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

Differentiation Formula for Integrals Over Sets Given by Inclusion

Uryasev, S.P.

IIASA Working Paper

WP-88-059

July 1988

Uryasev SP (1988). Differentiation Formula for Integrals Over Sets Given by Inclusion. IIASA Working Paper. IIASA, Laxenburg, Austria: WP-88-059 Copyright © 1988 by the author(s). http://pure.iiasa.ac.at/id/eprint/3146/

Working Papers 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. All rights reserved. Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage. All copies must bear this notice and the full citation on the first page. For other purposes, to republish, to post on servers or to redistribute to lists, permission must be sought by contacting repository@iiasa.ac.at

WORKING PAPER

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INCLUSION

Stanislav Uryas'ev

July 1988
WP-88-59

IIASA
International Institute for Applied Systems Analysis

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INCLUSION

Stanislav Uryas'ev

July 1988
WP-88-59

Working Papers are interim reports on work of the International Institute for Applied Systems Analysis and have received only limited review. Views or opinions expressed herein do not necessarily represent those of the Institute or of its National Member Organizations.

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

Formulae for differentiation with respect to the parameter of an integral over the set given by an inclusion are proposed. Such formulae are useful for solving chance constrained optimization problems. Using these formulae one can compute the gradient (or stochastic quasi-gradient) of the chance constraint and consequently apply gradient (or stochastic quasi-gradient) algorithm for the optimization.

Alexander B. Kurzhanski
Chairman
System and Decision Sciences Program

CONTENTS

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INCLUSION

Stanislav Uryas'ev

1. INTRODUCTION

Let the function

$$F(x) = \int_{f(x,\,y) \in B} p(x,\,y)\,\mathrm{d}y \tag{1}$$

be defined on the Euclidean space R^n, where $f: R^n \times R^m \to R$ and $p: R^n \times R^m \to R$ are some functions and $B \subseteq R$. To solve optimization problems containing the functions in the form (1), a differentiation formula for the function (1) is needed. One of the sources of such problems are chance constrained stochastic programming problems. For example, let

$$F(x) = P\{f(x,\,\varsigma(\omega)) \le 0\} \tag{2}$$

be a probability function, where $\varsigma(\omega)$ is a random vector in the space R^m. The random vector $\varsigma(\omega)$ has a probability density $p(x,\,y)$ depending on a parameter $x \in R$. The function (2) can be represented in the form (1), where $B = \{t : t \le 0\}$.

Differentiation formulae for function (1) are described in the papers of E. Raik [5], N. Roenko [6]. Special cases of probability functions (2) with normal and gamma distributions, have been investigated in the papers of A. Prékopa, T. Szantái [3], [4].

The gradient expression given in [5], [6] have the form of surface integrals, and are inconvenient from the computational point of view since the measure of a surface in the space R^m is equal to zero.

In the article of S. Uryas'ev [8] another type of formula was considered, where the gradient is an integral over a volume. For some applications such formulae are more convenient, because stochastic quasi-gradient algorithms [1] can be used for the minimization of function (1). In article [8] the formula for the gradient was proved under assumption that the set

$$\mu(x) = \{y \in R^m : f(x,\,y) \in B\} \tag{3}$$

is bounded.

The boundedness of the set $\mu(x_0)$ is rather strict assumption. For example, if we consider a linear function $f(x, y) = x_1y_1 + x_2y_2$, then for any nonempty set $B \subset R$ the set $\mu(x)$ is not bounded. In the present paper we prove analogous results with the weakest assumptions.

2. THE GRADIENT FORMULA FOR THE CASE WITH BOUNDED SET

The first result considers the case with bounded set $\mu(x) \cap T(x)$ where

$$T(x) = \{y \in R^m : p(x, y) \neq 0\} \ .$$

Let us denote V some bounded neighborhood of the point $x_0 \in R^n$, cl the closure sign and L^T a transposed matrix (or a vector) L; let $A(V) = (\bigcup_{x \in V} \mu(x)) \cap (\bigcup_{x \in V} T(x))$; $G = \mathrm{cl}(V \times A(V))$.

The expression for the gradient of the function (1) is given in the following theorem.

