International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

Constrained Optimization of Discontinuous Systems

Ermoliev, Y.M. & Norkin, V.I.

IIASA Working Paper

WP-96-078

July 1996

Working Paper

CONSTRAINED OPTIMIZATION OF DISCONTINUOUS SYSTEMS

Yuri M. ERMOLIEV
Vladimir I. NORKIN

WP-96-78
July 1996

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

CONSTRAINED OPTIMIZATION OF DISCONTINUOUS SYSTEMS

Yuri M. ERMOLIEV
Vladimir I. NORKIN

WP-96-78
July 1996

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 807 □ Fax: +43 2236 71313 □ E-Mail: info@iiasa.ac.at

Abstract

In this paper we extend the results of Ermoliev, Norkin and Wets [8] and Ermoliev and Norkin [7] to the case of constrained discontinuous optimization problems. In contrast to [7] the attention is concentrated on the proof of general optimality conditions for problems with nonconvex feasible sets. Easily implementable random search technique is proposed.

Key words: Discontinuous Systems, Necessary Optimality Conditions, Averaged Functions, Mollifier Subgradients, Stochastic Optimization.

Contents

CONSTRAINED OPTIMIZATION OF DISCONTINUOUS SYSTEMS

Yuri M. ERMOLIEV
Vladimir I. NORKIN

1 Introduction

In this paper we elaborate further results of Ermoliev, Norkin and Wets [8] and Ermoliev and Norkin [7] to a general constrained discontinuous optimization problem:

$$\text{minimize } F(x) \tag{1}$$

$$\text{subject to } x \in K \subseteq R^n, \tag{2}$$

where $F(x)$ is a (strongly) lower semicontinuous function, K is a compact set.

As we showed in [7] the class of strongly lower semicontinuous functions is appropriate for modeling and optimization of abruptly changing systems without instantaneous jumps and returns. In particular, we analyzed risk control problems, optimization of stochastic networks and discrete event systems, screening irreversible changes and stochastic pollution control. Another important application may be stochastic jumping processes describing risk reserves of interdependent insurance and reinsurance companies. In a rather general form the risk reserves can be understood as "reservoirs", where risk premiums are continuously flowing in and random claims at random time moments abruptly draining them out. A sample path of such process is a strongly lower semicontinuous function with random jumps at claim occurrence times.

In a sense the main aim of this article is to provide proofs of necessary optimality conditions for general discontinuous constrained optimization problems discussed in [7]. In Section 2 we analyze situations when the expectation function belongs to the class of strongly lower semicontinuous functions. General idea of discontinuous optimization is presented in Section 3. Optimality conditions for discontinuous functions and general constraints are analysed in Section 4. Section 5 outlines possible computational procedures.

2 Some classes of discontinuous functions

In nonsmooth analysis different classes of continuous functions are introduced and studied. The same is necessary for discontinuous functions. We basically restrict possible discontinuity to the case of strongly lower semicontinuous functions, which seem to be most important for applications.

Definition 2.1 *A function $F: R^n \longrightarrow R^1$ is called strongly lower semicontinuous at x, if it is lower semicontinuous at x and there exists a sequence $x^k \longrightarrow x$ with F continuous at x^k (for all k) such that $F(x^k) \longrightarrow F(x)$. The function F is called strongly lower semicontinuous (strongly lsc) on $X \subseteq R^n$ if this holds for all $x \in X$.*

Definition 2.2 *Lower semicontinuous function $F: R^n \longrightarrow R^1$ is called directionally continuous at x if there exists an open (direction) set $D(x)$ containing sequences $x^k \in D(x)$, $x^k \longrightarrow x$ such that $F(x^k) \longrightarrow F(x)$. Function $F(x)$ is called directionally continuous if this holds for any $x \in R^n$.*

Definition 2.3 *Function $F(x)$ is called piecewise continuous if for any open set $A \subset R^n$ there is another open set $B \subset A$ on which $F(x)$ is continuous.*

Proposition 2.1 *If function $F(x)$ is piecewise continuous and directionally continuous then it is strongly lower semicontinuous.*

Proof. By definition of piecewise continuity for any open vicinity $V(x)$ of x we can find an open set $B \subset D(x) \cap V(x)$ on which function F is continuous. Hence there exists sequence $x^k \in D(x)$, $x^k \longrightarrow x$ with F continuous at x^k. By definition of directional continuity $F(x^k) \longrightarrow F(x)$.$\square$

Properties of directional continuity, piecewise continuity and strong lower semicontinuity can be easily verified for one dimensional function. For instance, if one dimensional function $F(x)$, $x \in R$, is (i) lower semicontinuous, (ii) continuous almost everywhere in R and (iii) at each point of discontinuity $x \in R$ function $F(x)$ is continuous either from the left or from the right, then $F(x)$ is strongly lsc. Next proposition clarifies the structure of multidimensional discontinuous functions of interest.

Proposition 2.2 *If $F(x) = F_0(F_1(x_1), \ldots, F_m(x_m))$, where $x = (x_1, \ldots, x_m)$, $x_i \in R^{n_i}$, function $F_0(\cdot)$ is continuous and functions $F_i(x_i)$, $i = 1, \ldots, n$ are strongly lsc (directionally continuous), then the composite function $F(x)$ is also strongly lsc (directionally continuous). If $F(x) = F_0(F_1(x), \ldots, F_m(x))$, $x \in R^n$, where $F_0(\cdot)$ is continuous and $F_i(x)$, $i = 1, \ldots, m$, are piecewise continuous, then $F(x)$ is also piecewise continuous.*

In particular, strong lsc, directional continuity and piecewise continuity are preserved under continuous transformations.

