Schlossplatz 1
A-2361 Laxenburg, Austria

E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

**Interim Report** **IR-03-033**

# Risk and Extended Expected Utility Functions: Optimization Approaches

*Yuri Ermoliev(ermoliev@iiasa.ac.at)*
*Vladimir Norkin(norkin@d130.icyb.kiev.ua)*

**Approved by**

*Leen Hordijk(hordijk@iiasa.ac.at)*
Director, IIASA

August 2003

***Interim Reports*** on work of the International Institute for Applied Systems Analysis receive only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute, its National Member Organizations, or other organizations supporting the work.

## Abstract

The proper analysis of polices under uncertainties has to deal with "hit-or-miss" type situations by using appropriate risk functions, which can also be viewed as so-called extended expected utility functions. Formally this often requires the solution of dynamic stochastic optimization problems with discontinuous indicator functions of such events as ruin, underestimating costs and overestimating benefits. The available optimization techniques, in particular formulas for derivatives of risk functions, may not be applicable due to explicitly unknown probability distributions and essential discontinuities. The aim of this paper is to develop a solution technique by smoothing the risk function over certain parameters, rather than over decision variables as in the classical distribution (generalized functions) theory. For smooth approximations we obtain gradients in the form of expectations of stochastic vectors which can be viewed as a form of stochastic gradients for the original risk function. We pay special attention to optimization of risk functions defined on trajectories of discrete time stochastic processes with stopping times, which is critically important for analyzing regional vulnerability against catastrophes.

Y. Ermoliev
V. Norkin

# Contents

# Risk and Extended Expected Utility Functions: Optimization Approaches

*Yuri Ermoliev(ermoliev@iiasa.ac.at)*[*]
*Vladimir Norkin(norkin@d130.icyb.kiev.ua)*[**]

## 1 Introduction

The proper analysis of polices under uncertainties has to deal with "hit-or-miss" type situations by using appropriate risk functions (see, e.g., discussion in [13], [15]), which can also be viewed as so-called extended expected utility functions. Formally this often requires the solution of dynamic stochastic optimization problems with discontinuous indicator functions of such events as ruin, underestimating costs and overestimating benefits. The available optimization techniques, in particular formulas for derivatives of risk functions, may not be applicable due to explicitly unknown probability distributions and essential discontinuities. The aim of this paper is to develop a solution technique by smoothing the risk function over certain parameters, rather than over decision variables as in the classical distribution (generalized functions) theory. For smooth approximations we obtain gradients in the form of expectations of stochastic vectors which can be viewed as a form of stochastic gradients for the original risk function. We pay special attention to optimization of risk functions defined on trajectories of discrete time stochastic processes with stopping times, which is critically important for analyzing regional vulnerability against catastrophes (see, e.g., [10]-[13]).

Any decision involving uncertainties leads to multiple outcomes with possible favorable and unfavorable consequences. For example, investments in conventional or new technologies may lead to considerable profits under favorable scenarios. But the cost of unfavorable scenarios, e.g., due to global warming, may be environmental degradation and economic stagnation. The notion of risk functions is used to represent tradeoffs and interdependencies between different outcomes and decisions, which often leads to specific stochastic optimization (STO) problems. We discuss this in some details in Section 2. In particular, Section 2 outlines connections between the so-called chance constrains, ruin (survival) probabilities, Value-at-Risk (VaR), and Conditional-Value-at-Risk (CVaR), which are important for applications in quality (e.g., air) control, reliability theory, insurance, finance, catastrophic risk management, and sustainable developments (land use, energy). The standard stochastic optimization models are formulated by using expectations

$$F(x) = \mathbf{E}f(x,\omega) = \int f(x,\omega)d\mathbf{P}(\omega) \qquad (1)$$

of some goal functions $f(x,\omega)$ for a given decision $x$ and variables $\omega$ which are determined by environment affecting the consequences of $x$. It is assumed that $x$ belongs to a feasible

[*]International Institute for Applied Systems Analysis (IIASA), Laxenburg, Austria
[**]Institute of Cybernetics of the Ukrainian Academy of Sciences, 03680 Kiev, Ukraine

set $X \subseteq R^n$ and $\omega$ is an elementary event (scenario) of a probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Such a space gives a description of all possible uncertainties $\Omega$ and all observable events (possible scenarios) $A \in \mathcal{F}$ with associated probability measure $\mathbf{P}$.

There are various shortcomings in representation (1). One of them is connected with the analysis of low probability events, for example, $A(x) = \{\omega : f(x,\omega) \geq c\}$ for large $c$. The sources of risk are often characterized as the violation of certain constraints or regulations, such as constraints on permissible loads, stresses, demands and supplies, etc. Therefore we can think of all favorable and unfavorable events for a given $x$ as a partitioning of $\Omega$, $\Omega = \cup_{i=1}^m A_i$, where each element $A_i$ is given as

$$A_i(x) = \{\omega \in \Omega | g^i(x,\omega) \leq 0\}, \quad i = 1, \ldots, m,$$

with some, in general vector valued function $g^i(x,\omega)$. Here we assume that number $m$ is fixed and does not depend on $x$ and $\omega$. Function (1) can be rewritten as

$$F(x) = \sum_{i=1}^{m} \mathbf{E}\{f(x,\omega)|A_i\}\mathbf{P}(A_i(x)),$$

where $\mathbf{E}\{\cdot|\cdot\}$ is the symbol of the conditional expectation.

If $A_i(x)$ is a so-called "low probability – high consequence" (catastrophic) event, the contribution of the corresponding term into an overall expectation function may be not sensible. Therefore we need indicators which are more selective to unfavorable or favorable low probability situations, such as, e.g., conditional expectations, i.e. the function

$$\mathbf{E}\{f(x,\omega)|A_i(x)\} = \frac{\mathbf{E}[f(x,\omega)I\{A_i(x)\}]}{\mathbf{E}I\{A_i(x)\}},$$

where $I\{A\}$ is the indicator function of $A$:

$$I\{A\} = \begin{cases} 1 & if\ \omega \in A, \\ 0 & otherwise. \end{cases}$$

From a formal point of view various important models can be described by using expectations of the type

$$\mathbf{E}[f(x,\omega)I\{g(x,\omega) \leq 0\}] \tag{2}$$

for some random function $f$ and random vector function $g$. The optimization of function (2) can be viewed as a basic subproblem to deal with a vast variety of applications. For example, the case of $f \equiv 1$ reduces to the probability function

$$\mathbf{E}I\{g(x,\omega) \leq 0\} = \mathbf{P}\{g(x,\omega) \leq 0\}, \tag{3}$$

which is often the object of optimization or a constraint function in the study of ruin, safety and survival of systems.

Although formally an optimization problem involving functions of type (2) has the form (1), there is a number peculiarities requiring new concepts. A main issue is the implicit dependence of the integrand on the policy variable $x$. In particular, it restricts the straightforward use of the sample mean approximations. Among other peculiarities there is a possible discontinuity of function (2) and its singularity with respect to low probability events (see, for example, discussion in [14], [15]).

Formulas for explicit differentiation of probability (and similar) functions and corresponding optimization procedures are available in [31], [34], [30], [36], [23], [24], [20], [35], and reviewed in [21]. According to these results gradients of probability functions are represented as surface and/or volume integrals. Both representations require an explicit form of probability density function for the random variable $\omega$ (and even its derivatives in case of volume integral) that is not always available, and even the probability space may be unknown. Besides, the probability function (and other indicators, see [15]) can easily be nonsmooth, and then the available explicit differentiation formulas are certainly not applicable.

**Example 1.1** (nonsmooth probability function). Assume that $g(x,\omega) = x + \omega$, where $x \in R$ and $\omega$ is uniformly distributed on the interval $[0,1]$. Then probability function

$$\mathbf{P}\{0 \le g(x,\omega) \le 1\} = \begin{cases} 0, & |x| \ge 1, \\ 1 - |x|, & |x| \le 1, \end{cases}$$

is nonsmooth at $x = 0, \pm 1$.

In this paper we develop another approach (close to, but different from [17]) to optimization of possibly nonsmooth risk functions of type (2), (3). Namely, we first uniformly approximate these functions by smoothing them over certain parameters, rather than over decision variables as in [17]. Then for approximations we obtain (by interchange of differentiation and expectation operators) explicit formulas for gradients in the form of expectations of stochastic gradients. We pay special attention to risk functions defined on trajectories of discrete time stochastic processes which may depend on stopping times. The basical "mollified", stochastic gradients for the original problem obtained are used for designing an iterative stochastic optimization procedure.

Section 2 shows that many important performance functions of a stochastic system with vector of outputs $f(x,\omega)$ can be expressed in the form $U(x) = \mathbf{E}u(f(x,\omega))$, where $u(\cdot)$ is some possibly discontinuous "utility" function. These functions can be called risk functions or extended expected utility functions. In particular, some functions depending on the stopping time, e.g., expected shortfall of risk processes, can be expressed in such form. In Section 3 we study conditions of continuity and Lipschitz continuity of risk function $U(x)$. We analyse randomly perturbed system $f(x,\omega) + \epsilon\eta$ and corresponding utility $U_\epsilon(x) = \mathbf{E}u(f(x,\omega) + \epsilon\eta)$, where $\eta$ is an independent of $\omega$ random variable, and $\epsilon$ is a small perturbation parameter. Functions $U_\epsilon(x)$ can be viewed as the result of kernel smoothing of the function $U(x,y) = \mathbf{E}U(f(x,\omega) + y)$ over parameter $y$. It appears that functions $U_\epsilon(x)$ are smooth (or generalized differentiable) in $x$ and we establish in Section 4 formulas for their (sub)gradients in the form of expectation of stochastic gradients $\xi_\epsilon(x,\omega)$, $\mathbf{E}\xi_\epsilon(x,\omega) \in \partial U_\epsilon(x)$. We also establish conditions of uniform (in $x$) convergence of $U_\epsilon(x)$ to $U(x)$ as $\epsilon \to 0$. Section 5 analyses necessary optimality conditions for minimization of $U(x)$ on a compact set $X$ in terms of the so-called mollifier subdifferential $\partial_m U(x)$. The subdifferential $\partial_m U(x)$ is constructed as a set of all cluster points of (sub)gradients $\partial U_{\epsilon_\nu}(x^\nu)$ when $x^\nu \to x$, $\epsilon_\nu \to 0$. It appears that under a certain regularity condition on function $U(x,y)$ the subdifferential $\partial_m U(x)$ is included in Clarke's subdifferential $\partial U(x)$ of $U(x)$. Section 6 outlines the solution procedure for an arising limit extremal problem, i.e. minimization of $U(x)$ on $X$ by using (sampled) stochastic quasigradients $\xi_{\epsilon_\nu}(x,\omega)$ of $U_{\epsilon_\nu}(x)$ at iteration $\nu$, $\epsilon_\nu \longrightarrow 0$ for $\nu \longrightarrow 0$. Section 6 concludes with some general remarks on the so-called integrated risk management.