THEOREM 1 *Let:*

1 *the set G be bounded;*

2 *the function $f: R^n \times R^m \to R$ have continuous partial deviatives $f_x(x, y)$, $f_y(x, y)$, $f_{xy}(x, y)$, $f_{yy}(x, y)$ on an open neighborhood of the set G;*

3 *the function $p: R^n \times R^m \to R$ have continuous partial derivatives $p_x(x, y)$, $p_y(x, y)$ on an open neighborhood of the set G;*

4 *$\|f_y(x, y)\| \geq \gamma > 0$ on the set G.*

Then the function $F(x)$, given by the formula (1) is differentiable at the point x_0 and the gradient is equal to

$$F_x(x_0) = \int_{\mu(x_0)} [p_x(x_0, y) - \Lambda(x_0, y)\nabla_y]\,dy \ , \tag{4}$$

where

$$\Lambda(x, y) = p(x_0, y)\|f_y(x_0, y)\|^{-2} f_x(x_0, y) f_y^T(x_0, y) \tag{5}$$

PROOF Let $\sigma > 0$, define

$$T = \{z \in R^m : \|z - y\| \leq \sigma, \ y \in \mathrm{cl}(\bigcup_{x \in V} T(x))\} \ .$$

Taking into account the definition of T and $T(x)$ we have $p(x, y) = 0$ for $x \in V$, $y \in \mu(x) \cap (R^m \setminus T)$. For this reason

$$F(x) = \int_{\mu(x)} p(x, y)\, \mathrm{d}y = \int_{\mu(x) \cap T} p(x, y)\, \mathrm{d}y$$

if $x \in V$. Denote

$$\chi(\Delta x, z, \alpha) \overset{\text{def}}{=} f(x_0 + \Delta x, z + \alpha f_z(x_0, z)) - f(x_0, z) \ .$$

In the integral

$$F(x_0 + \Delta x) = \int_{\mu(x_0 + \Delta x) \cap T} p(x_0 + \Delta x, y)\, \mathrm{d}y \tag{6}$$

make the change of variables

$$y = z + \alpha(z, \Delta x) f_z(x_0, z)$$

where the function $\alpha(z, \Delta x)$ is given by the equation

$$\chi(\Delta x, z, \alpha) = 0 \ . \tag{7}$$

The theorem will depend on the following three lemmas.

LEMMA 1 *There exists a neighborhood U of the point x_0 and a unique function $\alpha(z, \Delta x)$ on the set $A(U) \times (U - x_0)$ which satisfies the equation (7) and the condition $\alpha(z, 0) = 0$. The function $\alpha(z, \Delta x)$ is continuously differentiable on $A(U) \times (U - x_0)$.*

PROOF Let us verify the conditions of the implicit function theorem (see, for example, [2], p. 454). By virtue of condition 2 of the theorem the function χ is continuous with respect to the all variables and the partial derivatives $\chi_{\Delta x}, \chi_z, \chi_\alpha$ exist and are continuous for sufficiently small $\Delta x, \alpha$, and for all $z \in A(V)$. The function χ at the point $(0, z, 0)$ equals zero. The derivative χ_α does not equal zero at this point (see the condition 4 of our theorem) since

$$\chi_\alpha(0, z, o) = \|f_z(x_0, z)\|^2 \geq \gamma > 0 \ .$$

All the conditions of the implicit function theorem are verified and the lemma is proved. □

Below we use the following definitions: $\varphi(z, \Delta x) = \alpha(z, \Delta x) f_z(x_0, z)$ where $\alpha(z, \Delta x)$ satisfies the equation (7); $a_{\Delta x}(z) = z + \varphi(z, \Delta x)$;

$$h(x_0 + \Delta x) = \{y \in R^m : f(x_0 + \Delta x, y) \in B, \, y \in T\} \ ;$$

$$h_{\Delta x}(x_0) = \{z \in R^m : f(x_0, z) \in B, a_{\Delta x}(z) \in T\} \quad .$$

LEMMA 2 *The mapping* $a_{\Delta x} : h_{\Delta x}(x_0) \to h(x_0 + \Delta x)$ *is an injection if* $\|\Delta x\|$ *is sufficiently small.*

PROOF First we show that if $z \in h_{\Delta x}(x_0)$ then $a_{\Delta x}(z) \in h(x_0 + \Delta x)$. Taking into account equation (7) and the definition of $h_{\Delta x}(x_0)$ we have

$$f(x_0, z) = f(x_0 + \Delta x, z + \varphi(z, \Delta x)) \in B$$

$$a_{\Delta x}(z) \in T$$

Consequently

$$f(x_0 + \Delta x, a_{\Delta x}(z)) \in B$$

$$a_{\Delta x}(z) \in T$$

and $a_{\Delta x}(z) \in h(x_0 + \Delta x)$.