Proof is evident.

The next proposition gives a sufficient condition for a mathematical expectation function $F(x) = \mathbf{E}f(x,\omega)$ to be strongly lower semicontinuous.

Proposition 2.3 *Assume function $f(\cdot,\omega)$ is locally bounded around x by an integrable (in ω) function, piecewise continuous around x and a.s. directionally continuous at x with direction set $D(x,\omega) = D(x)$ (independent of ω). Suppose ω takes only a finite or countable number of values. Then $F(x) = \mathbf{E}f(x,\omega)$ is strongly lsc at x.*

Proof. Lower semicontinuity of F follows from Fatu lemma. The convergence of $F(x^k)$ to $F(x)$ for $x^k \longrightarrow x$, $x^k \in D(x)$ follows from Lebesgue's dominant convergence theorem. Hence F is directionally continuous at x in $D(x)$. It remains to show that in any open set $A \subset R^n$ which is close to x there are points of continuity of F. For the case when ω takes finite number of values $\omega_1,\ldots,\omega_m$ with probabilities $p_1,\ldots,p_m$ the function $F(\cdot) = \sum_{i=1}^m p_i f(\cdot,\omega_i)$ is clearly piece-wise continuous. For the case when ω takes a countable number of values there is a sequence of closed balls $B_i \subsetneq B_{i-1} \subset A$ convergent to some point $y \in A$ with $f(\cdot,\omega_i)$ continuous on B_i. We shall show that $F(\cdot) = \sum_{i=1}^\infty p_i f(\cdot,\omega_i)$ is continuous at y. By assumption $|f(x,\omega_i)| \leq C_i$ for $x \in A$ and $\sum_{i=1}^\infty p_i C_i < +\infty$. Then

$$F(x) - F(y) = \sum_{i=1}^{\infty} p_i(f(x,\omega_i) - f(y,\omega_i)) = \sum_{i=1}^{m} p_i(f(x,\omega_i) - f(y,\omega_i)) + \delta_m(x,y),$$

$$|\delta_m(x,y)| \leq \sum_{i=m+1}^{\infty} 2p_iC_i \quad x,y \in A.$$

Thus for any $x^k \longrightarrow y$

$$\limsup_k F(x^k) \leq F(y) + \sum_{i=m+1}^{\infty} 2p_iC_i;$$

$$\liminf_k F(x^k) \geq F(y) - \sum_{i=m+1}^{\infty} 2p_iC_i.$$

Since $\sum_{i=m+1}^\infty 2p_iC_i \longrightarrow 0$ as $m \longrightarrow \infty$ then $\lim_k F(x^k) = F(y)$.□

Let us remark that functions of the form $f(x,\omega) = f(x-\omega)$, $x,\omega \in R^n$, with $f(\cdot)$ piecewise and directionally continuous have $D(x)$ independent of ω.

Propositions 2.1-2.3 provide a certain calculous for strongly lsc functions.

3 Averaged functions and mollifier subgradients

In order to optimize discontinuous functions we approximate them by so-called averaged functions which are often considered in optimization theory (see Yudin [23], Hasminski [12], Antonov and Katkovnik [1], Zaharov [24], Katkovnik and Kulchitsky [14] Nikolaeva [18], Archetti and Betrò [2], Warga [25], Katkovnik,[13], Gupal [9], [10], Gupal and Norkin [11], Rubinstein [22], Batuhtin and Maiboroda [4], Mayne and Polak [16], Mikhalevich, Gupal and Norkin [17], Ermoliev and Gaivoronski [6], Kreimer and Rubinstein [15], Batuhtin [3], Ermoliev, Norkin and Wets [8]). The convolution of a discontinuous function with appropriate mollifier (probability density function) improves continuity and differentiability, but on the other hand increases computational complexity of resulting problems since it transfers a deterministic function $F(x)$ into an expectation function defined as multiple integral. Therefore, this operation is meaningful only in combination with appropriate stochastic optimization techniques. Our purpose is to introduce such technique and to develop a certain subdifferential calculous for discontinuous functions. Let us introduce necessary notions and facts which are generalized in the next section to the case of constrained problems.

Definition 3.1 *Given a locally integrable (discontinuous) function $F:\ R^n \longrightarrow R^1$ and a family of mollifiers $\{\psi_\theta:\ R^n \longrightarrow R_+,\ \ \theta \in R_+\}$ that by definition satisfy $\int_{R^n}\psi_\theta(z)dz = 1$, $\ supp\psi_\theta := \{z \in R^n |\ \psi_\theta(z) > 0\} \subseteq \rho_\theta \mathbf{B}$ with a unit ball $\mathbf{B}$, $\rho_\theta \downarrow 0$ as $\theta \downarrow 0$, the associated family $\{F^\theta,\ \theta \in R_+\}$ of averaged functions is defined by*

$$F^\theta(x) := \int_{R^n} F(x-z)\psi_\theta(z)dz = \int_{R^n} F(z)\psi_\theta(x-z)dz. \tag{3}$$

Mollifiers may also have unbounded support (see [8]).

