag{5.2}$$

It follows from (A2), (A5), (A8) that $\beta_i/\tau_i + \lambda_i \tau_i \|d^i\|^2 \to 0$ wp 1. Thus wp 1 we can find $k_{\min}(\omega) \geqslant j_1(\omega)$ such that if $i \geqslant k_{\min}$ then $\beta_i(\omega)/\tau_i(\omega) + \lambda_i(\omega)\, \tau_i(\omega) \|d^i(\omega)\|^2 \leqslant \gamma$. Therefore for $k_0 \geqslant k_{\min}$ we obtain from (5.2) the inequality

$$F(x^k(\omega)) - F(x^{k_0}(\omega)) \leqslant -\gamma \sum_{i=k_0}^{k-1} \tau_i(\omega) + \alpha_k(\omega) + \alpha_{k_0}(\omega)$$

which holds for all $k \in I(k_0, \varepsilon, \omega)$. The above inequality, combined with (A8), completes the proof. ■

THEOREM 2. *Let* (A1)–(A9) *hold. Then for P-almost all* ω *the sequence* $\{F(x^k(\omega))\}$ *converges and all accumulation points of the sequence* $\{x^k(\omega)\}$ *belong to* X^*.

Proof. We shall apply Theorem 1, setting $W(x) = F(x)$. Assumption (a) is satisfied since d^k is bounded ((A2)) and $\tau_k \to 0$ wp 1 ((A5), (A3)). Let $\{x^k(\omega)\}_{k \in \mathscr{K}} \to x'(\omega) \notin X^*$, and let (b) be false. Then for all $\varepsilon > 0$ we can find $j \in \mathscr{K}$ such that $x^k(\omega) \in U_{\varepsilon/2}(x^j(\omega))$ for all $k \geqslant j$. Hence $x^k(\omega) \in U_\varepsilon(x^{k_0}(\omega))$ for all $k \geqslant k_0 \geqslant j$. It follows from Lemma 3 that for sufficiently small $\varepsilon > 0$ and sufficiently large k_0 the inequality (5.1) holds for all $k \geqslant k_0$. Then boundedness of F on X contradicts (A7). Thus (b) is satisfied for almost all ω.

We shall verify (c). Since $s(k_0, \varepsilon, \omega) - 1 \in I(k_0, \varepsilon, \omega)$ and $\rho_k(\omega) \to 0$ then for all $\varepsilon \in (0, \varepsilon_0]$ we can find $j \in \mathscr{K}$ such that $s(k_0, \varepsilon, \omega) \in I(k_0, 2\varepsilon, \omega)$ for all $k_0 \geqslant j$, $k_0 \in \mathscr{K}$. It follows from Lemma 3 that for sufficiently small $\varepsilon > 0$ and sufficiently large k_0 we have

$$F(x^{s(k_0, \varepsilon, \omega)}(\omega)) \leqslant F(x^{k_0}(\omega)) - \gamma \sum_{i=k_0}^{s(k_0, \varepsilon, \omega)-1} \tau_i(\omega) + r(k_0). \tag{5.3}$$

On the other hand

$$\varepsilon < \|x^{s(k_0, \varepsilon, \omega)}(\omega) - x^{k_0}(\omega)\| \leqslant C \sum_{i=k_0}^{s(k_0, \varepsilon, \omega)-1} \tau_i(\omega). \tag{5.4}$$

Combining (5.3) and (5.4) we obtain the inequality

$$F(x^{s(k_0,\varepsilon,\omega)}(\omega)) \leqslant F(x^{k_0}(\omega)) - \gamma\varepsilon/C + r(k_0).$$

After transition to the limit with $k_0 \to \infty$, $k_0 \in \mathscr{K}$ we obtain the required inequality. ■