With Lemma 1 and the condition 2 of our theorem it follows that the function $a_{\Delta x}(z)$ is differentiable on $A(U)$. Let us now prove that two different points z and $z + \Delta z$ from $h_{\Delta x}(x_0)$ can not be mapped by $a_{\Delta x}(z)$ into the time point of the set $h(x_0 + \Delta x)$. We divide this fact into two statements:

a) for arbitrary $\delta > 0$ there exists $\epsilon > 0$ such that if $\|\Delta z\| \geq \delta$ and $\|\Delta x\| \leq \epsilon$ then the inequality

$$\|a_{\Delta x}(z + \Delta z) - a_{\Delta x}(z)\| \geq \|\Delta z\|/2 \tag{8}$$

holds;

b) there exist $\delta > 0$ and $\epsilon > 0$ such that if $\|\Delta z\| \leq \delta$, $\|\Delta x\| \leq \epsilon$ then inequality (8) holds.

We start with the statement a). We have

$$\|a_{\Delta x}(z + \Delta z) - a_{\Delta x}(z)\| = \|\Delta z + \varphi(z + \Delta z, \Delta x) - \varphi(z, \Delta x)\| \geq$$

$$\geq \|\Delta z\| - \|\varphi(z + \Delta z, \Delta x) - \varphi(z, \Delta x)\| \tag{9}$$

Next, we evaluate the last term of the inequality. Expanding the function $\chi(\Delta x, z, \alpha)$ into a Teylor-series with respect to Δx, $\alpha(z, \Delta x)$ at the point $(0, z, 0)$ we have

$$\chi(\Delta x, z, \alpha(z, \Delta x)) = \chi(0, z, o) + <\chi_{\Delta x}(\theta \Delta x, z, \theta \alpha(z, \Delta x)), \Delta x> +$$

$$+ <\chi_\alpha(\theta\Delta x,\, z,\, \theta\alpha(z,\, \Delta x))\alpha(z,\, \Delta x) =$$

$$= <f_x(x_0 + \theta\Delta x,\, z + \theta\varphi(z,\, \Delta x)),\, \Delta x> +$$

$$+ <f_z(x_0 + \theta\Delta x,\, z + \theta\varphi(z,\, \Delta x)),\, f_z(x_0,\, z) > \alpha(z,\, \Delta x) = 0 \ ,$$

where $0 \leq \theta \leq 1$. With the notations $\tilde{x} = x_0 + \theta\Delta x$, $\tilde{z} = z + \theta\varphi(z,\, \Delta x)$ we get

$$\alpha(z,\, \Delta x) = -\frac{<f_x(\tilde{x},\, \tilde{z}),\, \Delta x>}{<f_z(\tilde{x},\, \tilde{z}),\, f_z(x_0,\, z)>} \ . \tag{10}$$

From the relation (10) and the condition 4 of the theorem we obtain

$$|\alpha(z,\, \Delta x)| = O(||\Delta x||) \tag{11}$$

for $z \in A(U)$ and sufficiently small Δx. (This means there exists a constant C such that $|\alpha(z,\, \Delta x)| \leq C||\Delta x||$.) Conditions 1,2 of our theorem imply that the constant C does not depend upon $z \in A(U)$. Relation (11) implies

$$||\varphi(z,\, \Delta x)|| = O(||\Delta x||) \ . \tag{12}$$

If $||\Delta z|| > \delta$ then there exists $\epsilon > 0$ such that for $||\Delta x|| < \epsilon$

$$||\varphi(z + \Delta z,\, \Delta x)|| < ||\Delta z||/4 \ ,$$

$$||\varphi(z,\, \Delta x)|| < ||\Delta z||/4$$

and consequently

$$||\varphi(z + \Delta z,\, \Delta x) - \varphi(z,\, \Delta x)|| \leq ||\Delta z||/2 \tag{13}$$

Substitution of the last inequality in (9) gives (8).