Example 3.1 *Assume $F(x) = Ef(x,\omega)$. If $f(x,\omega)$ is such that $\mathbf{E}_\omega|f(x,\omega)|$ exists and grows in the infinity not faster than some polynom of x and random vector η has standard normal distribution, then for $\xi_\theta(x,\eta,\omega) = \frac{1}{\theta}[f(x+\theta\eta,\omega) - f(x,\omega)]\eta$ or $\xi_\theta(x,\eta,\omega) = \frac{1}{2\theta}[f(x+\theta\eta,\omega) - f(x-\theta\eta,\omega)]\eta$, $\theta > 0$, we have $\nabla F^\theta(x) = \mathbf{E}_{\eta\omega}\xi_\theta(x,\eta,\omega)$. The finite difference approximations $\xi_\theta(x,\eta,\omega)$ are unbiased estimates of $\nabla F^\theta(x)$. As in [7], we can call them stochastic mollifier gradient of $F(x)$.*

Definition 3.2 *(See, for example, Rockafellar and Wets [19]). A sequence of functions $\{F^k :\ R^n \longrightarrow \overline{R}\}$ epi-converges to $F:\ R^n \longrightarrow \overline{R}$ relative to $X \subseteq R^n$ if for any $x \in X$*

(i) $\liminf_{k\to\infty} F^k(x^k) \geq F(x)$ for all $x^k \longrightarrow x$, $x^k \in X$;

(ii) $\lim_{k\to\infty} F^k(x^k) = F(x)$ for some sequence $x^k \longrightarrow x$, $x^k \in X$.

The sequence $\{F^k\}$ epi-converges to F if this holds relative to $X = R^n$.

For example, if $g:\ R^n \times R^m \longrightarrow \overline{R}$ is (jointly) lsc at $(\overline{x},\overline{y})$ and is continuous in y at $\overline{y}$, then for any sequence $y^k \longrightarrow \overline{y}$, the corresponding sequence of functions $F^k(\cdot) = g(\cdot,y^k)$ epi-converges to $F(\cdot) = g(\cdot,y)$.

The following important property of epi-convergent functions shows that constrained optimization of a discontinuous function $F(x)$ can be in principle carried out through optimization of approximating epi-convergent functions $F^k(x)$.

Theorem 3.1 *If sequence of functions $\{F^k : R^n \longrightarrow \overline{R}\}$ epi-converges to $F : R^n \longrightarrow \overline{R}$ then for any compact $K \subset R^n$*

$$\lim_{\epsilon\downarrow 0}(\liminf_k(\inf_{K_\epsilon} F^k)) = \lim_{\epsilon\downarrow 0}(\limsup_k(\inf_{K_\epsilon} F^k)) = \inf_K F, \tag{4}$$

where $K_\epsilon = K + \epsilon\mathbf{B}$, $\mathbf{B} = \{x \in R^n \| \|x\| \le 1\}$. If $F^k(x^k_\epsilon) \le \inf_{K_\epsilon} F^k + \delta_k$, $x^k_\epsilon \in K_\epsilon$, $\delta_k \downarrow 0$ as $k \longrightarrow \infty$, then

$$\limsup_{\epsilon\downarrow 0}(\limsup_k x^k_\epsilon) \subseteq argmin_K F, \tag{5}$$

where $(\limsup_k x^k_\epsilon)$ denotes the set X_ϵ of cluster points of the sequence $\{x^k_\epsilon\}$ and $(\limsup_{\epsilon\downarrow 0} X_\epsilon)$ denotes the set of cluster points of the family $\{X_\epsilon,\ \epsilon \in R_+\}$ as $\epsilon \downarrow 0$.

Proof. Note that $(\inf_{K_\epsilon} F^k)$ monotonously increases (non decreases) as $\epsilon \downarrow 0$, hence the same holds for $\liminf_{k\to\infty} \inf_{K_\epsilon} F^k$ and $\limsup_{k\to\infty} \inf_{K_\epsilon} F^k$. Thus limits over $\epsilon \downarrow 0$ in (4) exist.

Let us take arbitrary sequence $\epsilon_m \downarrow 0$, indices k^s_m and points x^s_m such that under fixed m

$$\liminf_k(\inf_{K_{\epsilon_m}} F^k) = \lim_{s\to\infty}(\inf_{K_{\epsilon_m}} F^{k^s_m}) = \lim_{s\to\infty} F^{k^s_m}(x^s_m).$$

Thus

$$\lim_{\epsilon\downarrow 0}(\limsup_k(\inf_{K_\epsilon} F^k)) \ge \lim_{\epsilon\downarrow 0}(\liminf_k(\inf_{K_\epsilon} F^k)) = \lim_{m\to\infty}\lim_{s\to\infty} F^{k^s_m}(x^s_m) = \lim_{m\to\infty} F^{k^{s_m}_m}(x^{s_m}_m)$$

for some indices s_m. By property (i) of epi-convergence $\lim_{m\to\infty} F^{k^{s_m}_m}(x^{s_m}_m) \ge \inf_K F$. Hence

$$\lim_{\epsilon\downarrow 0}(\limsup_k(\inf_{K_\epsilon} F^k)) \ge \lim_{\epsilon\downarrow 0}(\liminf_k(\inf_{K_\epsilon} F^k)) \ge \inf_K F.$$

Let us proof the opposite inequality. Since F is lower semicontinuous, then $F(x) = \inf_K F$ for some $x \in K$. By condition (ii) of epi-convergence there exists sequence $x^k \longrightarrow x$ such that $F^k(x^k) \longrightarrow F(x)$. For k sufficiently large $x^k \in K_\epsilon$, hence $\inf_{K_\epsilon} F^k \le F^k(x^k)$ and

$$\lim_{\epsilon\downarrow 0}(\liminf_k(\inf_{K_\epsilon} F^k)) \le \lim_{\epsilon\downarrow 0}(\limsup_k(\inf_{K_\epsilon} F^k)) \le F(x) = \inf_K F.$$

The proof of (4) is completed.