## 2 Examples

Let us discuss some important problems described by functions of type (2).

### 2.1 Chance constrained problem

The problem

$$f(x) \longrightarrow \min_x \tag{4}$$

subject to chance constraint

$$\mathbf{P}\{g(x,\omega) \leq 0\} \geq 1-\beta, \tag{5}$$

can be approximated by the following simple recourse problem (with penalty parameter $N$):

$$F(x) = f(x) + N\mathbf{E}\max\{0, g(x,\omega)\} \longrightarrow \min_x, \tag{6}$$

where $\mathbf{E}\max\{0, g(x,\omega)\} = \mathbf{E}g(x,\omega)I\{g(x,\omega) \geq 0\}$. In particular, in papers [10], [11] such replacement was used for insurance portfolio optimization under constraints on the probability of insolvency. The random term $\max\{0, g(x,\omega)\}$ can be interpreted as ex-post borrowing for positive excess losses $g(x,\omega)$ and $N$ is the price for such borrowing. It appears that problems (4)-(5) and (6) are closely connected (see discussion in [11], [18], [37]). For instance, according to [10], [11] an optimal value of (4)-(5) can be approximated by an optimal value of (6) with some large penalty parameter $N(\beta)$.

### 2.2 Value at risk and conditional value at risk

An important special case of problem (4) – (6) for financial applications (see, for example, [5]) is the minimization of the value at risk (or VaR, $\beta$-quantile, see, for example, [20]):

$$Q_\beta(x) = \min\{y|\ \mathbf{P}\{g(x,\omega) \leq y\} = 1-\beta\} \longrightarrow \min_x. \tag{7}$$

Instead of (7) we can again solve a penalty problem (6):

$$F(x,y) = y + N\mathbf{E}\max\{0, g(x,\omega) - y\} \longrightarrow \min_{x,y}, \tag{8}$$

which is a special case of stochastic minmax problems [9]. From the optimal conditions for this problem follows that the optimal $y$ (for a given $x$) is $1/N$ quantile of random variable $g(x,\omega)$ (see, e.g., [9], p. 416).

The Conditional Value at Risk (CVaR) is defined as

$$C_\beta(x) = \frac{1}{\beta}\mathbf{E}g(x,\omega)I\{g(x,\omega) - Q_\beta(x) \geq 0\}. \tag{9}$$

The minimization of $C_\beta(x)$ under natural assumptions [32] is equivalent to the following convex optimization problem

$$y + \frac{1}{\beta}\mathbf{E}\max\{0, g(x,\omega) - y\}\} \longrightarrow \min_{y,x}, \tag{10}$$

i.e., CVaR minimization (9) has the form of (8) with $N = \frac{1}{\beta}$.

Problem (8) has the following economic interpretation. Assume again that $g(x,\omega)$ represents stochastic excess losses depending on decision variable $x$ and stochastic parameter $\omega$. These losses are covered by ex-ante borrowing $y$ (for the price 1) and ex-post borrowing $\max\{0, g(x,\omega) - y\}$ (for price $N$). These provide more flexibility compared with the control only by decisions $x$ as in (4), (5) (see also [12] for more general formulations).

### 2.3 Risk process

Consider a classical discrete time risk process (see, for instance, [1], [2], [11]) describing the evolution of reserves $R_t(x)$ of an insurance company:

$$R_{t+1}(x) = R_0 + \Pi_t(x) - C_t(x), \quad t = 0, 1, \ldots, T, \tag{11}$$

where $R_0 \geq 0$ is the initial capital of the company, $\Pi_t(x)$ are aggregate premiums and $C_t(x)$ are random aggregate outcomes up to time $t$, e.g., claims, taxes, dividends, etc., and $x$ is a decision vector. Functions $\Pi_t(x)$, $C_t(x)$ are assumed to be continuously differentiable (or at least generalized differentiable [26], [16]) with respect to $x$. They are random but the dependence on random factors is not indicated for the simplicity of notation. Components of vector $x$ may include parameters of portfolio of assets and insurance contracts (see [11] for details).

The problem is to optimize the performance of a company over time horizon $[0, T]$ which is described by a number of performance functions, for example:

random *stopping time*

$$\tau(x) = \max\{t \in [0, T] : \ R_s(x) \geq 0 \ \ \forall\, s, 0 \leq s < t\}, \tag{12}$$

which is called *default time* when $\tau(x) < T$ or $R_{\tau(x)=T}(x) < 0$;

the *probability of insolvency (ruin)* on time interval $[0, T]$:

$$\begin{aligned} \psi_T(x) &= \mathbf{P}\{\tau(x) < T,\ R_T(x) < 0 \ \ if\ \tau(x) = T\} \\ &= 1 - \mathbf{P}\{R_t(x) \geq 0, \ \ 0 \leq t \leq T\} \\ &= \mathbf{E}(1 - I\{R_t(x) \geq 0, \ \ 0 \leq t \leq T\}); \end{aligned} \tag{13}$$

*partial expected profit* (on survived trajectories)

$$F_T(x) = \mathbf{E} R_T(x) I\{R_t(x) \geq 0,\ 0 \leq t \leq T\}; \tag{14}$$

*expected shortfall* (negative depth of insolvency)

$$\begin{aligned} H_T(x) &= \mathbf{E}\min\{0, R_{\tau(x)}(x)\} \\ &= \mathbf{E}\sum_{t=0}^{T} R_t(x) I\{R_\tau \geq 0,\ 0 \leq \tau < t; R_t(x) < 0\}; \end{aligned} \tag{15}$$

*stability criterion*

$$\begin{aligned} S_T(x) &= \mathbf{P}\{R_t(x) \geq (1-\epsilon)\mathbf{E}R_t(x),\ 0 \leq t \leq T\} \\ &= \mathbf{E} I\{R_t(x) \geq (1-\epsilon)\mathbf{E}R_t(x),\ 0 \leq t \leq T\}, \quad 0 < \epsilon \leq 1. \end{aligned} \tag{16}$$

The stability criterion estimates the probability that the company does not operate much worse than the average trajectory. The structure of this criterion is similar to (13). Let us note that function $\tau(x)$ may be discontinuous in $x$. This may cause discontinuities of all functions (13) – (16).

**Assumption P.** *For any fixed* $x \in X$, $t \in [0, T]$ *and* $c, \delta \geq 0$

*(i)* $\mathbf{P}\{R_t(x) = c\} = 0$;

*(ii)* $\mathbf{P}\{R_t(x) \in [c - \delta, c + \delta]\} \leq L\delta$ *for some constant* $L > 0$.

We show in section 3 that under assumption P(i) the above indicators are continuous, and under assumption P(ii) they are Lipschitz continuous in $x$.

### 2.4 Discontinuous utility functions

With the explicit introduction of uncertainties and risks the overall performance of a decision $x$ becomes a tradeoff between different socioeconomic and environmental indicators (costs, benefits, incomes, damages) and indicators of risks. The classical example is the mean-variance efficient strategies providing a tradeoff between expected returns and the variance. Unfortunately, the concept of the mean-variance efficient strategies may be misleading and even wrong for nonnormal probability distributions (especially for catastrophic risks) which require more sophisticated risk indicators and corresponding concepts of robust strategies. More precisely, in practice a given decision $x$ results in different outcomes $f(x,\omega) = (f_1(x,\omega), \ldots, f_m(x,\omega))$ affected by some uncertain (random) variables $\omega$. Formally, the overall performance of $x$ can be often summarized in the form of an expectation function

$$U(x) = \mathbf{E}u(f_1(x,\omega), \ldots, f_m(x,\omega)),$$

where $u(\cdot)$ is a "utility" function defined on $f \in R^m$. The mean-variance efficient solutions maximizing $\mathbf{E}f(x,\omega) - N\mathbf{E}[f(x,\omega) - \mathbf{E}f(x,\omega)]^2$, $N > 0$, can also be obtained from the maximization of the following type of function:

$$\max_{x,y} \mathbf{E}\left[f(x,\omega) - N\left(f(x,\omega) - y\right)^2\right].$$

This representation convexifies the problem for $f(x,\omega) = -|f(x,\omega)|$, where $|f(\cdot,\omega)|$ is a convex (cost) function.