Let us now prove the statement b). Denote by

$$\tilde{\alpha} = -\frac{<f_x(x_0,\, z),\, \Delta x>}{||f_z(x_0,\, z)||^2} \, ,\ \tilde{\varphi} = \tilde{\alpha} f_z(x_0,\, z) \ .$$

From the condition 2 of the theorem and the relation (10) we get

$$\alpha(z,\, \Delta x) = \tilde{\alpha} + o(||\Delta x||) \ , \tag{14}$$

$$\varphi(z,\, \Delta x) = \tilde{\varphi} + o(||\Delta x||) \ , \tag{15}$$

where $o(||\Delta x||)/||\Delta x|| \to 0$ if $||\Delta x|| \to 0$. Next, we prove

$$\varphi_z(z,\, \Delta x) = \tilde{\varphi}_z + o(||\Delta x||) \ . \tag{16}$$

To justify (16) it is sufficient to show that

$$\alpha_{z_i}(z, \Delta x) = \tilde{\alpha}_{z_i} + o(\|\Delta x\|),\ i = 1, \ldots, m \quad . \tag{17}$$

We differentiate the identity $\chi(\Delta x, z, \alpha(z, \Delta x)) = 0$ with respect to z and get

$$\begin{aligned} 0 = \chi_z(\Delta x, z, \alpha(z, \Delta x)) &= f_z(x_0 + \Delta x, z + \varphi(z, \Delta x)) + \\ &+ <f_z(x_0 + \Delta x, z + \varphi(z, \Delta x)), f_z(x_0, z)>\alpha_z(z, \Delta x) + \\ &+ \alpha(z, \Delta x) f_{zz}(x_0, z) f_z(x + \Delta x, z + \varphi(z, \Delta x)) - f_z(x_0, z) \quad . \end{aligned} \tag{18}$$

Consequently, combining the Teylor-series expansion with the relations (14), (15), we obtain

$$\begin{aligned} 0 &= <f_{z_i x}(x_0, z), \Delta x> + <f_{z_i z}(x_0, z), \varphi(z, \Delta x)> + \|f_z(x_0, z)\|^2 \alpha_{z_i}(z, \Delta x) + \\ &+ \alpha(z, \Delta x)<f_{z_i z}(x_0, z), f_z(x_0, z)> + o(\|\Delta x\|) = \\ &= <f_{z_i x}(x_0, z), \Delta x> + 2<f_{z_i z}(x_0, z), \tilde{\varphi}> + \|f_z(x_0, z)\|^2 \alpha_{z_i}(z, \Delta x) + o(\|\Delta x\|) \end{aligned}$$

and

$$\alpha_{z_i}(z, \Delta x) = - \frac{<f_{z_i x}(x_0, z), \Delta x> + 2<f_{z_i z}(x_0, z), \tilde{\varphi}>}{\|f_z(x_0, z)\|^2} + o(\|\Delta x\|)$$

With

$$\tilde{\alpha}_{z_i} = - \frac{<f_{z_i x}(x_0, z), \Delta x> + 2<f_{z_i z}(x_0, z), \tilde{\varphi}>}{\|f_z(x_0, z)\|^2} \tag{19}$$

and the above we have the relation (17). Now

$$\tilde{\varphi}_z = (\tilde{\alpha} f_z(x_0, z))_z = \tilde{\alpha}_z f_z(x_0, z) + \tilde{\alpha} f_{zz}(x_0, z)$$

and

$$|\tilde{\alpha}| = O(\|\Delta x\|),\ \|\tilde{\alpha}_z\| = O(\|\Delta x\|) \quad ,$$

hence

$$\|\tilde{\varphi}_z\| = O(\|\Delta x\|) \quad . \tag{20}$$

With relations (16), (20) we have

$$\|\varphi_z(z, \Delta x)\| = O(\|\Delta x\|) \tag{21}$$

This implies

$$\|\varphi(z + \Delta z, \Delta x) - \varphi(z, \Delta x)\| = \|\langle\varphi_z(z + \nu\Delta z, \Delta x), \Delta z\rangle\| \leq$$

$$\leq \|\varphi_z(z + \nu\Delta z, \Delta x)\| \, \|\Delta z\| \leq O(\|\Delta x\|)\|\Delta z\| \quad ,$$

where $0 \leq \nu \leq 1$. Thus for sufficiently small $\|\Delta x\|$, the inequality (13) holds, and consequently inequality (8) holds. □

LEMMA 3 *The mapping* $a_{\Delta x}: h_{\Delta x} \rightarrow h(x_0 + \Delta x)$ *is a surjection for sufficiently small* $\|\Delta x\|$.