Now prove (5). Let $x^k_\epsilon \in K_\epsilon$ and $F^k(x^k_\epsilon) \le \inf_{K_\epsilon} F^k + \delta_k$, $\delta_k \downarrow 0$. Denote $X_\epsilon = \limsup_k x^k_\epsilon \subseteq K_\epsilon$. Let $\epsilon_m \downarrow 0$, $x_{\epsilon_m} \in X_{\epsilon_m}$ and $x_{\epsilon_m} \longrightarrow x \in K$ as $m \longrightarrow \infty$. By construction of X_ϵ for each fixed m there exist sequences $x^{k^s_m}_m \longrightarrow x_{\epsilon_m}$ satisfying $F^{k^s_m}(x^{k^s_m}_m) \le \inf_{K_{\epsilon_m}} F^{k^s_m} + \delta_{k^s_m}$, $\delta_{k^s_m} \downarrow 0$ as $s \longrightarrow \infty$. By property (i)

$$F(x_{\epsilon_m}) \le \liminf_s F^{k^s_m}(x^{k^s_m}_m) \le \liminf_s(\inf_{K_{\epsilon_m}} F^{k^s_m}) \le \limsup_k(\inf_{K_{\epsilon_m}} F^k).$$

Due to lower semicontinuity of F and (4) we obtain

$$F(x) \leq \liminf_{m \longrightarrow \infty} F(x_{\epsilon_m}) \leq \liminf_{\epsilon_m \downarrow 0} (\limsup_{k} (\inf_{K_{\epsilon_m}} F^k)) = \inf_{K} F,$$

hence $x \in \mathrm{argmin}_K F$, that proves (5).□

Remark that in Theorem 3.1 we could relax constraint set K in different ways, for instance, if $K = \{x \in R^n | \ G(x) \leq 0\}$ with some lower semicontinuous function $G(x)$, then we could define $K_\epsilon = \{x \in R^n | \ G(x) \leq \epsilon\}$, $\epsilon \geq 0$.

Let us illustrate the result of Theorem 3.1 by the following example.

Example 3.2 *Consider a discontinuous optimization problem*

$$\min_{x \geq 0} \left[F(x) = \begin{cases} 0, & x \leq 0, \\ 1, & x > 0 \end{cases} \right] \tag{6}$$

Let $F^\theta(x)$ be a family of averaged functions for F associated with a family of mollifiers $\psi(y/\theta)$, $\theta > 0$, where mollifier $\psi(\cdot)$ is symmetric with respect to point $y = 0$. Obviously, functions $F^\theta(x)$ epi-converge to F and $\min_{x \geq 0} F^\theta(x) = F^\theta(0) = 1/2$. If we don't relax constraint set $\{x | \ x \geq 0\}$ then optimization of approximate functions $F^\theta(x)$ over set $\{x | \ x \geq 0\}$ leads to a wrong result

$$\lim_{\theta \to 0} \min_{x \geq 0} F^\theta(x) = \frac{1}{2}.$$

The relaxation according to Theorem 3.1 leads to the true optimal value of the problem:

$$\lim_{\theta \to 0} \min_{x \geq -\epsilon} F^\theta(x) = 0$$

and thus

$$\lim_{\epsilon \to 0} (\lim_{\theta \to 0} \min_{x \geq -\epsilon} F^\theta(x)) = 0 = \min_{x \geq 0} F(x).$$

The following statement jointly with Theorem 3.1 shows that the averaged functions can be used for optimization of discontinuous functions.

Theorem 3.2 *(Ermoliev et al. [8]). For any strongly lower semicontinuous, locally integrable function $F : \ R^n \longrightarrow R$, any associated sequence of averaged functions $\{F^{\theta_k}, \ \theta_k \downarrow 0\}$ epi-converges to F.*

Jointly with Propositions 2.1, 2.3 Theorem 3.2 gives sufficient conditions for average functions to epi-converge to original discontinuous expectation function.

A subdifferential calculous for nonsmooth and discontinuous functions can be developed on the basis of their mollifier approximations.

Definition 3.3 *Let function $F: R^n \longrightarrow R$ be locally integrable and $\{F^k := F^{\theta_k}\}$ be a sequence of averaged functions generated from F by means of the sequence of mollifiers $\{\psi^k := \psi_{\theta_k} : R^n \longrightarrow R\}$ where $\theta_k \downarrow 0$ as $k \longrightarrow \infty$. Assume that the mollifiers are such that the averaged functions F^k are smooth (of class C^1). The set of ψ-mollifier subgradients (subdifferential) of F at x is by definition*

$$\partial_\psi F(x) := \limsup_k \{\nabla F(x^k) | \, x^k \longrightarrow x\},$$

i.e. $\partial_\psi F(x)$ consists of the cluster points of all possible sequences $\{\nabla F^k(x^k)\}$ such that $x^k \longrightarrow x$.

The subdifferential $\partial_\psi F(x)$ has the following properties (see Ermoliev, Norkin and Wets [8]):
$\partial_\psi F(x) = \partial F(x)$ for convex functions $F(x)$;
$conv.hull\partial_\psi F(x) = \partial_{Clarke} F(x)$ for locally Lipschitzian function $F(x)$;
$\partial_\psi F(x) = \partial_{Warga} F(x)$ for continuous functions.