6. The construction of gradient estimates ξ^k.

The construction of the sequence $\{F_k\}$ is based on the properties of Lipschitz functions collected in Proposition 1. Let $\|\cdot\|_*$ be an arbitrary norm in R^n and let $Q=\{u \in R^n: \|u\|_* \leqslant 1\}$, $S=\{u \in R^n: \|u\|_* = 1\}$. Let $h: R^n \to R^1$ be a nonnegative integrable function with support Q and let $\int_{R^n} h(u)\,du = 1$. Let $p>0$. We define

$$\hat{F}(x,p) = \int_{R^n} h(u)\, F(x-pu)\, du. \tag{6.1}$$

It follows from (c) of Proposition 1 that $\hat{F}(\cdot, p)$ is differentiable for all $p>0$ such that $X+pQ \subset \mathscr{X}$. The gradient is given by

$$\nabla_x \hat{F}(x,p) = \int_{R^n} h(u)\, \nabla F(x-pu)\, du. \tag{6.2}$$

Hence, by virtue of points (a) and (d) of Proposition 1, for any compact set $X \subset \mathscr{X}$,

$$\lim_{p \downarrow 0} \max_{x \in X} \min_{v \in \partial F(x)} \|v - \nabla_x \hat{F}(x,p)\| = 0. \tag{6.3}$$

Next, applying to (6.1) the rules of differentiation of the convolution of distributions, we obtain

$$\nabla_x \hat{F}(x,p) = \frac{1}{p} \int_{R^n} \nabla h(u)\, F(x-pu)\, du, \tag{6.4}$$

where $\nabla h = \left(\frac{\partial h}{\partial u_1}, \frac{\partial h}{\partial u_2}, \ldots, \frac{\partial h}{\partial u_n}\right)$ is a generalized vector-function (see e.g. [7]). A particularly simple and important for practice case of (6.4) is that with $h(u)$ constant within Q. Let V be the volume of Q and let σ be the area of S. Then we obtain from (6.4) the equalities

$$\nabla_x \hat{F}(x,p) = \frac{\sigma}{Vp} \int_S \frac{1}{\sigma} N(u)\, F(x+pu)\, dS \tag{6.5}$$

and

$$\nabla_x \hat{F}(x,p) = \frac{\sigma}{2Vp} \int_S \frac{1}{\sigma} N(u)\, [F(x+pu) - F(x-pu)]\, dS, \tag{6.6}$$

where $N(u)$ is the outer normal to S at $u \in S$ (see [7]).

Let us define $F_k(x) = \hat{F}(x, p_k)$ and let $p_k \downarrow 0$. It follows from (6.1) that (3.6) holds with $\alpha_k = \text{const}\cdot p_k$, and (3.7) holds with $\beta_k = \text{const}\cdot|p_k - p_{k+1}|$. Next, it follows from (6.6) that (3.8) holds with $\lambda_k = \text{const}/p_k$, and (3.9) holds with $\mu_k = \text{const} \times$

$|p_k - p_{k+1}|/p_k$. Then conditions (A5) and (A8) take on the form: (A5') $\tau_k/p_k\, p_k + |p_k + - p_{k+1}|/p_k\, p_k \to 0$ and (A8') $|p_k - p_{k+1}|/\tau_k \to 0$. Of course, it is very difficult to compute the value of $F_k(x)$ at a given x. However, the algorithm from § 3 does not use the values of $F_k(x^k)$; the only information required are the vectors ξ^k which satisfy (3.5). The formulae (6.1), (6.2), (6.5) and (6.6) make it possible to construct various estimates of $\nabla F_k(x^k)$, satisfying the required conditions. We shall show four typical examples for the stochastic case, i.e. the case when $F(x) = Ef(x, \theta)$. In order to satisfy (A1) we shall assume that the inequality

$$|f(x, \theta) - f(y, \theta)| \leqslant \tilde{L}\|x - y\| \tag{6.7}$$

holds with some $\tilde{L} > 0$ for all $x, y \in \mathscr{X}$ and P-almost all θ. In what follows $\theta^{k,i}$ denote samples of θ, such that $\theta^{k,i}$ is independent of all x^j, d^j, z^j $(j \leqslant k)$.

Example 1 (analytical gradients)

Assume that for given x, θ it is possible to compute an element of $\partial_x f(x, \theta)$. Then we can define $\xi^k = v^k$, where $v^k \in \partial_x f(x^k - p_k u^k, \theta^{k,0})$ and u^k is drawn randomly from a uniform distribution over Q. It follows from (6.2) that (3.5) holds with $b^k = 0$.