Traditionally the utility function is assumed to be continuous and differentiable. It is easy to see that all risk functions discussed in this section can be represented in the same form but with nonsmooth and even discontinuous utility functions. For example, if $u(\cdot)$ is the indicator function for the event $\{f \in R^m | \ f \geq c\}$, then

$$U(x) = \mathbf{P}\{f(x,\omega) \geq c\}. \qquad (17)$$

If

$$u(f_1, f_2) = f_1 I\{f_2 \geq 0\} = \begin{cases} 0, & f_2 < 0, \\ f_1, & f_2 \geq 0 \end{cases},$$

then we obtain function (2)

$$U(x) = \int_{f_2(x,\omega)\geq 0} f_1(x,\omega)\mathbf{P}(d\omega). \qquad (18)$$

In the particular case $f_1(x,\omega) \equiv f_2(x,\omega) = f(x,\omega)$

$$U(x) = \mathbf{E}\max\{0, f(x,\omega)\mathbf{P}(d\omega).$$

Functions $U(x)$ with nonsmooth and discontinuous integrand $u(\cdot)$ can be used as a unified concept to analyze quite different risk management problems. In short, we can call such $U(x)$ the risk functions and $u(f)$ the sample risk function or (extended) utility function. We can call $U(x)$ also extended expected utility function. Note that although indicators (13), (14), (15) are defined through stopping time $\tau(x)$, they can also be expressed in the form $\mathbf{E}u(R_0, R_1(x), \ldots, R_T(x))$ with some discontinuous function $u(\cdot)$.

## 3 Risk functions

Consider the following risk function given in the form of extended expected utility

$$U(x) = \mathbf{E}u(f(x,\omega)), \tag{19}$$

where $f: R^n \times \Omega \longrightarrow R^m$ is a continuous in $x$ and measurable in $\omega$ vector function, $u: R^m \longrightarrow R^1$ is a Borel (extended utility) function, $\mathbf{E}$ (or $\mathbf{E}_\omega$) denotes mathematical expectation over measure $\mathbf{P}$ (or $\mathbf{P}_\omega$) on $\Omega$. In general, as we discussed in Section 2, function $u(\cdot)$ may be discontinuous on a set $D \subset R^m$.

**Proposition 3.1** (Continuity of risk function). *Assume that*

*(i) $f(x,\omega)$ is a.s. continuous at point $x$,*

*(ii) $\mathbf{P}\{f(x,\omega) \in D\} = 0$,*

*(iii) $u(f(y,\omega)) \le M(\omega)$ for all $y$ from a vicinity of $x$ with integrable function $M(\omega)$.*

*Then function $U(x)$ is continuous.*

The proposition follows from Lebesgue's dominance convergence theorem.

Denote

$$D_\delta = \{y \in R^m | \; dist(y,D) \le \delta\}, \quad dist(y,D) = \inf_{z \in D} \|y - z\|.$$

**Proposition 3.2** (Lipschitz continuity). *Assume that*

*(i) $u(\cdot)$ is uniformly Lipschitzian in any ball outside the discontinuity set $D$;*

*(ii) $f(x,\omega)$ are a.s. Lipschitzian in $x \in X$ uniformly in $\omega$;*

*(iii) $\mathbf{P}\{f(x,\omega) \in D_\delta\} \le C\delta$ for all $x \in X$, $y \in R^m$, $\delta > 0$ and some constant $C$;*

*(iv) $u(f(x,\omega) + y) \le M$ for all $x \in X$, $y \in R^m$ and some constant $M$.*

*Then function $U(x,y) = \mathbf{E}u(f(x,\omega) + y)$ is Lipschitz continuous in $(x,y) \in X \times R^m$, and hence risk function $U(x) = U(x,0)$ is Lipschitzian in $x \in X$.*

**Proof.** Let $L_u$ and $L_f$ be Lipschitz constants for $u$ and $f$, respectively. For given $x_1$, $x_2$, $y_1$, $y_2$ define $x_\lambda = x_1 + \lambda(x_2 - x_1)$, $y_\lambda = y_1 + \lambda(y_2 - y_1)$ with $\lambda \in [0,1]$, $\delta = L_f\|x_2 - x_1\| + \|y_2 - y_1\|$, $\Omega_{2\delta} = \{\omega \in \Omega | \; (f(x_1,\omega) + y_1) \in D_{2\delta}\}$, Obviously, $\|f(x_\lambda,\omega) + y_\lambda - f(x_1,\omega) - y_1\| \le L_f\lambda\|x_2 - x_1\| + \lambda\|y_2 - y_1\| \le \delta$. Note that if $(f(x_1,\omega) + y_1) \in D_{2\delta}$, then $(f(x_2,\omega) + y_2) \in D_{3\delta}$, and if $(f(x_1,\omega) + y_1) \overline{\in} D_{2\delta}$, then $(f(x_\lambda,\omega) + y_\lambda) \overline{\in} D_\delta$ for any $\lambda \in [0,1]$. We have

$$\begin{aligned} U(x_2,y_2) - U(x_1,y_1) &= \left(\int_{\Omega_{2\delta}} + \int_{\Omega\setminus\Omega_{2\delta}}\right) [u(f(x_2,\omega) + y_2) \\ &\qquad\qquad - u(f(x_1,\omega) + y_1)]\mathbf{P}(d\omega) \\ &\le M\mathbf{P}\{(f(x_2,\omega) + y_2) \in D_{3\delta}\} \\ &\quad + M\mathbf{P}\{(f(x_1,\omega) + y_1) \in D_{2\delta}\} \\ &\quad + \int_{\Omega\setminus\Omega_{2\delta}} L_u\|f(x_2,\omega) + y_2 - f(x_1,\omega) - y_1\|\mathbf{P}(d\omega) \\ &\le (5MC + L_u)(L_f\|x_2 - x_1\| + \|y_2 - y_1\|). \; \square \end{aligned}$$

If function $u(\cdot)$ is discontinuous then it can be approximated in different ways by continuous functions $u_\epsilon(\cdot)$ for some parameter $\epsilon$ in such a way that $u_\epsilon(y) \longrightarrow u(y)$ as $\epsilon \longrightarrow 0$ for all $y \overline{\in} D$. Then function $U(x)$ is approximated by functions

$$U_\epsilon(x) = \mathbf{E}u_\epsilon(f(x,\omega)). \tag{20}$$

**Proposition 3.3** (Convergence of approximations). *Assume that*

*(i) $\lim_{\delta \to 0} \mathbf{P}\{f(x,\omega) \in D_\delta\} = 0$, pointwise (uniformly) in $x \in X$;*

*(ii) $\lim_{\epsilon \to 0} u_\epsilon(z) = u(z)$, uniformly in $z \overline{\in} D_\delta$ for any $\delta > 0$;*

*(iii) $u(f(x,\omega))$ and $u_\epsilon(f(x,\omega))$ are bounded by an integrable in square function $M(\omega)$ uniformly in $x \in X$ and $\epsilon > 0$.*

*Then $\lim_{\epsilon \to 0} U_\epsilon(x) = U(x)$ pointwise (uniformly) in $x \in X$.*

**Proof.** Define $\Omega_1 = \{\omega \in \Omega |\ f(x,\omega) \in D_\delta\}$ and $\Omega_2 = \Omega \setminus \Omega_1$. Then

$$\begin{aligned}
|U(x) - U_\epsilon(x)| &\leq \left(\int_{\Omega_1} + \int_{\Omega_2}\right) |u(f(x,\omega)) - u_\epsilon(f(x,\omega))| \mathbf{P}(d\omega) \\
&\leq 2\int_{\Omega_1} M(\omega)\mathbf{P}(d\omega) + \int_{\Omega_2} |u(f(x,\omega)) - u_\epsilon(f(x,\omega))| \mathbf{P}(d\omega) \\
&\leq 2\left(\int_{\Omega} M^2(\omega)\mathbf{P}(d\omega)\right)^{1/2} \mathbf{P}\{f(x,\omega) \in D_\delta\} \\
&\quad + \sup_{y \in R^m \setminus D_\delta} |u(y) - u_\epsilon(y)|. \qquad (21)
\end{aligned}$$

The first term on the right-hand side of (21) can be made arbitrarily small by choosing $\delta$ small enough due to (i), (iii). For a given $\delta$ the second term on the right-hand side of (21) can be made arbitrary small by choosing $\epsilon$ small enough due to (ii).□

One way to construct approximations $U_\epsilon(x)$ is to consider stochastically disturbed performance indicators

$$f_\epsilon(x,\omega,\eta) = f(x,\omega) + \epsilon\eta,$$

where $\epsilon$ is a small positive parameter, $\eta \in R^m$ is a random vector independent of $\omega$ with density $K(\cdot)$. The corresponding disturbed risk function takes the form

$$\begin{aligned}
U_\epsilon(x) &= \mathbf{E}_\eta \mathbf{E}_\omega u(f_\epsilon(x,\omega,\eta)) \\
&= \mathbf{E}_\omega \mathbf{E}_\eta u(f(x,\omega) + \epsilon\eta) \\
&= \mathbf{E}_\omega u_\epsilon(f(x,\omega)),
\end{aligned}$$

where $u_\epsilon(f)$ is the so-called smoothed (or mollified) utility function

$$u_\epsilon(y) = \mathbf{E}_\eta u(y + \epsilon\eta) = \frac{1}{\epsilon^m}\int u(z) K\left(\frac{z-y}{\epsilon}\right) dz$$

used in kernel density estimation (see, for example, [7]), in probability function optimization (see [22], [27]) and in nonsmooth optimization (see [25], [17] and references therein).

**Proposition 3.4** (Convergence of mollified utilities at continuity points). *Let $u(x)$ be a real-valued Borel measurable function on $R^m$, $K(x)$ be a bounded, integrable, real valued density function on $R^m$ and one of the following holds*

*(i) $u(\cdot)$ is bounded on $R^m$;*

*(ii) $K(\cdot)$ has a compact support;*

*(iii) $\|y\|K(y) \longrightarrow 0$ as $\|y\| \longrightarrow \infty$, where $\|\cdot\|$ denotes the Euclidean norm on $R^m$.*

*Then $u_\epsilon(y) \longrightarrow u(y)$ as $\epsilon \longrightarrow 0$ at any continuity point of $u(\cdot)$.*

The statement of the proposition under assumption (i) can be found in [3], and under (ii), (iii) it is available in [6].