PROOF Let us take some point $y \in h(x_0 + \Delta x)$. We will prove that there exist a point $z^* \in Z \overset{\text{def}}{=} \{z : f(x_0, z) = f(x_0 + \Delta x, y)\}$ and a scalar $\alpha^* > 0$ such that $y = z^* + \alpha^* f(x_0, z^*)$. It is not difficult to see that z^* can be taken any point of the set

$$Z^* \overset{\text{def}}{=} \underset{z \in Z}{\operatorname{argmin}} \|z - y\| \quad .$$

Indeed, if $\tilde{z} \in Z^*$ then the vector $y - \tilde{z}$ should be collinear with the vector $f_z(x_0, \tilde{z})$, otherwise the set Z intersects with the interior of the set

$$\{z : \|z - y\| \leq \|\tilde{z} - y\|\}$$

and $\tilde{z} \notin Z^*$. Lemma 1 implies that the set Z^* consists only of the one point z^* and $y = a_{\Delta x}(z^*)$. Since $y \in h(x_0 + \Delta x)$ then $f(x_0, z^*) = f(x_0 + \Delta x, y) \in B$ and $y = a_{\Delta x}(z^*) \in T$, consequently $z^* \in h_{\Delta x}(x_0)$. □

Taking into account Lemmas 2 and 3 we can change variables $y = a_{\Delta x}(z)$ in the integral (6)

$$F(x_0 + \Delta x) = \int_{h(x_0 + \Delta x)} p(x_0 + \Delta x, y)\,dy = \int_{\substack{f(x_0 + \Delta x, y) \in B, \\ y \in T}} p(x_0 + \Delta x, y)\,dy =$$

$$= \int_{\substack{f(x_0 + \Delta x, a_{\Delta x}(z)) \in B, \\ a_{\Delta x}(z) \in T}} p(x_0 + \Delta x, a_{\Delta x}(z)) J\,dz =$$

$$= \int_{\substack{f(x_0, z) \in B, \\ a_{\Delta x}(z) \in T}} p(x_0 + \Delta x, a_{\Delta x}(z)) J\,dz =$$

$$= \int_{h_{\Delta x}(x_0)} p(x_0 + \Delta x, a_{\Delta x}(z)) J\,dz \quad . \tag{22}$$

where J is the Jacobian of the mapping $a_{\Delta x}(z)$. Since $a_{\Delta x}(z) = z + \varphi(z, \Delta x) \to z$ for $\|\Delta x\| \to 0$ and

$$p(x, y) = 0 \quad \text{for } x \in V,\, y \notin \bigcup_{x \in V} T(x)$$

then

$$F(x_0 + \Delta x) = \int\limits_{\substack{f(x_0, z) \in B, \\ a_{\Delta x}(z) \in T}} p(x_0 + \Delta x,\, a_{\Delta x}(z)) J \,\mathrm{d}z =$$

$$= \int\limits_{\substack{f(x_0, z) \in B, \\ z \in T}} p(x_0 + \Delta x,\, a_{\Delta x}(z)) J \,\mathrm{d}z \quad . \tag{23}$$

Let us now compute J

$$J = \begin{vmatrix} (y_1)_{z_1} & \cdots & (y_1)_{z_m} \\ (y_2)_{z_1} & \cdots & (y_2)_{z_m} \\ & \vdots & \\ (y_m)_{z_1} & \cdots & (y_m)_{z_m} \end{vmatrix}$$

We have

$$(y_i)_{z_i} = 1 + (\varphi_i(z, \Delta x))_{z_i},\ i = 1, \ldots, m \quad ,$$

$$(y_i)_{z_j} = (\varphi_i(z, \Delta x))_{z_j},\ i \neq j;\ i = 1, \ldots, m;\ j = 1, \ldots, m \quad .$$

With (21) we obtain

$$J = 1 + \sum_{i=1}^{m} (\varphi_i(z, \Delta x))_{z_i} + O(\|\Delta x\|^2) \tag{24}$$

Then with the above and (16)

$$J = 1 + \sum_{i=1}^{m} (\tilde{\varphi}_i)_{z_i} + o(\|\Delta x\|) \quad . \tag{25}$$