Theorem 3.3 *(Ermoliev et al. [8]). Suppose that $F : R^n \longrightarrow R$ is strongly lower semicontinuous and locally integrable. Then for any sequence $\{\psi_{\theta_k}\}$ of smooth mollifiers, we have $0 \in \partial_\psi F(x)$ whenever x is a local minimizer of F.*

4 Necessary optimality conditions

Theorem 3.3 can be used for constrained optimization problems if exact penalties are applicable. Unfortunately, this operation can practically remove some important minimums of the original problem. Consider the following example:

$$\min_{x \geq 0} \left[F(x) = \begin{cases} 1, & x > 0, \\ 0, & x \leq 0. \end{cases} \right].$$

Point $x = 0$ is a reasonable minimum of the problem. We could replace this problem, for example, by the following one:

$$\min_{x \geq 0} \left[\overline{F}(x) = \begin{cases} F(x), & x \geq 0, \\ -x + 1, & x < 0. \end{cases} \right].$$

The penalty function $\overline{F}(x)$ has single discontinuity point $x = 0$, where $\overline{F}$ achieves its global minimum $\overline{F}(0) = 0$. Thus penalty functions may have isolated minimums, which are difficult to discover.

Besides, we also encounter the following difficulties. Consider

$$\min\{ \sqrt[3]{x} | \, x \geq 0\}. \tag{7}$$

In any reasonable definition of gradients the gradient of the function $\sqrt[3]{x}$ at point $x = 0$ equals to $+\infty$. Hence to formulate necessary optimality conditions for such problems and possibly involving discontinuities we need a special notion which incorporates infinite quantities. An appropriate notion is a cosmic vector space $\overline{R^n}$ introduced by Rockafellar and Wets [20]. Denote $R_+ = \{x \in R | \; x \geq 0\}$ and $\overline{R_+} = R_+ \cup \{+\infty\}$.

Definition 4.1 *Define a (cosmic) space $\overline{R^n}$ as a set of pairs $\overline{x} = (x, a)$, where $x \in R^n$, $\|x\| = 1$ and $a \in \overline{R_+}$. All pairs of the form $(x, 0)$ are considered identical and are denoted as $\overline{0}$.*

A topology in the space $\overline{R^n}$ is defined by means of cosmically convergent sequences.

Definition 4.2 *Sequence $(x_k, a_k) \in \overline{R^n}$ is called (cosmically) convergent to an element $(x, a) \in \overline{R^n}$ (denoted c-lim$_{k\to\infty}(x_k, a_k)$) if either* $\lim_k a_k = a = 0$ *or there exist both limits* $\lim_k x_k \in R^n$, $\lim_k a_k \in \overline{R^n}$ *and* $x = \lim_k x_k$, $a = \lim_k a_k \neq 0$, *i.e.*

$$c\text{-}lim_k(x_k, a_k) \quad = \quad \begin{cases} (\lim_k x_k, \lim_k a_k) & if \quad (\lim_k a_k) < +\infty, \\ (\lim_k x_k, +\infty) & if \quad a_k \longrightarrow +\infty, \\ (\lim_k x_k, +\infty) & if \quad a_k = +\infty. \end{cases}$$

Denote

$$\text{c-Limsup}_k(x_k, a_k) = \{(x, a) \in \overline{R^n} | \; \exists \{k_m\} : \; (x, a) = \text{c-lim}_{k\to\infty}(x_{k_m}, a_{k_m})\}.$$

For closed set $K \subset R^n$ denote a tangent cone

$$T_K(x) = \limsup_{\tau} \frac{K - x}{\tau},$$

to the set K at point x, normal cones

$$\hat{N}_K(x) = \{v \in R^n | < v, \omega > \leq 0 \;\text{ for all } \omega \in T_K(x)\},$$

$$N_K(x) = \limsup_{\overline{x}\to x} \hat{N}_K \overline{x},$$

and extended normal cone

$$\overline{N}_K(x) = \{(y, b) \in \overline{R^n} | y \in N_K(x), \|y\| = 1, b \in \overline{R_+}\}.$$

For what follows we need the following closeness property of normal cone mapping $(x, \epsilon) \longrightarrow N_{K_\epsilon}(x)$.

Lemma 4.1 *Let $K_\epsilon = K + \epsilon \times B$, $B = \{x \in R^n | \; \|x\| \leq 1\}$. Then for any sequences $x \longrightarrow \overline{x} \in K$ and $\epsilon \longrightarrow 0$,*

$$\limsup_{x\to\overline{x}, \epsilon\to 0} N_{K_\epsilon}(x) \subseteq N_K(\overline{x}).$$

Proof. For $x \in R^n$ define $y(x) \in K$ such that

$$||y(x) - x|| = \inf_{y' \in K} ||x - y||.$$

Let us show that $T_K(y(x)) \subseteq T_{K_\epsilon}(x)$. Let $w \in T_K(y(x))$, i.e.

$$w = \lim_{\nu \longrightarrow \infty} \frac{y^\nu - y(x)}{\tau_\nu}, \text{ where } y^\nu \in K,\ y^\nu \longrightarrow y(x),\ \ \tau_\nu \longrightarrow 0.$$