**Proposition 3.5** (Uniform convergence outside discontinuity points ). *Assume that*

*(i) $u(\cdot)$ is a Borel function with closed set $D$ of discontinuity points;*

*(ii) density $K(\cdot)$ has a compact support.*

*Then $u_\epsilon(y)$ uniformly converges to $u(y)$ outside arbitrary vicinity of $D$.*

**Proof.** We have to show that $u_{\epsilon_k}(y^k) \longrightarrow u(y)$ for any sequences $\epsilon_k \longrightarrow 0$ and $y^k \longrightarrow y \overline{\in} D$. From here a uniform convergence of $u_\epsilon(\cdot)$ to $u(\cdot)$ follows in any compact $A$ such that $A \cap D = \emptyset$. Represent

$$u_{\epsilon_k}(y^k) = \int_{S(K)} u(y^k + \epsilon_k z) K(z) dz,$$

where $S(K) = \{z|\ K(Z) > 0\}$ denotes support of density $K(\cdot)$. Since $D$ is closed and $y\overline{\in}D$ there exists $\delta > 0$ such that $\{z|\ \|z-y\| \leq \delta\} \cap D = \emptyset$. In the $V_\delta = \{z|\ \|z-y\| \leq \delta\}$ function $u(\cdot)$ is continuous and thus bounded. For any $z \in S(K)$ by (ii) $\lim_k(y^k + \epsilon_k z) = y$. Thus by Lebesgue dominance convergence theorem

$$\begin{aligned} \lim_k u_{\epsilon_k}(y^k) &= \textstyle\int_{S(K)} \lim_k u(y^k + \epsilon_k z) K(z) dz \\ &= \textstyle\int_{S(K)} u(y) K(z) dz = u(y). \square \end{aligned}$$

**Example 3.1** (Partial smoothing). If in (18) we disturb only function $f_2$ then

$$\begin{aligned} U_\epsilon(x) &= \mathbf{E}_\eta \mathbf{E}_\omega f_1(x,\omega) I_{f_2(x,\omega)+\epsilon\eta \geq 0} \\ &= \mathbf{E}_\omega f_1(x,\omega) \mathbf{E}_\eta I_{f_2(x,\omega)+\epsilon\eta \geq 0} \\ &= \mathbf{E}_\omega f_1(x,\omega)(1 - \mathcal{F}(-f_2(x,\omega)/\epsilon)), \end{aligned}$$

where $\mathcal{F}$ is a cumulative distribution function of random variable $\eta$.

**Proposition 3.6** (Uniform convergence under partial smoothing). *Assume that conditions of Proposition 3.5 are fulfilled and*

*(i) function* $\mathbf{E}|f_1(x,\omega)|$ *is bounded on* $X$*;*

*(ii)* $\mathbf{P}\{|f_2(x,\omega)| \leq \delta\} \longrightarrow 0$ *as* $\delta \longrightarrow 0$ *uniformly in* $x \in X$.

*Then* $U_\epsilon(x)$ *converges to* $U(x)$ *uniformly in* $x \in X$.

**Proof.** For arbitrary numbers $C$, $\delta$ estimate the difference

$$\begin{aligned} |U_\epsilon(x) - U(x)| &\leq \mathbf{E}|f_1(x,\omega)| \cdot |1 - \mathcal{F}(-f_2(x,\omega)/\epsilon) - I_{f_2(x,\omega)\geq 0}| \\ &= \left( \int_{f_1(x,\omega) \geq C} + \int_{f_1(x,\omega) < C} \right) |f_1(x,\omega)| \times \\ &\quad \times |1 - \mathcal{F}(-f_2(x,\omega)/\epsilon) - I_{f_2(x,\omega)\geq 0}| \mathbf{P}(d\omega) \\ &\leq \int_{f_1(x,\omega) \geq C} |f_1(x,\omega)| \mathbf{P}(d\omega) \\ &\quad + C\mathbf{E}|1 - \mathcal{F}(-f_2(x,\omega)/\epsilon) - I_{f_2(x,\omega)\geq 0}| \\ &\leq \int_{f_1(x,\omega) \geq C} |f_1(x,\omega)| \mathbf{P}(d\omega) + C\mathbf{P}\{|f_2(x,\omega)| \leq \delta\}. \\ &\quad + C \sup_{|y| \geq \delta} |1 - \mathcal{F}(-y/\epsilon) - I_{y \geq 0}| \end{aligned} \tag{22}$$

The first term on the right-hand side of (22) is made arbitrarily small by taking $C$ sufficiently large by (i). The second term for given $C$ is made small by taking $\delta$ sufficiently small by (ii). Given $C$ and $\delta$ the third term can be made small by taking $\epsilon$ small by Proposition 3.5. $\square$

**Example 3.2** (Smoothing probability function ). Consider probability function

$$U(x) = \mathbf{P}_\omega\{f_1(x,\omega) \leq 0, \ldots, f_m(x,\omega) \leq 0\}$$

and its approximation

$$U_\epsilon(x) = \mathbf{P}_{\omega,\eta}\{f_1(x,\omega) + \epsilon\eta_1 \leq 0, \ldots, f_m(x,\omega) + \epsilon\eta_m \leq 0\},$$

where $\eta = (\eta_1, \ldots, \eta_m)$, $\epsilon > 0$ is a random vector variable with the cumulative distribution function $\mathcal{F}$ and distribution $\mathbf{P}_\eta$, $\mathbf{P}_{\omega,\eta}$ is the product of measures $\mathbf{P}_\omega$ and $\mathbf{P}_\eta$. Then

$$U_\epsilon(x) = \mathbf{E}_\omega \mathcal{F}\left( -\frac{1}{\epsilon} f_1(x,\omega), \ldots, -\frac{1}{\epsilon} f_m(x,\omega) \right).$$

We can also approximate by using $\eta_i = \eta$, $i = 1, \ldots, m$, where random variable $\eta$ has the cumulative distribution function $\mathcal{F}$. Then

$$\begin{aligned} U_\epsilon(x) &= \mathbf{P}_\omega \mathbf{P}_\eta \{\eta \le -f_1(x,\omega)/\epsilon, \ldots, \eta \le -f_m(x,\omega)/\epsilon\} \\ &= \mathbf{P}_\omega \mathbf{P}_\eta \{\eta \le -\tfrac{1}{\epsilon} \max_{1 \le i \le m} f_i(x,\omega)\} \\ &= \mathbf{E}_\omega \mathcal{F}\left(-\tfrac{1}{\epsilon} \max_{1 \le i \le m} f_i(x,\omega)\right). \end{aligned}$$

If functions $u_\epsilon$ and $f(x,\omega)$ in (20) are continuously (or generalized) differentiable, then compound function $u_\epsilon(f(x,\omega))$ is also continuously (generalized) differentiable with (sub)differential $\partial_x u_\epsilon(f(x,\omega))$, which can be calculated by a chain rule (see [16], [26] for the nondifferentiable case).

If (sub)differential $\partial_x u_\epsilon(f(x,\omega))$ is majorized by an integrable (Lipschitz) constant, $\xi_\epsilon(x,\omega)$ is a measurable selection of $\partial_x u_\epsilon(f(x,\omega))$, then function $F_\epsilon(x)$ is also (generalized) differentiable with (sub)differential

$$\partial U_\epsilon(x) = \mathbf{E} \partial_x u_\epsilon(f(x,\omega)) \ni \mathbf{E} \xi_\epsilon(x,\omega). \tag{23}$$

For optimization of $F_\epsilon(x)$ one can apply specific stochastic gradient methods (see Section 6) based on samples of $\xi_\epsilon(x,\omega)$ with $\epsilon \longrightarrow 0$. For a given $\epsilon$ it is also possible to use the sample mean optimization methods.

# 4 Stochastic smoothing of risk processes

To optimize risk functions we can apply mollifiers [17]) over decision variables $x$. Similarly, we can mollify risk process over some parameters, for example, initial state. In addition to smoothing effects, which are usually weaker than in the first case, the significant advantage of the parametric smoothing is the possibility to obtain fast statistical estimators of the risk functions and their derivatives [12].

Beside standard risk process (11) consider a process with random initial capital $R_0 + \epsilon\eta$ [12]:

$$Q_t(x,\epsilon) = R_0 + \epsilon\eta + \Pi_t(x) - C_t(x) = R_t(x) + \epsilon\eta, \quad 0 \le t \le T, \tag{24}$$

where $\eta$ is an independent of all claims $C_t(x)$ one-dimensional random variable with a continuously differentiable distribution function

$$\mathcal{F}(y) = \mathbf{P}_\eta\{\eta < y\},$$

$\epsilon$ is a small (smoothing) parameter ($\epsilon \to 0$).

We can think of (24) as risk process (11) with disturbed initial values $R_0$ or $R_1(x)$. Through dynamic equation (24) the disturbance $\epsilon\eta$ is transferred to further values $R_t(x)$, $t \ge 1$, of the process. Similarly we can independently disturb all $R_t(x)$, $0 \le t \le T$, and interpret these disturbances as the presence of insignificant lines of business of the insurance company.

In subsection 2.3 we introduced important performance functions of process (11): probability of insolvency $\psi_T(x)$, partial expected profit $F_T(x)$, expected shortfall $H(x)$. Under assumption P(i) they are continuous, and under P(ii) they are Lipschitz continuous. Here we consider the same performance functions also for the disturbed process (24). Under assumption P(ii) by the results of section 3 (Propositions 3.3, 3.5) these approximates converge uniformly in $x$ to the original undisturbed performance functions as the disturbance goes to zero. The smoothing effects enable us to derive their subdifferentials.