Example 2 (naive gradient estimation)

If it is not possible to compute an element of $\partial_x f(x, \theta)$ then we can define

$$\xi^k = \frac{1}{\Delta_k} \sum_{i=1}^{n} [f(x^k - p_k u^k + \Delta_k e^i, \theta^{k,i}) - f(x^k - p_k u^k - \Delta_k e^i, \theta^{k,i})]\, e^i$$

where u^k is drawn randomly from a uniform distribution over Q and e^i is the unit vector of the i-th coordinate. It follows from (6.1) that (3.5) holds with bias b^k of the range Δ_k/p_k.

Example 3 (gradient estimation on the sphere)

Let $\|u\|_* = \|u\|$. Then $\sigma = nV$ and $N(u) = u$ in (6.5), (6.6). Let $n_k > 0$ for all $k \geqslant 0$ and let

$$\xi^k = \frac{n}{2n_k p_k} \sum_{i=1}^{n_k} [f(x^k + p_k u^{k,i}, \theta^{k,i}) - f(x^k - p_k u^{k,i}, \theta^{k,i})]\, u^{k,i},$$

where $u^{k,i}$ $(i = 1, 2, \ldots, n_k)$ are drawn randomly from a uniform distribution over S. It follows from (6.6) that (3.5) holds with $b^k = 0$.

Example 4 (gradient estimation on the cube)

Let $\|u\|_* = \|u\|_\infty = \max_{1 \leqslant i \leqslant n} |u_i|$. Then $\sigma = nV$ in (6.5), (6.6) and $N(u) = \text{sign}(u) \cdot e^r$, where $|u_r| = \|u\|_\infty$. Thus, basing on (6.5), we can construct the following algorithm for the estimation of $\nabla F_k(x^k)$. A random point u^k is drawn from a uniform distribution in Q and $2n$ points in S are defined by projecting u^k on the faces of the hypercube Q:

$$\begin{aligned} u_+^{k,i} &= (u_1^k, \ldots, u_{i-1}^k, \quad 1, u_{i+1}^k, \ldots, u_n^k), \\ u_-^{k,i} &= (u_1^k, \ldots, u_{i-1}^k, -1, u_{i+1}^k, \ldots, u_n^k). \end{aligned} \qquad i=1, \ldots, n$$

Finally, the vector

$$\xi^k = \frac{1}{2p_k} \sum_{i=1}^{n} [f(x^k + p_k u_+^{k,i}, \theta^{k,i}) - f(x^k + p_k u_-^{k,i}, \theta^{k,i})] e^i$$

is computed. Since $N(u_+^{k,i}) = e^i$ and $N(u_-^{k,i}) = -e^i$, it follows from (6.5) that ξ^k satisfies (3.5) with $b^k = 0$.

Let us note that in all the above examples (A2) holds by virtue of (6.7); that is why we have taken the same $\theta^{k,i}$ in square brackets (this is possible in off-line computations).

It is also possible to use in (6.1) nonuniform functions $h(\cdot)$ which may yield new interesting algorithms for the estimation of $\nabla_x \hat{F}(x, p)$.

Finally, let us note that the problem of estimating the derivatives of smooth functions was discussed extensively in [10], and in [8] the method from Example 4 was evaluated by direct differrentiation of $\hat{F}(\cdot, p)$.

7. Some definite algorithms

In this section we shall discuss methods for the construction of mappings D_k satisfying conditions from § 3. Our aim is to show that only slight modifications are necessary to develop stochastic analogues of many deterministic feasible direction algorithms.

It will be assumed throughout this section that the sequence $\{F_k\}$ is defined as in § 6. We shall also assume that the gradients $\nabla g_i(\cdot)$ of the constraint functions of (P) are Lipschitz continuous.