Substituting (25) into the relation (23) yields

$$F(x_0 + \Delta x) = \int\limits_{\mu(x_0) \cap T} p(x_0 + \Delta x, z + \varphi(z, \Delta x)) \left[1 + \sum_{i=1}^{m} (\tilde{\varphi}_i)_{z_i} + o(\|\Delta x\|) \right] J \,\mathrm{d}z =$$

$$= \int\limits_{\mu(x_0) \cap T} [p(x_0, z) + \langle p_x(x_0, z), \Delta x \rangle + \langle \rho_z(x_0, z), \tilde{\varphi} \rangle] \,\mathrm{d}z +$$

$$+ \int_{\mu(x_0)\cap T} p(x_0, z) \sum_{i=1}^{m} (\tilde{\varphi}_i)_{z_i} \mathrm{d}z + o(\|\Delta x\|) =$$

$$= \int_{\mu(x_0)\cap T} [p(x_0, z) + <p_x(x_0, z), \Delta x> + <p(x_0, z)\tilde{\varphi}, \nabla_z>] \mathrm{d}z + o(\|\Delta x\|) =$$

$$= \int_{\mu(x_0)\cap T} [p(x_0, z) + <p_x(x_0, z), \Delta x> - <\Lambda(x_0, z)\nabla_z, \Delta x>] \mathrm{d}z + o(\|\Delta x\|) \ .$$

Thus

$$F(x_0 + \Delta x) - F(x_0) = \int_{\mu(x_0)\cap T} <p_x(x_0, z) - \Lambda(x_0, z)\nabla_z, \Delta x> \mathrm{d}z + o(\|\Delta x\|) \ .$$

The last relation implies

$$F_x(x_0) = \int_{\mu(x_0)\cap T} [p_x(x_0, z) - \Lambda(x_0, z)\nabla_z] \mathrm{d}z \ .$$

Since $p_x(x_0, z) - \Lambda(x_0, z) = 0$ for $z \notin T$ then

$$F_x(x_0) = \int_{\mu(x_0)} [p_x(x_0, z) - \Lambda(x_0, z)\nabla_z] \mathrm{d}z$$

and this proves the theorem. □

3. THE GRADIENT FORMULA FOR THE CASE WITH UNBOUNDED SET

Next we prove that with some additional assumptions the formula (4) is true without the boundedness of the set $\mu(x_0) \cap T(x_0)$.

We use here the following designations: V is a bounded neighborhood of the point $x_0 \in R^n$; $A(V) = \bigcup_{x \in V} \mu(x)$; $G = \mathrm{cl}(V \times A(V))$; and $B^r \subset R^m$ is the ball with center at the point 0 and radius r. We introduce the function $p^r : R^n \times R^m \to R$. Outside the ball B^r the function p^r coincides with p i.e.

$$p(x, y) = p^r(x, y) \quad \text{for} \;\; y \notin B^r, x \in V \ , \tag{26}$$

and for $x \in V$ and y inside the ball B^r the function $p^r(x, y)$ can be set equal zero except in the neighborhood of the boundary of the B^r. Define

$$F^r(x) = \int_{\mu(x)} p^r(x, y) \, \mathrm{d}y \ .$$

We introduced the function F^r to estimate the integral

$$\int_{\mu(x)\setminus B^r} p(x,\, y)\,\mathrm{d}y \quad .$$

If $\left| \int_{\mu(x)\cap B^r} p^r(x,\, y)\,\mathrm{d}y \right|$ is a small value then

$$F^r(x) \approx \int_{\mu(x)\setminus B^r} p(x,\, y)\,\mathrm{d}y \quad .$$

Let us define

$$\Lambda(x,\, y) = p(x,\, y)\|f_y(x,\, y)\|^{-2} f_x(x,\, y) f_y^T(x,\, y) \quad ,$$

$$\Lambda^r(x,\, y) = p^r(x,\, y)\|f_y(x,\, y)\|^{-2} f_x(x,\, y) f_y^T(x,\, y) \quad .$$

We may now prove the following theorem.