Denote $x^\nu = y^\nu + (x - y(x)) \in K_\epsilon$. Then by definition

$$w = \lim_{\nu \longrightarrow \infty} \frac{x^\nu - x}{\tau_\nu} \in T_{K_\epsilon}(x)$$

and thus $T_K(y(x)) \subseteq T_{K_\epsilon}(x)$. This inclusion implies $\hat{N}_{K_\epsilon}(x) \subseteq \hat{N}_K(y(x))$ and $N_{K_\epsilon}(x) \subseteq N_K(y(x))$. Hence

$$\limsup_{x \to \overline{x}, \epsilon \to 0} N_{K_\epsilon}(x) \subseteq \limsup_{x \to \overline{x}} N_K(y(x)) \subseteq N_K(\overline{x}). \Box$$

Corollary 4.1 *For extended normal cones we have the same closeness property,*

$$\limsup_{x \to \overline{x}, \epsilon \to 0} \overline{N}_{K_\epsilon}(x) \subseteq \overline{N}_K(\overline{x}).$$

Remark. We could use another sort of relaxation for set K. Suppose K is convex and is given by an inequality constraint:

$$K = \{x \in R^n |\ G(x) \le 0\}$$

with some convex function $G(x)$. Consider a relaxed set

$$K_\epsilon = \{x \in R^n |\ G(x) \le \epsilon\}.$$

Normal cones to K_ϵ and $K = K_0$ are formed by subdifferentials $\partial G(x)$, $x \in K_\epsilon$, of function G,

$$N_{K_\epsilon}(x) = \begin{cases} \{\lambda \partial G(x) |\ \lambda \ge 0\} & \text{if } G(x) = \epsilon, \\ 0, & \text{if } G(x) < \epsilon, \end{cases} \qquad \epsilon \ge 0.$$

Now closeness property of mapping $(x, \epsilon) \longrightarrow N_{K_\epsilon}$, stated in Lemma 4.1 follows from closeness of subdifferential mapping $x \longrightarrow \partial G(x)$.

Definition 4.3 *Let function $F:\ R^n \longrightarrow R$ be locally integrable and $\{F^k := F^{\theta_k}\}$ be a sequence of averaged functions generated from F by convolution with mollifiers $\{\psi^k := \psi_{\theta_k}:\ R^n \longrightarrow R\}$ where $\theta_k \downarrow 0$ as $k \longrightarrow \infty$. Assume that the mollifiers are such that the averaged functions F^k are smooth (of class C^1). The set of the extended ψ-mollifier subgradients of F at x is by definition*

$$\overline{\partial}_\psi F(x) := c\text{-}Limsup_k \left\{ \left(\frac{\nabla F^k(x^k)}{||\nabla F^k(x^k)||}, ||\nabla F^k(x^k)|| \right) |\ x^k \longrightarrow x \right\},$$

where expression $\frac{\nabla F^k(x^k)}{||\nabla F^k(x^k)||}$ is replaced by any unit vector if $\nabla F^k(x^k) = 0$, i.e. $\overline{\partial}_\psi F(x)$ consists of the cluster points (in cosmic space $\overline{R^n}$) of all possible sequences $\{(\frac{\nabla F^k(x^k)}{||\nabla F^k(x^k)||}, ||\nabla F^k(x^k)||)\}$ such that $x^k \longrightarrow x$. The full (extended) Ψ-mollifier subgradient set is $\partial_\Psi F(x) := \cup_\psi \partial_\psi F(x)$ where ψ ranges over all possible sequences of mollifiers that generate smooth averaged functions.

The extended mollifier subdifferential $\overline{\partial}_{\psi}F(x)$ is always a non-empty closed set in $\overline{R^n}$.

Now we can formulate necessary optimality conditions for constrained discontinuous optimization problem: $\min\{F(x)|\ x \in K\}$, where $F(x)$ may have the form of the expectation.

Theorem 4.1 *Let K be a closed set in R^n. Assume that a locally integrable function F has a local minimum relative to K at some point $x \in K$ and there is a sequence $x^k \in K$, $x^k \longrightarrow x$ with F continuous at x^k and $F(x^k) \longrightarrow F(x)$. Then, for any sequence $\{\psi^k\}$ of smooth mollifiers, one has*

$$- \overline{\partial}_{\psi}F(x) \cap \overline{N}_K(x) \neq \emptyset, \tag{8}$$

where $-\overline{\partial}_{\psi}F(x) = \{(-g,a) \in \overline{R^n}|\ (g,a) \in \overline{\partial}_{\psi}F(x)\}$.

Proof. Let x be a local minimizer of F on K. For a sufficiently small compact neighborhood V of x, define $\phi := F(z) + \|z-x\|^2$. The function ϕ achieves its global minimum on $(K \cap V)$ at x. Consider also the averaged functions

$$\phi^k(z) = \int_{R^n} \phi(y-z)\psi^k(y)dy = F^k(z) + \beta^k(x,z),$$

where

$$F^k(z) = \int_{R^n} F(y-z)\psi^k(y)dy, \beta^k(x,z) = \int |y-z-x|^2\psi^k(y)dy.$$