Example 1

Let $x \in X$, $z \in Z$. We define the subproblem SP (x, z)

$$\max \eta$$

$$\text{subject to} \quad \begin{aligned} &\langle z, s \rangle + \eta \leqslant 0, \\ &g_i(x) + \langle \nabla g_i(x), s \rangle + \eta \leqslant 0 \quad \text{for } i=1, \ldots, m, \\ &\|s\|_* \leqslant 1. \end{aligned}$$

Let $\eta(x, z)$, $s(x, z)$ denote any solution of SP (x, z) and let $D_k(x, z) = D(x, z) = \{d \in R^n: d = \eta(x, z)\, s(x, z)\}$. Let X^* be the set of all stationary points of (P). Evidently

$$X^* = \{x^* \in X: \min_{v \in \partial F(x^*)} \eta(x^*, v) = 0\}.$$

We shall prove that the map D satisfies assumptions of § 3. Let us verify (A9). If $x' \notin X^*$ then there exists $\gamma > 0$ such that $\eta(x', v) \geq 2\gamma$ for all $v \in \partial F(x')$. Let us consider the point-to-set map $T\colon X \times [0, 1] \to \mathscr{P}(R^1)$, $T(x, \delta) = \eta\left(x, U_{2\delta}(\partial F(x))\right)$. It follows from Proposition 1, (d) that the map $(x, \delta) \to \left(x, U_{2\delta}(\partial F(x))\right)$ is closed and uniformly bounded on $X \times [0, 1]$. The function $\eta(\cdot, \cdot)$ is continuous. Thus the map T is closed [22]. Since $T(x', 0) \subset [2\gamma, \infty)$ then there exists $\delta > 0$ such that $\eta(x, z) \geq \gamma$ for all $x \in U_\delta(x') \cap X$, $z \in U_{2\delta}(\partial F(x))$. By virtue of (6.3) we can find $j \geq 0$ such that $U_\delta(\nabla F_k(x)) \subset U_{2\delta}(\partial F(x))$ for all $k \geq j$ and all $x \in X$. Therefore $\langle z, d \rangle \leq -\gamma^2$ if $d \in D(x, z)$, $x \in U_\delta(x') \cap X$, $z \in U_\delta(\nabla F_k(x))$ and $k \geq j$. Since $\nabla F_k(x) \in U_\delta(z)$ then for sufficiently small δ we have $\langle \nabla F_k(x), d \rangle \leq -\frac{1}{2}\gamma^2$ which completes the proof of (A9).

The only question to be clarified is the possibility of choosing a sequence $\{\tau_k\}$ in such a way that (3.4), (A7), (A8) hold. Let us make the denotations $s^k = s(x^k, z^k)$, $\eta_k = \eta(x^k, z^k)$, $d^k = \eta_k s^k$. For any $\tau \geq 0$ we have the inequalities

$$g_i(x^k + \tau d^k) \leq g_i(x^k) + \tau \eta_k \langle \nabla g_i(x^k), s^k \rangle + L_g \tau^2 \eta_k^2 \|s^k\|^2 \leq$$
$$\leq (1 - \tau \eta_k)\, g_i(x^k) + \tau \eta_k^2 (-1 + L_g \tau \|s^k\|^2)$$

where L_g is the Lipschitz constant for $\nabla g_i(\cdot)$. Since η_k and $\|s^k\|$ are bounded, we can find a constant $C > 0$ such that

$$g_i(x^k + \tau d^k) \leq (1 - C\tau)\, g_i(x^k) + \tau \eta_k^2 (-1 + C\tau),$$

which shows that all $\tau_k \leq 1/C$ are feasible and (3.4), (A7), (A8) are consistent. Let us note that this feature has been achieved by the re-scaling of directions $d^k = \eta_k s^k$ [2].

The above method is a stochastic analogue of the method from [21]. Another stochastic version was analysed in [11], [12] for smooth convex problems.

Example 2

Let $x \in X$, $z \in Z$, $\varepsilon \geq 0$. We define the subproblem SP (x, z, ε)

$$\max \eta$$

subject to

$$\langle z, s \rangle + \eta \leq 0,$$
$$\langle \nabla g_i(x), s \rangle + \eta \leq 0 \quad \text{for} \quad i \in I(x, \varepsilon),$$
$$\|s\|_* \leq 1,$$

where $I(x, \varepsilon) = \{1 \leq i \leq m\colon g_i(x) \geq -\varepsilon\}$. We denote by $\eta(x, z, \varepsilon)$, $s(x, z, \varepsilon)$ any solution of SP (x, z, ε). Let $\varepsilon_k \downarrow 0$ wp 1 and let $D_k(x, z) = \{d \in R^n\colon d = \eta(x, z, \varepsilon_k)\, s(x, z, \varepsilon_k)\}$. Evidently, the set X^* of all stationary points of (P) has the form