### 4.1 The probability of ruin

Define measure $\mathbf{P}$ as the product of $\mathbf{P}_\omega$ and $\mathbf{P}_\eta$, $\mathbf{P} = \mathbf{P}_\eta \times \mathbf{P}_\omega$. Then the probability of ruin till moment $T$ of the disturbed risk process $\{Q_t(x,\epsilon) = R_t(x) + \epsilon\eta,\ t = 0, 1, \dots, T\}$ is

$$\begin{aligned}
\psi_T(x,\epsilon) &= 1 - \mathbf{P}\{Q_t(x,\epsilon) \geq 0, \quad 0 \leq t \leq T\} \\
&= 1 - \mathbf{P}\{\eta \geq -R_t(x)/\epsilon,\ \ 0 \leq t \leq T\} \\
&= 1 - \mathbf{P}\{\eta \geq \max_{0 \leq t \leq T} -R_t(x)/\epsilon\} \\
&= \mathbf{P}\{\eta < -\min_{0 \leq t \leq T} R_t(x)/\epsilon\} \\
&= \mathbf{E}_\omega \mathcal{F}\{-\min_{0 \leq t \leq T} R_t(x)/\epsilon\},
\end{aligned}$$

with a subdifferential (see Clarke [4], Theorems 2.3.9, 2.3.12, 2.7.2)

$$\partial\psi_T(x,\epsilon) = -\mathbf{E}_\omega \mathcal{F}'\{-R_{t^*}(x)/\epsilon\} \cdot \nabla R_{t^*}(x)/\epsilon|_{t^* \in t^*(x)}\,, \tag{25}$$

where $t^*(x) = argmin_{0 \leq t \leq T} R_t(x)$, and functions $R_t(x)$ are assumed continuously differentiable in $x$.

### 4.2 Partial expected profit

Partial expected profit at time $T$ (on survived disturbed trajectories) is given by the formula:

$$\begin{aligned}
F_T(x,\epsilon) &= \mathbf{E}_\omega \mathbf{E}_\eta Q_T(x,\epsilon) I\{Q_t(x,\epsilon) \geq 0, \quad 0 \leq t \leq T\} \\
&= \mathbf{E}_\omega \mathbf{E}_\eta R_T(x) I\{Q_t(x,\epsilon) \geq 0,\ \ 0 \leq t \leq T\} \\
&\quad + \mathbf{E}_\omega \mathbf{E}_\eta \epsilon\eta I\{Q_t(x,\epsilon) \geq 0,\ \ 0 \leq t \leq T\} \\
&= \mathbf{E}_\omega R_T(x)(1 - \mathcal{F}(-\min_{0 \leq t \leq T} R_t(x)/\epsilon)) \\
&\quad + \epsilon \mathbf{E}_\omega \mathcal{H}(-\min_{0 \leq t \leq T} R_t(x)/\epsilon),
\end{aligned}$$

with subdifferential

$$\begin{aligned}
\partial F_T(x,\epsilon) &= \mathbf{E}_\omega (1 - \mathcal{F}(-R_{t^*}(x)/\epsilon)) \nabla R_T(x) \\
&\quad + \mathbf{E}_\omega \left(\mathcal{F}'(-R_{t^*}(x)/\epsilon)\right. \\
&\quad \left. - \epsilon \mathbf{E}_\omega \mathcal{H}'(-R_{t^*}(x)/\epsilon)\right) \nabla R_{t^*}(x)/\epsilon|_{t^* \in t^*(x)}\,,
\end{aligned} \tag{26}$$

where $I\{A\}$ is the indicator function for event $A$, $\mathcal{H}(y) = \int_{\eta \geq y} \eta d\mathcal{F}(\eta), t^*(x) = argmin_{0 \leq t \leq T} R_t(x)$.

### 4.3 Expected shortfall

Consider the expected shortfall

$$H_T(x,\epsilon) = \mathbf{E}_{\omega\eta} \min\{0, Q_{\tilde{\tau}(x)}(x,\epsilon)\},$$

$$\tilde{\tau}(x) = \max\{t \in [0,T] : Q_s(x,\epsilon) \geq 0,\ 0 \leq s < t\ \}.$$

Function

$$\begin{aligned}
H_T(x,\epsilon) &= \mathbf{E}_{\omega\eta} \textstyle\sum_{t=0}^T \min\{0, Q_t(x,\epsilon) I(\min_{0 \leq \tau < t} Q_\tau \geq 0)\} = \\
&= \mathbf{E}_{\omega\eta} \textstyle\sum_{t=0}^T Q_t(x,\epsilon) I(-\min_{0 \leq \tau < t} R_\tau(x)/\epsilon \leq \eta \leq -R_t(x)/\epsilon) \\
&= \mathbf{E}_\omega \textstyle\sum_{t=0}^T R_t(x)[\mathcal{F}(-R_t(x)/\epsilon) - \mathcal{F}(-\min_{0 \leq \tau < t} R_\tau/\epsilon)] \\
&\quad + \epsilon \mathbf{E}_\omega \textstyle\sum_{t=0}^T \mathcal{J}(-\min_{0 \leq \tau < t} R_\tau(x)/\epsilon, -R_t(x)/\epsilon),
\end{aligned}$$

where $\mathcal{J}(y,z) = \int_y^z \eta d\mathcal{F}(\eta)$, $\min_{0 \leq \tau < t}\{\cdot\}|_{t=0} = +\infty$, is a generalized differentiable function (see [26], [25]) as constructed from continuously differentiable functions by means of min, max, composition and expectation operations, with subdifferential

$$
\begin{aligned}
\partial H_T(x,\epsilon) &= \mathbf{E}_\omega \sum_{t=0}^{T}[\mathcal{F}(-R_t(x)/\epsilon) - \mathcal{F}(-\min_{0\le\tau<t} R_\tau(x)/\epsilon)]\nabla R_t(x) \\
&\quad + \mathbf{E}_\omega \sum_{t=0}^{T} R_t(x)[\partial\mathcal{F}(-R_t(x)/\epsilon) - \partial\mathcal{F}(-\min_{0\le\tau<t} R_\tau(x)/\epsilon)] \\
&\quad + \epsilon\mathbf{E}_\omega \sum_{t=0}^{T} \partial\mathcal{J}(-\min_{0\le\tau<t} R_\tau(x)/\epsilon, -R_t(x)/\epsilon) \\
&= \mathbf{E}_{\omega\eta}\nabla R_t(x)|_{t=\tilde{\tau}(x)} \\
&\quad + \mathbf{E}_\omega \sum_{t=0}^{T} R_t(x)[\partial\mathcal{F}(-R_t(x)/\epsilon) - \partial\mathcal{F}(-\min_{0\le\tau<t} R_\tau(x)/\epsilon)] \\
&\quad + \epsilon\mathbf{E}_\omega \sum_{t=0}^{T} \partial\mathcal{J}(-\min_{0\le\tau<t} R_\tau(x)/\epsilon, -R_t(x)/\epsilon). \qquad (27)
\end{aligned}
$$

If functions $\{R_t(x),\ 0 \le t \le T\}$ are continuously differentiable with respect to decision variables $x$, then approximations $\psi_T(x,\epsilon)$, $F_T(x,\epsilon)$, $H_T(x,\epsilon)$ are generalized differentiable and thus can be optimized by the method of Section 6, based on Monte Carlo simulations only of trajectories of the process $\{R_t(x),\ 0 \le t \le T\}$.

# 5 Optimality conditions

In this section we give necessary conditions of local optimality (Proposition 5.4) and sufficient conditions for stationarity (Corollary 5.2) of the risk functions given in the form of extended expected utility function $U(x)$. We derive them in terms of mollifier subdifferential $\partial_m U(x)$. We basically follow the approach from [17], the difference consists in the following. In [17] the original (may be discontinuous) deterministic function was approximated by a family of continuously differentiable functions through random disturbances of decision variables. Here we deal with the implicitly known expected utility function depending on some parameters with possibly discontinuous integrand. We are able to approximate this function by smoothing over parameters only by a family of (maybe nonsmooth) Lipschitzian functions.

## 5.1 Mollifier subdifferential

Consider a family of Lipschitzian functions $U_\epsilon(x)$ that approximate a continuous function $U(x)$ on $X$ as $\epsilon \to 0$. Denote $\partial U_\epsilon(x)$ and $N_X(x)$ Clarke's subdifferential of $U_\epsilon(x)$ and normal cone to set $X$ at point $x \in X$, respectively (see [4], [33]).

**Definition 5.1.** For the approximation family $M = \{U_\epsilon(x),\ \epsilon > 0\}$ (similar to [17]) define *mollifier subdifferential*

$$\partial_m U(x) = \mathrm{L}imsup_{x^\nu \to x, \epsilon_\nu \searrow 0} \partial U_{\epsilon_\nu}(x^\nu), \qquad (28)$$

where the right-hand side consists of all cluster points of all such sequences $g^\nu \in \partial U_{\epsilon_\nu}(x^\nu)$ that $x^\nu \to x$, $\epsilon_\nu \searrow 0$. Let us also define *mollifier derivative in direction* $l$

$$U'_m(x;l) = \limsup_{x^\nu \to x, \epsilon_\nu \searrow 0} U^o_{\epsilon_\nu}(x^\nu; l),$$

where

$$U_\epsilon^o(x;l) = \limsup_{\tilde{x}\to x, \lambda \searrow 0} \frac{1}{\lambda}[U_\epsilon(\tilde{x}+\lambda l) - U_\epsilon(\tilde{x})]$$

is Clarke's generalized derivative of $U_\epsilon(\cdot)$ at point $x$ in direction $l$.