In [8] it is shown that (i) functions ϕ^k are continuously differentiable, (ii) they epi-converge to ϕ relative to $K \cap V$ and (iii) their global minimums z^k on $K \cap V$ converge to x as $k \longrightarrow \infty$. For sufficiently large k the following necessary optimality condition is satisfied:

$$-\nabla F^k(z^k) = n(z^k) \in N_K(z^k), \quad z^k \in K.$$

If $\nabla F^{k_m}(z^{k_m}) = 0$ for some $\{z^{k_m} \longrightarrow x\}$ then also $\overline{0} \in \overline{\partial}_{\psi}F(x)$ and $\overline{0} \in \overline{N}_K(x)$. If $\nabla F^{k_m}(z^{k_m}) \longrightarrow g \neq 0$ for some $\{z^{k_m} \longrightarrow x\}$ then

$$-\frac{\nabla F^{k_m}(z^{k_m})}{\|\nabla F^{k_m}(z^{k_m})\|} \longrightarrow -\frac{g}{\|g\|} \in N_K(x),$$

and $(\frac{g}{\|g\|}, \|g\|) \in \overline{\partial}_{\psi}F(x)$, $\quad (-\frac{g}{\|g\|}, \|g\|) \in \overline{N}_K(x)$. If $\limsup_k \|\nabla F^k(z^k)\| = +\infty$ then for some $\{z^{k_m} \longrightarrow x\}$

$$-\frac{\nabla F^{k_m}(z^{k_m})}{\|\nabla F^{k_m}(z^{k_m})\|} \longrightarrow -g \in N_K(x),$$

and $(g, +\infty) \in \overline{\partial}_{\psi}F(x)$, $\quad (-g, +\infty) \in \overline{N}_K(x)$.□

Next proposition shows that optimality conditions are also satisfied for limits X' of some local minimizers x_ϵ of relaxed problems $\min\{F(x)|\ x \in K_\epsilon = K + \epsilon\mathbf{B}\}$.

Proposition 4.1 *Let x_ϵ be a local minimizer such that there exists sequence $x_\epsilon^k \longrightarrow x_\epsilon$, $x_\epsilon^k \in K_\epsilon$ with F continuous at x_ϵ^k and $F(x_{\epsilon^k}) \longrightarrow F(x_\epsilon)$ as $k \longrightarrow \infty$. Assume $x_{\epsilon_m} \longrightarrow x$ for some $\epsilon_m \downarrow 0$ as $m \longrightarrow \infty$. Then (8) is satisfied at x.*

Proof follows from Theorem 4.1 and closeness of (extended) mollifier subdifferential mapping $x \longrightarrow \overline{\partial}_\psi F(x)$ and (extended) normal cone mapping $(x, \epsilon) \longrightarrow \overline{N}_{K_\epsilon}(x)$.

Proposition 4.2 *If F is strongly lsc and the constraint set K is compact then the set X^* of points, satisfying necessary optimality condition (8), is nonempty and contains at least one global minimizer of F in K.*

Proof. Construct a sequence of differentiable averaged functions F^k epi-converging to F (what is possible by Theorem 3.2). Relax constraint set K, i.e. define $K_\epsilon = K + \epsilon \times B$, where $B = \{x| \ \|x\| \leq 1\}$. Find a global minimizer x_ϵ^k of F^k over K_ϵ. For x_ϵ^k we have necessary optimality condition (see Rockafellar and Wets [21]):

$$-\nabla F^k(x_\epsilon^k) \in N_{K_\epsilon}(x_\epsilon^k).$$

We can assume that $x_\epsilon^k \longrightarrow y_\epsilon \in K_\epsilon$. From here it follows

$$-\overline{\partial} F(y_\epsilon) \cap \overline{N}_{K_\epsilon}(y_\epsilon) \neq \emptyset.$$

Now let $y_\epsilon \longrightarrow y \in K$, $\epsilon \longrightarrow 0$. By Theorem 3.1 $y \in X^*$. Then by closeness of mappings $\overline{\partial} F(\cdot)$ and $\overline{N}_{K_\epsilon}(\cdot)$ we finally obtain

$$\overline{\partial} F(y) \cap \overline{N}_K(y) \neq \emptyset. \square$$

Now let us come back to problem (7) and show how the developed theory resolves the exposed difficulties.

Example 4.1 *Consider again an optimization problem:* $\min\{\sqrt[3]{x}| \ x \geq 0\}$. *Then we have*

$$\overline{\partial}_\psi \sqrt[3]{x}|_{x=0} = (+1, +\infty), \ \ \overline{N}_{x\geq 0}(0) = \cup_{a \in \overline{R_+}}(-1, a)$$

and thus

$$-\overline{\partial}_\psi \sqrt[3]{x}|_{x=0} \cap \overline{N}_{x \geq 0}(0) = (-1, +\infty) \neq \emptyset.$$

5 On numerical optimization procedures

Theorem 4.1 and Propositions 4.1, 4.2 immediately give at least the following idea for the approximate solution of problem (1), (2). Let us fix a small smoothing parameter θ and a small constraint relaxation parameter ϵ, choose a mollifier $\psi_\theta(\cdot) = \psi(\cdot/\theta)$ and instead of original discontinuous optimization problem consider a relaxed smoothed optimization problem:

$$\min[F^\theta(x)| \;\; x \in K_\epsilon], \tag{9}$$

where $F^\theta(x)$ is defined by (3). Then stochastic gradient method to solve (9) has the form:

x^0 is an arbitrary starting point;

$$x^{k+1} = \Pi_{K_\epsilon}(x^k - \rho_k \xi_\theta(x^k)), \quad k = 0, 1, \ldots;$$

where $\mathbf{E}\{\xi_\theta(x^k)|x^k\} = \nabla F^\theta(x^k)$, Π_{K_ϵ} denotes the orthogonal projection operator on the set K_ϵ, positive step multipliers ρ_k satisfy conditions

$$\sum_{k=0}^{\infty} \rho_k = +\infty, \qquad \sum_{k=0}^{\infty} \rho_k^2 < +\infty, \tag{10}$$

Vectors $\xi_\theta(x^k)$ can be called stochastic mollifiers gradients.