$$X^* = \{x^* \in X\colon \min_{v \in \partial F(x^*)} \eta(x, v, 0) = 0\}.$$

Let us verify (A9). If $x' \notin X^*$ then there exists $\gamma > 0$ such that $\eta(x', v, 0) \geq 2\gamma$ for all $v \in \partial F(x')$. We can find $\bar{\delta} > 0$ and $\bar{\varepsilon} > 0$ such that $I(x, \bar{\varepsilon}) = I(x', 0)$ for $x \in U_{\bar{\varepsilon}}(x') \cap X$. Let us define the point-to-set map $T\colon X \times [0, \bar{\delta}] \to \mathscr{P}(R^1)$, $T(x, \delta) =$

$=(x, U_{2\delta}(\partial F(x)), \bar{\varepsilon})$. Since $I(x, \bar{\varepsilon})$ is constant within $U_\delta(x')$, then $\eta(\cdot, \cdot, \bar{\varepsilon})$ is continuous over $(U_{\bar{\delta}}(x') \cap X) \times Z$. Proceeding as in Example 1, we can prove that the map T is closed on $(U_{\bar{\delta}}(x') \cap X) \times [0, 1]$. Therefore we can find $\delta \in (0, \bar{\delta}]$ such that $\eta(x, z, \bar{\varepsilon}) \geqslant \gamma$ for all $x \in U_\delta(x') \cap X$, $z \in U_{2\delta}(\partial F(x))$. Next, $\eta(x, z, \varepsilon_k) \geqslant \geqslant \eta(x, z, \bar{\varepsilon})$ for $\varepsilon_k \in [0, \bar{\varepsilon}]$. Finally, $U_\delta(\nabla F_k(x)) \subset U_{2\delta}(\partial F(x))$ for sufficiently large k. Therefore $\langle z, d \rangle \leqslant -\gamma^2$ if $d \in D_k(x, z)$, $x \in U_\delta(x') \cap X$, $z \in U_\delta(\nabla F_k(x))$ and k is large enough. For sufficiently small δ we have also $\langle \nabla F_k(x), d \rangle \leqslant -\frac{1}{2}\gamma^2$ which proves (A9).

The question of τ_k-feasibility of the directions d^k is more complicated than in Example 1. For $i \in I(x^k, \varepsilon_k)$ we have

$$g_i(x^k + \tau d^k) \leqslant g_i(x^k) + \tau\eta_k \langle \nabla g_i(x^k), s^k \rangle + L_g \tau^2 \eta_k^2 \|s^k\|^2 \leqslant g_i(x^k) + \tau\eta_k^2(-1 + C_1 \tau)$$

where C_1 is a bound for $L_g \|s^k\|^2$. If $i \notin I(x^k, \varepsilon_k)$ then in general we can guarantee only that

$$g_i(x^k + \tau d^k) \leqslant -\varepsilon_k + C_2 \tau$$

for some constant C_2. Thus, to be sure that (3.4) and (A7), (A8) are consistent we should impose on the sequence $\{\varepsilon_k\}$ conditions (A7), (A8) with τ_k replaced by ε_k.

The method analysed in Example 2 is a stochastic version of the classical feasible direction algorithm [17], [22], [23]. Versions similar to curs were suggested in [1], [8] but the lack of re-scaling of directions and the lack of any assumptions on the sequence $\{\varepsilon_k\}$ resulted in certain inaccuracies, connected with incosistency of some assumptions.