THEOREM 2 *Let:*

1 *the function* $f\colon R^n \times R^m \to R$ *have that continuous partial derivatives* $f_x(x,\, y)$, $f_y(x,\, y)$, $f_{xy}(x,\, y)$, $f_{yy}(x,\, y)$ *on an open neighborhood of the set* G*;*

2 $\|f_y(x,\, y)\| > 0$ *on the set* G*;*

3 *the function* $p : R^n \times R^m \to R$ *have continuous partial derivatives* $p_x(x,\, y)$, $p_y(x,\, y)$ *on an open neighborhood of the set* G*;*

4 *for each* $r > 0$ *the function* $p^r : R^n \times R^m \to R$ *have continuous partial derivatives* $p_x^r(x,\, y)$, $p_y^r(x,\, y)$ *on an open neighborhood of the set*

$$\mathrm{cl}(V \times (A(V) \cap B^r))$$

and

$$p(x,\, y) = p^r(x,\, y) \quad \text{for } x \in V,\, y \notin B^r \;;$$

5 *for each* $\epsilon > 0$ *there exist* $R > 0$ *and* $\delta > 0$ *such that if* $\|\Delta x\| \le \delta$ *and* $r \ge R$ *then*

$$|F^r(x_0 + \Delta x) - F^r(x_0)| \le \epsilon\|\Delta x\| \;;$$

6 *for each* $r > 0$ *the integral*

$$Q(r) \overset{\text{def}}{=} \int_{\mu(x_0)} [p_x^r(x_0,\, y) - \Lambda^r(x_0,\, y)\nabla_y]\,\mathrm{d}y$$

exists and $|Q(r)| \to 0$ *for* $r \to +\infty$.

Then the function $F(x)$*, given by the formula (1) is differentiable at the point* x_0 *and the gradient is equal to*

$$F_x(x_0) = \int_{\mu(x_0)} [p_x(x, y) - \Lambda(x, y)\nabla_y]\, dy \ .$$

PROOF Let us take some $\epsilon > 0$. Applying the assumption 5 and 6 of the theorem we see that there exists $R > 0$ and $\delta_1 > 0$ such that if $||\Delta x|| \leq \delta_1$ and $r \geq R$

$$|F^r(x_0 + \Delta x) - F^r(x_0)| \leq \epsilon ||\Delta x||$$

and

$$|Q(r)| \leq \epsilon$$

Consequently

$$|F^r(x_0 + \Delta x) - F^r(x_0) - <Q(r), \Delta x>| \leq 2\epsilon ||\Delta x|| \qquad (27)$$

The function

$$F(x) - F^r(x) = \int_{\mu(x)} (p(x, y) - p^r(x, y))\, dy$$

satisfies the conditions of Theorem 1. For this reason there exists δ_2 such that if $||\Delta x|| \leq \delta_2$ then

$$|F(x_0 + \Delta x) - F^r(x_0 + \Delta x) - F(x_0) + F^r(x_0) -$$

$$- < \int_{\mu(x_0)} [p_x(x, y) - \Lambda(x, y)\nabla_y]\, dy - Q(r), \Delta x>| \leq \epsilon ||\Delta x|| \ . \qquad (28)$$

Let $\delta = \min(\delta_1, \delta_2)$. If $||\Delta x|| \leq \delta$ then applying (28) and (27) we get

$$|F(x_0 + \Delta x) - F(x_0) - < \int_{\mu(x_0)} [p_x(x, y) - \Lambda(x, y)\nabla_y]\, dy, \Delta x>| \leq$$

$$\leq |F(x_0 + \Delta x) - F^r(x_0 + \Delta x) - F(x_0) + F^r(x_0) -$$

$$- < \int_{\mu(x_0)} [p_x(x, y) - \Lambda(x, y)\nabla_y]\, dy - Q(r), \Delta x>| +$$

$$+ |F^r(x_0 + \Delta x) - F^r(x_0) - <Q(r), \Delta x>| \leq 3\epsilon ||\Delta x|| \ .$$

Because $\epsilon > 0$ was arbitrary the last inequality implies the statement of the theorem. □

4. LINEAR CASE

We consider the linear case in more detail. Let $x \in R^n$ and

$$F(x) = P\{c \leq A(\omega)x \leq B\} \ ,$$

where $A(\omega)$ is a random $k \times n$ matrix; the vectors c, b belongs to R^k, P is a probability measure; and the rows $a^j(\omega)$, $j = 1, \ldots, k$ of the random matrix $A(\omega)$ are independent and have the probability density functions $p^j(a^j)$, $j = 1, \ldots, k$. Denote by

$$\varphi^j(x) = P\{c_j \leq \langle a^j(\omega), x\rangle \leq b_j\} = \int_{c_j \leq \langle a^j, x\rangle \leq b_j} p^j(a^j)\, \mathrm{d}(a^j) \ .$$