Obviously, mapping $\partial_m U(x)$ is closed, mollifier derivative $U'_m(x)$ is convex and positively homogeneous in $l$. Define a convex set

$$G_m(x)\} = \{g | \langle g, l\rangle \leq U'_m(x;l) \ \ \forall l \in R^n\}.$$

**Proposition 5.1** (Characterization of the mollifier subdifferential). *Let family $M = \{U_\epsilon(x),\ \epsilon > 0\}$ be uniformly locally Lipschitzian. Then*

$$co\{\partial_m U(x)\} = G_m(x),$$

*where $co\{\cdot\}$ denotes a convex hull.*

**Proof.** Fix an arbitrary point $x$ and direction $l$. By definition there exist such sequences $\epsilon_\nu \to 0$, $x^\nu \to x$ that $U^o_{\epsilon_\nu}(x^\nu;l) \longrightarrow U'_m(x;l)$. By definition of Clarke's subdifferential $U^o_\epsilon(x^\nu;l) = \langle g^\nu, l\rangle$ for some $g^\nu \in \partial U(x^\nu)$. Without loss of generality we can assume that $g^\nu \to g \in \partial_m U(x)$. Then

$$U'_m(x;l) = \langle g, l\rangle \leq \sup_{g\in\partial_m U(x)} \langle g, l\rangle,$$

and hence $G_m(x) \subseteq co\partial_m U(x)$.

Let us now prove the opposite relations. By definition for any $g \in \partial_m U(x)$ there exist such sequences $\epsilon_\nu \to 0$, $x^\nu \to x$ and $g^\nu \to g$ that $g^\nu \to \partial U_{\epsilon_n u}(x^\nu)$. By definition of Clarke's generalized gradients $\langle g^\nu, l\rangle \leq U^o(x^\nu;l)$. Then

$$\langle g, l\rangle \leq \limsup_\nu U^o(x^\nu;l) \leq U'_m(x;l),$$

and hence $\sup_{g\in\partial_m U(x)}\langle g, l\rangle \leq U'_m(x;l)$ and $\partial_m U(x) \subseteq G_m(x)$. □

## 5.2 Regularity

Beside (19) define function

$$U(x,y) = \mathbf{E}_\omega u(f(x,\omega)+y),$$

where parameter $y \in R^m$. Obviously, $U(x) = U(x,0)$ and $U_\epsilon(x) = \mathbf{E}_\eta U(x, \epsilon\eta)$. Under conditions of Proposition 3.2 function $U(x,y)$ is Lipschitz continuous in $(x,y)$.

**Definition 5.2** Define derivative of $U(x,y)$ at point $(x,y)$ in direction $(l_x, l_y) \in R^n \times R^m$ as

$$U'(x,y;l_x,l_y) = \lim_{\lambda\to+0} \frac{1}{\lambda}[U(x+\lambda l_x, y+\lambda l_y) - U(x,y)]$$

(if the limit exists), Clarke's *generalized derivative* [4] as

$$U^o(x,y;l_x,l_y) = \limsup_{\tilde{x}\to x, \tilde{y}\to y, \lambda\searrow 0} \frac{1}{\lambda}[U(\tilde{x}+\lambda l_x, \tilde{y}+\lambda l_y) - U(\tilde{x},\tilde{y})],$$

and *partial generalized derivative* of $U$ at $(x,y)$ in the direction $l_x \in R^n$ as

$$U^o_x(x,y;l_x) = \limsup_{\tilde{x}\to x, \lambda\searrow 0} \frac{1}{\lambda}[U(\tilde{x}+\lambda l_x, y) - U(\tilde{x},y)].$$

**Definition 5.3** Function $U(x,y)$ is called *Clarke regular* if for any $(l_x, l_y)$

$$U^o(x,y;l_x,l_y) = U'(x,y;l_x,l_y),$$

and *regular in* $x$ if for any $l_x$

$$U^o(x,y;l_x,0) = U^o_x(x,y;l_x).$$

**Proposition 5.2** (Calculous for regular in $x$ functions). *(i) Regular by Clarke function $U(x,y)$ is regular in $x$.*

*(ii) $U(x,y)$ is regular in $x$ iff $-U(x,y)$ is regular in $x$ .*

*(iii) Convex and concave in $(x,y)$ functions $U(x,y)$ are regular in $x$.*

**Proof.** (i) The statement follows from inequalities:

$$U'(x,y;l_x,0) \le U^o_x(x,y;l_x) \le U^o(x,y;l_x,0).$$

(ii) Suppose $-U(x,y)$ is regular in $x$ and show that $U(x,y)$ is regular in $x$. By Clarke [4], prop.2.1.1(c), $U^o(x,y;l_x,l_y) = (-U)^o(x,y;-l_x,-l_y)$, and similarly we have $U^o_x(x,y;l_x) = (-U)^o_x(x,y;-l_x)$. Let $x^\nu \to x$, $\lambda_\nu \to +0$ are such that

$$U^o_x(x,y;l_x) = \lim_{\nu\to+\infty} \frac{1}{\lambda_\nu}[U(x^\nu + \lambda_\nu l_x, y) - U(x^\nu, y)].$$

Then

$$\begin{array}{rcl} U^o_x(x,y;l_x) & = & \lim_{\nu\to+\infty} \frac{1}{\lambda_\nu}[-U((x^\nu+\lambda_\nu l_x) - \lambda_\nu l_x, y) - (-U(x^\nu + \lambda_\nu l_x, y))] \\ & \le & \limsup_{\tilde{x}\to x, \tilde{y}\to y, \lambda\searrow 0} \frac{1}{\lambda}[-U(\tilde{x} - \lambda l_x, \tilde{y}) - (-U(\tilde{x},\tilde{y}))] \\ & = & (-U)^o(x,y;-l_x,0) \end{array}$$

By regularity $(-U)^o(x,y;-l_x,0) = (-U)^o_x(x,y;-l_x)$. Thus

$$\begin{array}{rcl} U^o_x(x,y;l_x) & = & (-U)^o(x,y;-l_x,0) = U^o(x,y;l_x,0) \\ & = & (-U)^o_x(x,y;-l_x) = U^o_x(x,y;l_x). \end{array}$$

(iii) Since convex functions are Clarke regular ([4], prop. 2.3.6(b)), then by (i) they are regular in $x$, and by this and (ii) concave functions are also regular in $x$. □

**Example 5.1** (Regularity of integral functionals). If $U(x,y) = \mathbf{E}_\omega u(f(x,\omega)+y)$ and functions $u(f(\cdot,\omega)+\cdot)$ are Lipschitzian and Clarke regular with integrable Lipschitz constant, then $U(x,y)$ is also Lipschitzian and Clarke regular in $(x,y)$ (see [4], prop.2.7.2), and by Proposition 5 .2 is regular in $x$ .

Function $U(x,y) = \mathbf{E}_\omega u(f(x,\omega)+y)$ can be Lipschitzian and regular in $x$ even for discontinuous utilities $u(\cdot)$.

**Example 5.2** (A regular probability function). Let the mapping $f(x,\omega) = \phi(x) + \omega$, where vector random variable $\omega \in R^m$ has Lipschitzian c.d.f. $\mathcal{F}$ with constant $L_\mathcal{F}$, continuously differentiable mapping $\phi(\cdot): R^n \to R^m$ is such that equation $\phi(x) = y$ has a solution for any $y$. Then function

$$U(x,y) = \mathbf{P}\{\phi(x) + \omega + y \le 0\} = \mathcal{F}(-\phi(x) - y)$$

is regular in $x$.

Indeed, let sequences $\lambda_\nu \to 0$, $x^\nu \to x$, $y^\nu \to 0$ be such that generalized derivative in direction $l_x$

$$U^o(x,0;l_x,0) = \lim_{\nu\to\infty} \frac{1}{\lambda_\nu}[U(x^\nu + \lambda_\nu l_x, y^\nu) - U(x^\nu, y^\nu)].$$

From equations $\phi(\tilde{x}) = \phi(x^\nu) + y^\nu$ let us find solutions $\tilde{x}^\nu \to x$. Then

$$\begin{array}{rcl} U^o(x,0;l_x,0) & = & \lim_{\nu\to\infty} \frac{1}{\lambda_\nu}[\mathcal{F}(-\phi(x^\nu+\lambda_\nu l_x) - y^\nu) - \mathcal{F}(-\phi(x^\nu) - y^\nu)] \\ & \le & \limsup_{\nu\to\infty} \frac{1}{\lambda_\nu}[\mathcal{F}(-\phi(\tilde{x}^\nu+\lambda_\nu l_x)) - \mathcal{F}(-\phi(\tilde{x}^\nu))] \\ & & +\limsup_{\nu\to\infty} \frac{1}{\lambda_\nu}[\mathcal{F}(-\phi(x^\nu+\lambda_\nu l_x) - y^\nu) - \mathcal{F}(-\phi(\tilde{x}^\nu+\lambda_\nu l_x))] \\ & \le & U^o_x(x,0;l_x) \\ & & +\limsup_{\nu\to\infty} \frac{1}{\lambda_\nu}[\mathcal{F}(-\phi(x^\nu) - \lambda_\nu\phi'(x^\nu)l_x) - o(\lambda_\nu\|l_x\|) - y^\nu) \\ & & \qquad -\mathcal{F}(-\phi(\tilde{x}^\nu) - \lambda_\nu\phi'(\tilde{x}^\nu)l_x) - \tilde{o}(\lambda_\nu\|l_x\|))] \\ & \le & U^o_x(x,0;l_x) \\ & & +\limsup_{\nu\to\infty} L_{\mathcal{F}}(\phi'(x^\nu) - \phi(\tilde{x}^\nu))l_x \\ & & +\limsup_{\nu\to\infty} \frac{1}{\lambda_\nu} L_{\mathcal{F}}[|o(\lambda_\nu\|l_x\|)| + |\tilde{o}(\lambda_\nu\|l_x\|)|] \\ & = & U^o_x(x,0;l_x). \Box \end{array}$$

**Example 5.3** (Quasiconcavity and regularity of probability functions). Let function $f(x,\omega)$, $x \in R^n$, $\omega \in R^m$, be quasi-convex in $(x,\omega)$ and measure $\mathbf{P}_\omega$ $\alpha$-concave, $\alpha > -\infty$, (for instance, 0-concave, i.e. logarithmically concave). Then probability function $\mathbf{P}_\omega\{f(x,\omega) + y \le 0\}$ is $\alpha$-concave in $(x,y)$ and hence function $U(x,y) = 1 - \mathbf{P}_\omega\{f(x,\omega) + y \le 0\}$ is regular in $(x,y)$ (see [27], [28] for details).