Example 3

Let $x \in X$, $z \in Z$, $\varepsilon \geqslant 0$. We define the subproblem SP (x, z, ε):

$$\min \langle z, s \rangle$$

subject to $\quad \langle \nabla g_i(x), s \rangle + \varepsilon \leqslant 0$ for $i \in I(x, \varepsilon)$,

$$\|s\|_* \leqslant 1,$$

where $I(x, \varepsilon)$ is defined as in Example 2. We denote by $s(x, z, \varepsilon)$ any solution of SP (x, z, ε) and if the feasible set of SP (x, z, ε) is empty, then we set $s(x, z, \varepsilon) = 0$. Let $\varepsilon_k \downarrow 0$ wp 1 and let $D_k(x, z) = \{d \in R^n: d = s(x, z, \varepsilon_k)\}$. For simplicity we assume that for each point $x \in X$ the gradients $\nabla g_i(x)$ of the constraints binding at x are linearly independent. Then the set X^* of all stationary points of (P) has the form

$$X^* = \{x^* \in X: \max_{v \in \partial F(x^*)} \langle v, s(x^*, v, 0) \rangle\} = 0.$$

Let us verify (A9). If $x' \notin X^*$ then there exists $\gamma > 0$ such that $\langle v, s(x', v, 0) \rangle \leqslant -3\gamma$ for all $v \in \partial F(x^*)$. Under the assumed constraint qualification we can find $\bar{\delta} > 0$ and $\bar{\varepsilon} > 0$ such that $I(x, \bar{\varepsilon}) = I(x', 0)$ for $x \in U_{\bar{\delta}}(x') \cap X$ and the feasible set of SP (x, z, ε) is non-void for $x \in U_{\bar{\delta}}(x') \cap X$, $\varepsilon \in [0, \bar{\varepsilon}]$, $z \in Z$. Let us define the map T: $X \times [0, \bar{\delta}] \to \mathscr{P}(R^1)$, $T(x, \delta) = \{\langle z, s(x, z, \bar{\varepsilon}) \rangle : z \in U_{2\delta}(\partial F(x))\}$. It may be easily proved that T is closed on $(U_{\bar{\delta}}(x') \cap X) \cap [0, \bar{\delta}]$. Therefore we can find $\delta \in (0, \bar{\delta}]$ such that $\langle z, s(x, z, \bar{\varepsilon})\rangle \leqslant -2\gamma$ for all $x \in U_\delta(x') \cap X$, $z \in U_{2\delta}(\partial F(x))$. Next, $\langle z, s(x, z, \varepsilon_k)\rangle \leqslant \langle z, s(x, z, \bar{\varepsilon})\rangle$ for $\varepsilon_k \in [0, \bar{\varepsilon}]$ and also $U_\delta(\nabla F_k(x)) \subset U_{2\delta}(\partial F(x))$ for large k. Therefore $\langle z, d\rangle \leqslant -2\gamma$ if $d \in D_k(x, z)$, $x \in U_\delta(x') \cap X$, $z \in U_\delta(\nabla F_k(x))$ and k is large enough. For small δ we have also $\langle \nabla F_k(x), d\rangle \leqslant -\gamma$ which proves (A9).

Let us verify whether (3.4) and (A5)–(A8) are consistent. If $i \in I(x^k, \varepsilon_k)$ then

$$g_i(x^k + \tau d^k) \leqslant -\tau\varepsilon_k + C_1 \tau^2.$$

If $i \notin I(x^k, \varepsilon_k)$ then in general we have only

$$g_i(x^k + \tau d^k) \leqslant -\varepsilon_k + C_2 \tau.$$

In both cases we see that feasible steps τ_k will tend to zero as fast as the parameters ε_k. Therefore we should impose on the sequence $\{\varepsilon_k\}$ conditions (A7), (A8) with τ_k replaced by ε_k.

The above algorithm is a stochastic version of the method from [24] (see also [17]). Following the above-sketched manner one can easily prove convergence of stochastic nondifferentiable analogues of various classical feasible direction methods (see [20] for smooth stochastic examples).

References

[1] Bazhenov L. G. and Gupal A. M. On a certain stochastic analogue of the feasible direction algorithm. *Kibernetika* (Kiev), **9** (1973), 94–95 [in Russian].

[2] Brdyś M. and Ruszczyński A. Note on the application of Zangwill's theory to feasible direction method , report I.2.03. Instytut Automatyki, Politechnika Warszawska, 1978. [in Polish].

[3] Clarke F. H. Generalized gradient and application. *Trans. Amer. Math. Soc.*, **205** (1975), 247–262.

[4] Clarke F. H. A new approach to Lagrange multiplier. *Math. of Operations Res.* **1** (1976), 165–174.