In the view of the above assumptions

$$F(x) = \prod_{j=1}^{k} \varphi^j(x) \tag{29}$$

Let us compute $\varphi^j_x(x)$. With the formula (5) we get

$$\Lambda(x, a^j) = \begin{bmatrix} \lambda^1 \\ \vdots \\ \lambda^n \end{bmatrix}, \lambda^i = \|x\|^{-2} p^j(a^j) a^j_i x^T \ ,$$

$$\langle \lambda^i, \nabla_{a^j} \rangle = \|x\|^{-2} (\langle p^j_{a^j}(a^j), x\rangle a^j_i + p^j(a^j) x_i) \ .$$

Consequently

$$\varphi^j_x(x) = - \ \|x\|^{-2} \int_{c_j \leq \langle a^j, x\rangle \leq b_j} [\langle p^j_{a^j}(a^j), x\rangle a^j + p^j(a^j)x]\, \mathrm{d}(a^j) \ .$$

In the view of the preceding expression (29) it is easy to calculate $F_x(x)$.

5. ON THE MINIMIZATION OF THE INTEGRAL

Let us consider the problem of minimizing the function (1)

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

where $X \subset R^n$ is a convex closed set. To solve this problem one can use a gradient-based method. Note that for the computation of the gradient by formula (4) it is necessary to compute an n-dimentional integral. In order to avoid this, stochastic quasi-gradient algorithms can be used (see for example [1] [9]). One of the most simple stochastic quasi-gradient algorithms has the form

$$x^{s+1} = \Pi_X(x^s - \rho_s \xi^s)$$

where s is the number of the algorithm iteration; x^s is the approximation of the extremum on the s^{th} iteration; $\Pi_X(\cdot)$ is the orthoprojection operation on the convex set X; $\rho_s > 0$ is a step size; and ξ^s is a stochastic quasi-gradient i.e. the conditional expectation

$$M[\xi^s / x^0,\, \xi^0,\, x^1,\, \xi^1, \ldots, x^s] = F_x(x^s)$$

is equal to the gradient of the function $F(x)$ at the point x^s. In the case considered the stochastic quasi-gradient can be computed by the formula

$$\xi^s = (p_x(x^s, y^s) - \Lambda(x,^s, y^s)\nabla_y)p^{-1}(x^s, y^s)\Xi(y^s) \ .$$

where y^s is a sample of the probability vector y with density function $p(x^s, y)$ and

$$\Xi(y^s) = \begin{cases} 1, & \text{if} \quad f(x^s, y^s) \in B \ , \\ 0, & \text{if} \quad f(x^s, y^s) \notin B \ . \end{cases}$$

Ruszczynski, A. and Syski, W. [7] have used an analogous method for the minimization of the function

$$F(x) = P\{x - \eta(\omega) \in B\},\ x \in R^2 \ ,$$

where P is a probability measure, $\eta(\omega) \in R^2$ is a normally distributed random vector, and $B \subset R^2$ is some closed bounded set.

REFERENCES

[1] Ermoliev, Yu.: Stochastic quasi-gradient methods and their applications to system optimization. Stochastics, 4, (1983).

[2] Phikhtengoltz, G.: Course of differential and integral calculation II. Moskow, FM, (1962) (in Russian).

[3] Prékopa, A.: On probabilistic constrained programming, in: Proceedings of the Princeton Symposium on Mathematical Programming (Princton University Press, Princeton, N.J., 1970) 113–138.

[4] Prékopa, A. and T. Szántai: A new multivariate gamma distribution and its fitting to empirical streamflow data, Water Resources Research, 14 (1978) 19–24.

[5] Raik, E.: The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method, Izvestiya Akad. Nayk Est. SSR, Phis. Math., 24, 1 (1975) 3–6 (in Russian).

[6] Roenko, N.: Stochastic programming problems with integral functionals over multivalues mappings. Synopsis of thesis, USSR, Kiev, (1983) (in Russian).

[7] Ruszczynski, A. and W. Syski: A method of aggregate stochastic subgradient with on-line stepsize rule for convex stochastic programming problems. Math. Programming Study, 28, II, (1985), 113–131.

[8] Uryas'ev, S.: On the differentiability of an integral over the set given by inclusion. Kibernetika (Kiev) (1988), to appear (in Russian).

[9] Uryas'ev, S.: Stochastic quasi-gradient algorithms with adaptively controlled parameters. Austria, Laxenburg, IIASA, WP-86-32 (1986).