## 5.3 Optimality conditions

**Proposition 5.3** (Mollifier subdifferential as a subset of Clarke's subdifferential). *Assume that function $U(x,y)$ is Lipschitzian and regular in $x$, mollifier subdifferential $\partial_m U(x)$ of function $U(x) = U(x,0)$ is defined through functions $U_\epsilon(x) = \mathbf{E}_\eta U(x,\epsilon\eta)$ by (28), where $\eta$ is m-dimensional random vector with bounded support. Then*

$$\partial_m U(x) \subseteq \partial U(x),$$

*where $\partial U(x)$ is Clarke's subdifferential of Lipschitzian function $U(x)$.*

A similar relation for mollifier subdifferential was established in [17] (see also [33], par. 9.67, for sharper result).

**Corollary 5.1** (Convergence of subgradients). *Under conditions of Proposition 5.3*

$$\partial U_{\epsilon_\nu}(x^\nu) \longrightarrow \partial U(x), \;\; \forall \;\; \epsilon_\nu \to 0, \;\; x^\nu \to x. \tag{29}$$

**Proof of Proposition 5.3.** Fix any point $x$ and direction $l$. By definition of Clarke's generalized derivative there exist such sequences $\lambda_\nu \to 0$, $\tilde{x}^\nu \to x$ that

$$U^o_\epsilon(x;l) = \lim_{\nu\to+\infty} \frac{1}{\lambda}[U_\epsilon(\tilde{x}^\nu + \lambda_\nu l) - U_\epsilon(\tilde{x}^\nu].$$

Taking into account that $U_\epsilon(x) = \mathbf{E}_\eta U(x,\epsilon\eta)$ we obtain

$$\begin{array}{rcl} U^o_\epsilon(x;l) & = & \lim_{\nu\to+\infty} \frac{1}{\lambda_\nu}[U_\epsilon(\tilde{x}^\nu+\lambda_\nu l) - U_\epsilon(\tilde{x}^\nu)] \\ & \le & \mathbf{E}_\eta \limsup_{\nu\to+\infty} \frac{1}{\lambda_\nu}[U_\epsilon(\tilde{x}^\nu+\lambda_\nu l,\epsilon\eta) - U_\epsilon(\tilde{x}^\nu,\epsilon\eta)] \\ & \le & \mathbf{E}_\eta \limsup_{\tilde{x}\to x,\lambda\to 0} \frac{1}{\lambda}[U_\epsilon(\tilde{x}+\lambda l,\epsilon\eta) - U_\epsilon(\tilde{x},\epsilon\eta)] \\ & = & \mathbf{E}_\eta U^o_x(x,\epsilon\eta;l). \end{array}$$

By definition of mollifier derivative $U'_m(x;l)$ there exist such sequences $\epsilon_\nu \to 0$, $x^\nu \to x$ that $U'_m(x;l) = \lim_{\nu\to+\infty} U^o_{\epsilon_\nu}(x^\nu;l)$. Thus we obtain

$$\begin{array}{rcl} U'_m(x;l) & = & \lim_{\nu\to+\infty} U^o_{\epsilon_\nu}(x^\nu;l) \\ & \le & \mathbf{E}_\eta \limsup_{\nu\to+\infty} U^o_x(x^\nu,\epsilon_\nu\eta;l) \\ & \le & \mathbf{E}_\eta U^o(x,0;l,0) = U^o(x,0;l,0). \end{array}$$

From here by regularity assumption we obtain

$$U'_m(x;l) \le U^o(x,0;l,0) = U^o_x(x,0;l) = U^o(x;l)$$

and the desired inclusion. □

**Proposition 5.4** (Necessary optimality conditions). *Assume that*

*(i) functions $U_\epsilon(x)$ are Lipschitzian on $X$ with common Lipschitz constant for all $\epsilon > 0$;*

*(ii) functions $U_\epsilon(x)$ uniformly converge to $U(x)$ as $\epsilon \searrow 0$. Then at any local minimum $x^*$ of $U(x)$ on a compact set $X$*

$$0 \in \partial_m U(x^*) + N_X(x^*).$$

**Proof.** Define functions $\phi(z) = U(z) + \|z - x^*\|^2$ and $\phi_{\epsilon_\nu}(z) = U_{\epsilon_\nu}(z) + \|z - x^*\|^2$ for some sequence $\epsilon_\nu \searrow 0$. Let $B(x^*)$ be a ball around $x^*$ such that $U(z) \ge U(x^*)$ for all $z \in B(x^*) \cap X$. Obviously, $x^*$ is a unique global minimum of $\phi(z)$ on the set $B(x^*) \cap X$. Let functions $U_{\epsilon_\nu}(x)$ achieve their global minimums on $X$ at points $x^\nu$. By (ii) $x^\nu \longrightarrow x^*$ and by necessary optimality conditions [4] $0 = g^\nu + n^\nu$ for some $g^\nu \in \partial U_{\epsilon_\nu}(x^\nu)$, $n^\nu \in N_X(x^\nu)$. By (i) sequence $\{g^\nu\}$ has cluster points and let $g = \lim_k g^{\nu_k}$ be one of them. By construction $g \in \partial_m(x^*)$. Since $n^{\nu_k} = -g^{\nu_k}$ then by (i) sequence $\{n^{\nu_k}\}$ is bounded and thus has a cluster point $n$, which belongs to $N_X(x^*)$ by closedness of mapping $N_X(\cdot)$. Thus $0 = g + n \in \partial_m U(x^*) + N_X(x^*)$. □

**Corollary 5.2** (Sufficient condition for stationarity). *If under conditions of Proposition 5.3, 5.4* $0 \in \partial_m U(x^*)$ *then* $x^*$ *is a stationary point of function* $U(x) = \mathbf{E}_\omega U(x,0)$ *in the sense that* $0 \in \partial U(x^*)$ *and thus there is no such direction $l$ at $x^*$ that*

$$U(x + \lambda l) \le U(x) - \lambda\epsilon$$

*for all $x$ close to $x^*$, sufficiently small $\lambda$ and some $\epsilon > 0$.*

## 6 Stochastic optimization procedure

Let us consider the risk function in the form of extended expected utility function $U(x) = \mathbf{E}_\omega u(f(x,\omega))$, $u(\cdot)$ is some (possibly discontinuous) utility function. We are interested in solving the problem

$$U(x) \longrightarrow \min_{x \in X}. \tag{30}$$

For Lipschitzian function $U(x)$ and convex compact set $X$ we can define the attractor as the solution set satisfying necessary optimality conditions [4]

$$X^* = \{x^* \in X : \; 0 \in \partial U(x^*) + N_X(x^*)\},$$

where $\partial U(x)$ is Clarke's subdifferential of $U(x)$ and $N_X(x)$ is a normal cone to $X$ at point $x$. Unfortunately our problem $U(x)$ has, as a rule, a rather complex structure and no explicit form for subdifferntials $\partial U(x)$ is available. In sections $3-4$ we showed that $U(x)$ may be Lipschitz continuous and it can be approximated by (generalized [26], [25]) differentiable functions $U_\epsilon(x)$ uniformly in $x \in X$ in such a way that (see Corollary 5.1)

$$\partial U_{\epsilon_\nu}(x^\nu) \longrightarrow \partial U(x), \;\; \forall \;\; \epsilon_\nu \to 0, \;\; x^\nu \to x. \tag{31}$$

Let us assume that there exists such random vector function $\xi_\epsilon(x)$ that

$$\mathbf{E}\xi_\epsilon(x) \in \partial U_\epsilon(x), \qquad \sup_{\epsilon \in (0,\bar\epsilon],\, x \in X} \epsilon^2 \mathbf{E}\|\xi_\epsilon(x)\|^2 < +\infty \tag{32}$$

(see (23), (25) – (27) for particular examples). We are going to solve (30) through (possibly nonsmooth nonconvex) approximations $U_\epsilon(x)$, thus we are in the framework of the so-called limit extremal problems (see [8], [9] and references therein). Let $\{\epsilon_i\}$, $\{\rho_k\}$ be sequences of positive numbers such that

$$\lim_{i\to\infty} \epsilon_i = 0, \quad \lim_{k\to\infty} \rho_k = 0, \quad \sum_{k=0}^{\infty} \rho_k = +\infty. \tag{33}$$

Consider the following stochastic quasigradient (SQG) procedure:

*Step* 0: select $x^0 \in X$, set $i = 0$, $k = 0$, $k_i = 0$, $S = 0$;

*Step* 1: calculate

$$x^{k+1} = \Pi_X(x^k - \rho_k \xi_{\epsilon_i}(x^k)),$$

where $\Pi_X(\cdot)$ is the projection operator on the set $X$, and put $k := k+1$, $S := S + \rho_k$;

*Step* 2: if $S < \delta$ then go to Step 1, else put $i := i+1$, $k_i := k$, $S := 0$ and go to Step 1.

In this procedure we minimize function $U_{\epsilon_i}(x)$ by stochastic quasigradient method on iterations $k \in [k_i, k_{i+1})$, $\sum_{k=k_i}^{k_{i+1}-1} \rho_k \geq \delta > 0$, and then change $i$.