[5] Ermoliev Yu. M. Stochastic Programming Method. Nauka, Moskva, 1976. [in Russian].

[6] Fabian V. Stochastic approximation of constrained minima, Transactions of the 4th Prague Conference on Information Theory, Statistical Decision Functions and Random Processes 1965, Prague, 1967, 277–290.

[7] Gel'fand I. M. and Shilov G. E. Generalized Functions — volume 1: Properties and Operations, Academic Press, New York and London, 1964.

[8] Gupal A. M. Stochastic Methods for the Solution of Nonsmooth Extremal Problems. Naukova Dumka, Kiev, 1979. [in Russian].

[9] Huard P. (ed)., Point-to-set Maps and Mathematical Programming. *Mathematical Programming Study* 10: North–Holland, Amsterdam, 1979.

[10] Katkovnik V. Ya. Linear Estimates and Stochastic Optimization Problems. Nauka, Moskva, 1976. [in Russian].

[11] Kushner H. J. and Clark D. S. Stochastic Approximation Methods for Constrained and Unconstrained Systems. Springer–Verlag, New York, 1978.

[12] Kushner H. J. and Gavin T. Stochastic approximation type methods for constrained systems: algorithms and numerical results. *IEEE Transactions on Automatic Control*, **AC-19** (1974), 249–357.

[13] Lemarechal C. and Mifflin R. (eds), Nonsmooth Optimization. Pergamon Press, 1978.

[14] Mifflin R. Semismooth and semiconvex functions in constrained optimization. *SIAM J. Control and Optimization*, 15 (1977), 959–972.
[15] Nurminski E. A. Convergence conditions for nonlinear programming methods. *Kibernetika* (Kiev), 8 (1972), 79–81. [in Russian].
[16] Pervozvanskii A. A. Mathematical Models in Management. Nauka, Moskva, 1975. [in Russian].
[17] Polak E. Computational Methods in Optimization. Academic Press, New York, 1971.
[18] Poljak B. T. Nonlinear programming methods in the presence of noise. *Mathematical Programming*, 14 (1978), 87–97.
[19] Ruszczyński A. Design of decentralized control systems by stochastic programming methods, in: W. Findeisen (ed), Second Workshop on Hierarchical Control, Warszawa 1979.
[20] Ruszczyński A. Feasible direction methods for stochastic programming problems. *Mathematical Programming*, 19 (1980), 220-229.
[21] Topkis D. M. and Veinott A. F. On the convergence of some feasible directon algorithms. *SIAM J. Control*, 5 (1967) 268–279.
[22] Zangwill W. I. Nonlinear Programming: a Unified Approach. Prentice Hall, Englewood Cliffs, New York, 1969.
[23] Zoututendjk G. Methods of Feasible Directions. Elsevier, Amsterdam, 1960.
[24] Zukhovitskii S. T., Poljak R. A. and Primak M. E. Algorithms for the solution of a convex programming problem. *Doklady AN SSSR*, 153 (1963), 991–1000. [in Russian].

Received, June 1980

Stochastyczne metody kierunków dopuszczalnych dla niegładkich problemów optymalizacji stochastycznej

W artykule rozpatrywane są stochastyczne metody kierunków dopuszczalnych dla niegładkich i stochastycznych problemów optymalizacji. Przedstawiono model algorytmu wykorzystujący odwzorowania wielowartościowe oraz udowodniono zbieżność przy ogólnych założeniach dotyczących odwzorowań opisujących metodę. Na zakończenie, wykorzystując ogólną toerię stworzono stochastyczne, zorientowane na funkcje niegładkie odpowiedniki pewnych klasycznych algorytmów kierunków dopuszczalnych.

Стохастические методы допустимых направлений для негладких задач стохастической оптимизации

В статье рассматриваются стохастические методы допустимых направлений для негладких и стохастических задач оптимизации. Представлена модель алгоритма, использующая многозначные отображения, а также доказана сходимость при общих предносылках, касающихся отображений описывающих метод. В заключении, используя общую теорию, созданы стохастические, ориентированные на негладкие функции, аналоги некоторых классических алгоритмов допустимых направлений.