The convergence of such kind stochastic gradient method to a stationary set

$$X_\epsilon^\theta = \{x \in K_\epsilon | \; -\nabla F^\theta(x) \in N_{K_\epsilon}(x)\}$$

follows from results [5]. Now coming to the limit in $\theta \longrightarrow 0$ and then in $\epsilon \longrightarrow 0$ we see that limit points $[\limsup_\epsilon(\limsup_\theta X_\epsilon^\theta)]$ satisfy necessary optimality condition (8).

References

[1] Antonov G.E. and Katkovnik V.Ya. (1970), Filtration and smoothing in extremum search problems for multivariable functions, Avtomatika i vychislitelnaya tehnika, N.4, Riga. (In Russian).

[2] Archetti F. and Betrò B. (1975), Convex programming via stochastic regularization, *Quaderni del Dipartimento di Ricerca Operativa e Scienze Statistiche*, N 17, Università di Pisa.

[3] Batuhtin,B.D. (1994), On one approach to solving discontinuous extremal problems, Izvestia AN Rossii. Tehnicheskaia kibernetika (Communications of Russian Academy of Sciences. Technical Cybernetics), No. 3, pp.37-46. (In Russian).

[4] Batuhtin,B.D. and Maiboroda L.A. (1984), *Optimization of discontinuous functions*, Moscow, Nauka. (In Russian).

[5] Dorofeev P.A. (1986), A scheme of iterative minimization methods, U.S.S.R. Comput. Math. Math. Phys., Vol. 26, No. 2, pp.131-136. (In Russian).

[6] Ermoliev Yu. and Gaivoronski A. (1992), Stochastic programming techniques for optimization of discrete event systems, *Annals of Operations Research*, Vol. 39, pp.120-135.

[7] Ermoliev Yu.M. and Norkin V.I. (1995), On Nonsmooth Problems of Stochastic Systems Optimization, Working Paper WP-95-096, Int. Inst. for Appl. Syst. Anal., Laxenburg, Austria.

[8] Ermoliev Yu.M, Norkin V.I. and Wets R.J-B. (1995), The minimization of semi-continuous functions: Mollifier subgradients, SIAM J. Contr. and Opt., No.1, pp.149-167.

[9] Gupal A.M. (1977), On a method for the minimization of almost differentiable functions, Kibernetika, No. 1, pp.114-116. (In Russian, English translation in: Cybernetics, Vol. 13, N. 1).

[10] Gupal A.M. (1979), *Stochastic methods for solving nonsmooth extremal problems*, Naukova dumka, Kiev. (In Russian).

[11] Gupal A.M. and Norkin V.I. (1977), An algorithm for minimization of discontinuous functions, Kibernetika, No. 2, 73-75. (In Russian, English translation in: Cybernetics, Vol. 13, N. 2).

[12] Hasminski R.Z. (1965), Application of random noise in optimization and recognition problems, Problemy peredachi informatzii, Vol. 1, N. 3. (In Russian).

[13] Katkovnik V.Ya. (1976), *Linear Estimates and Stochastic Optimization Problems*, Nauka, Moscow. (In Russian).

[14] Katkovnik V.Ya. and Kulchitsky Yu. (1972), Convergence of a class of random search algorithms, *Automat. Remote Control*, No. 8, pp. 1321-1326. (In Russian).

[15] Kreimer J. and Rubinstein R.Y. (1992), Nondifferentiable optimization via smooth approximation: general analytical approach, Annals of Oper. Res., Vol. 39, pp.97-119.

[16] Mayne D.Q. and Polak E. (1984), Nondifferentiable optimization via adaptive smoothing, J. of Opt. Theory and Appl., Vol. 43, pp.601-613.

[17] Mikhalevich V.S., Gupal A.M. and Norkin V.I. (1987), *Methods of nonconvex optimization*, Nauka, Moscow. (In Russian).

[18] Nikolaeva N.D. (1974), On an algorithm for solving convex programming problems, Econom. i Matem. Methody, Vol. 10, pp. 941-946 (In Russian).

[19] Rockafellar R.T. and Wets R.J-B. (1984), Variational systems, an introduction, in: *Multifunctions and Integrands*, G.Salinetti, ed., Lecture Notes in Mathematics 1091, Springer-Verlag, Berlin, pp.1-54.

[20] Rockafellar R.T. and Wets R.J-B. (1991), Cosmic convergence, in: *Optimization and Nonlinear Analysis*, eds. A.Ioffe, M.Marcus and S.Reich, Pitman Research Notes in Mathematics Series 244, Longman Scientific & Technical, Essex, U.K., pp. 249-272.

[21] Rockafellar R.T. and Wets R.J-B. (1995), *Variational Analysis*, a monograph to be published in Springer-Verlag.

[22] Rubinstein R.Y. (1983), Smoothed functionals in stochastic optimization, Math. Oper. Res, Vol. 8, pp.26-33.

[23] Yudin D.B. (1965), Qualitative methods for analysis of complex systems I, Izvestia AN SSSR, Tehnich. Kibernetika, No. 1. (In Russian).

[24] Zaharov V.V. (1970), Integral smoothing method in multi-extremal and stochastic problems, Izvestia AN SSSR, Tehnich. Kibernetika, No. 4. (In Russian).

[25] Warga J. (1975), Necessary conditions without differentiability assumptions in optimal control, J. Diff. Equations, Vol. 15, pp.41-61.