Define $\overline{\epsilon}_k = \epsilon_i$ for $k \in [k_i, k_{i+1})$, $i = 0, 1, \ldots$, and assume

$$\sum_{k=0}^{\infty} \left(\frac{\rho_k}{\overline{\epsilon}_k}\right)^2 < \infty. \tag{34}$$

**Theorem 6.1** (Convergence of the stochastic quasigradient procedure).
*Assume that Lipschitz continuous function $U(x)$ is uniformly approximated by generalized differentiable functions $U_{\epsilon_i}(x)$ as $\epsilon_i \to 0$ on a convex compact set $X \subset R^n$ in such a way that conditions (31), (32) hold. Let sequence $\{x^k\}$ be constructed by SQG-procedure, where sequences $\{\epsilon_i\}$, $\{\rho_k\}$ satisfy (33), (34). Then a.s.*

*(i) cluster points of $\{x^k\}$ constitute a compact connected set and minimal in $U$ cluster points of $\{x^k\}$ belong to the attractor $X^*$;*

*(ii) if $U(X^*)$ does not contain intervals then all cluster points of $\{x^k\}$ belong to $X^*$ and sequence $\{U(x^k)\}$ has a limit in $U(X^*)$.*

The proof of the theorem is similar to the proof of the analogues result in [16] (for $U_{\epsilon_i}(x) \equiv U(x)$) which is based on the technique developed in [29] and further elaborated in [8], [25].

## Concluding remarks

Any decision involving uncertainties leads to multiple outcomes with possible positive and negative consequences. Explicit introduction of risks as a function of decisions leads to a risk function which can be used to impose additional constraints on the feasible set of decisions. A more comprehensive (integrated) approach specifies a set of new risk reduction and risk spreading alternatives besides the set of the original decisions. The set of the risk-related decisions may include insurance, securities, different risk mitigation and adaptation strategies. For example, together with investments in conventional $CO_2$-producing technologies it may include investments in $CO_2$-consuming technologies. The explicit introduction of risk significantly affects the original profile of gains and losses, e.g., risks may become profitable for construction sectors of the economy and insurance industry. This can be summarized in a form of expected welfare function (see [10, [11]), in particular, a form of (extended) expected utility function as is discussed in section 2. As a result, the risk management becomes a part of the welfare maximization problem and the need for additional costs on the risk reduction measures is easily justified from the perspective of the overall welfare analysis. In other words, the integrated approach can show that the explicit introduction of uncertainties and risk reduction measures is a welfare-generating strategy, although the risk management per se requires additional costs. This is the main point of the approaches proposed in [10]-[12] for catastrophic risk management. In connection with this the important methodological issue is risk-based welfare analysis. Section 2 shows that in general we can not rely on the concavity of the adjusted-to-risk welfare function and, hence, on the concept of the standard general equilibrium. Important emerging issues seem to be negotiations, bargaining processes and an appropriate concept of dynamic stochastic equilibrium. All these questions are beyond the scope of this paper (see, e.g., [19] for a discussion of some closely related issues), but the problems analyzed here will remain to be crucial for more general models.

## References

[1] Asmussen S. (2000), *Ruin Probabilities*, World Scientific, Singapore.

[2] Beard R.E., Pentikäinen T. and Pesonen E. (1984), *Risk Theory. The Stochastic Basis of Insurance*, London, Chapman and Hall.

[3] Cacoullas T. (1966), Estimation of a multivariate density, *Ann. Inst. Statist. Math.*, 18, pp. 179-189.

[4] Clarke F.H. (1983), *Optimization and Nonsmooth Analysis*, Wiley.

[5] Consiglio A., Cocco F. and Zenios S. (2000), Asset and Liability Modeling for Participating Policies with Guaranties, *Working Paper 00-41-C*, The Wharton School, University of Pennsylvania.

[6] Devroye L.P. (1979), On pointwise and the integral convergence of recursive kernel estimation of probability densities, *Util. Math.*, 15, pp.113-128.

[7] Devroye L.P. (1987), *A course in density estimation*, Boston e.a., Birhäuser.

[8] Ermoliev Y.M. (1976), *Methods of stochastic programming*, Moscow, Nauka.

[9] Ermoliev Yu. (1988), Stochastic quasigradient methods, in: *Numerical Techniques for Stochastic Optimization*, Yu. Ermoliev and R.J-B Wets (Eds.), Springer-Verlag, pp.141-185, 413-433.

[10] Ermoliev Yu.M., Ermolieva T.Y., MacDonald G. and Norkin V.I. (1998), On the design of catastrophic risk portfolios, *IIASA Interim Report IR-98-056/August*, Laxenburg, Austria.

[11] Ermoliev Y.M., Ermolieva T.Y., MacDonald G.J. and Norkin V.I. (2000), Stochastic Optimization of Insurance Portfolios for Managing Exposure to Catastrophic Risks, *Annals of Operations Research*, Vol. 99, pp. 207-225.

[12] Ermoliev Y.M., Ermolieva T.Y., MacDonald G. and Norkin V.I. (2001), Problems of catastrophic risks insurance, *Kibernetika i sistemnyi analiz* (in Russian, English translation in *Cybernetics and Systems Analysis*), N 2, 99-110.

[13] Ermolieva T., Ermoliev Y., Linnerooth-Bayer J. and Galambos I. (2001), The Role of Financial Instruments in Integrated Catastrophic Flood Management, In the Proceedings of the 8-th Annual Conference of the Multinational Financial Society, Garda, Italy, July.

[14] Ermoliev Yu. and Gaivoronski A.A. (1992), Stochastic quasigradient methods for optimization of discrete event systems *Annals of Operations Research*, Vol. 39, pp. 1-39.

[15] Ermoliev Yu.M. and Norkin V.I. (1997), On nonsmooth and discontinuous problems of stochastic systems optimization, *European J. of Operational Research*, Vol. 101, pp.230-244.

[16] Ermoliev Yu.M. and Norkin V.I. (1998), Stochastic generalized gradient method for solving nonconvex nonsmooth stochastic optimization problems, *Kibernetika i sistemny analiz*, N 2, 50-71 (In Russian, English translation in *Cybernetics and systems analysis*, V. 34, N 2, and in *IIASA Interim Report IR-97-021*, Int. Inst. for Appl. Syst. Analysis, Laxenburg, Austria, 1997).

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

[18] Kall P., Ruszczyński A. and Frauendorfer (1988), Approximation techniques in stochastic programming, *Numerical Techniques for Stochastic Optimization*, Yu. Ermoliev and R.J-B Wets (Eds.), Springer-Verlag, pp.33-64.

[19] Keyzer M., Ermoliev Y. and Norkin V. (2001), General equilibrium and welfare modeling in spatial continuum: a practical framework for land use planning, *Interim Report IR-01-033*, Intern. Inst. for Appl. Syst. Analysis, Laxenburg, Austria, 28 p.

[20] Kibzun A.I. and Kan Yu.S. (1996), *Stochastic Programming Problems with Probability and Quantile Functions*, John Willey & Sons.

[21] Kibzun A. and Uryasev S. (1998), Differentiability of probability function, *Stochastic Analysis and Applications*, Vol. 16, No. 6, pp. 1101-1128.

[22] Lepp R. (1983), Stochastic approximation type algorithm for maximization of probability function, *Izvestia Akademii Nauk Estonskoi SSR. Fizika i matematika (Communications of Academy of Sciences of Estonia. Physics and Math.)*, Vol. 32, No. 2, pp. 150-156.

[23] Marti K. (1995), Differentiation of Probability functions: The Transformation Method, *Computers Math. Applic.*, Vol. 30, No. 3-6, pp. 361-382.

[24] Marti K. (1996), Differentiation Formulas for Probability functions: The Transformation Method, *Math. Progr. J.*, Series B, Vol. 75, No. 2.

[25] Mikhalevich V.S., Gupal A.M. and Norkin V.I. (1987), *Methods of Nonconvex Optimization*, Moscow, Nauka.

[26] Norkin V.I. (1986), Stochastic Generalized Differentiable Functions in the Problem of Nonconvex Nonsmooth Stochastic Optimization, *Cybernetics* (English translation from Russian of *Kibernetika*), Vol. 22, N 6, pp. 804-809.

[27] Norkin V.I. (1993), The analysis and optimization of probability functions, *Working Paper WP-93-6*, Int. Inst. for Appl. Syst. Analysis (IIASA), Laxenburg, Austria.

[28] Norkin V.I. and Roenko N.V. (1991), $\alpha$-Concave Functions and Measures and Their Applications, *Kibernet. Sistem. Anal.*, No. 6, pp. 77-88 (in Russian, English translation in: *Cybernet. Sustems Anal.*, Vol. 27, pp. 860-869).

[29] Nurminski E.A. (1979), *Numerical methods for solving deterministic and stochastic nonlinear minimax problems*, Kiev, Naukova Dumka.

[30] Prékopa A. (1995) *Stochastic programming*, Kluver Academic Publishers.

[31] Raik E. (1975), Differentiability on parameters of the probability function and stochastic pseudogradient method for its optimization, *Izvestia Akademii Nauk Estonskoi SSR, Fizika i Matematika (Communications of the Estonian Academy of Sciences, Physics and Mathematics)*, Vol. 24, N. 1, pp. 3-6.

[32] Rockafellar R.T. and Uryasev S. (2000), Optimization of Conditional Value at Risk, *The Journal of Risk*, Vol. 2, No. 3, pp.21-41.

[33] Rockafellar R.T. and Wets R.J-B. (1998), *Variational Analysis*, Berlin, Springer-Verlag.

[34] Rubinstein R. (1992), Sensitivity Analysis of Discrete Event Systems by the "Push Out" Method, *Annals of Operations Research*, 39.

[35] Tretiakov G. (2002), Quasi-Gradient Algorithms for Solving Stochastic Programming Problems with Probability Function. New Formula for Probability Function Gradient, *Lecture Notes in Economics and Mathematical Systems*, Vol. 513.

[36] Uryasev S. (1995), Derivatives of probability functions and some Applications, *Annals of Operations Research*, 56, pp.287-311.

[37] Wets R.J-B. (1996), Challenges in stochastic programming, *Math. Progr.*, Vol.75, pp.115-135.