Working Paper

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INEQUALITIES

S. Uryas'ev

WP-91-38
October 1991

IIASA
International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INEQUALITIES

S. Uryas'ev

WP-91-38
October 1991

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Foreword

The investigation of a problem connected to probabilistic risk assessments for industrial plants led us to the need to optimize integrals calculated over sets that depend upon parameters. The problem was developed for two applications of tested and inspected components as an optimal control problem involving nonsmooth state transitions. In solving the optimization problem it is necessary to calculate the derivatives of an integral over a domain depending upon the parameters to be optimized. Up to date the theory of the differentiation these integrals is not fully developed. In the working paper a new general formula for differentiation of such integrals is proposed. These results were used for calculation of sensitivities for risk functions. This approach can have a wide application for the stochastic programming problems.

Comments about these mathematical results are invited.

Björn Wahlström
Leader
Social & Environmental Dimensions
of Technology Project

Contents

DIFFERENTIATION FORMULA FOR INTEGRALS OVER SETS GIVEN BY INEQUALITIES

S. Uryas'ev

1 Introduction

The optimization of operational strategies for inspected components[5],[6] can be reduced to optimization of the sum of integrals taken over sets that depend upon the parameters. To date, the theory for differentiation of such integrals is not fully developed. Here, we prove a general formula for the differentiation of an integral over the volume given by many inequalities. A gradient of the integral is represented as a sum of integrals taken over volume and over surface.

Let the function

$$F(x) = \int_{f(x,y) \leq 0} p(x,y)\, \mathrm{d}y \tag{1}$$

be defined on the Euclidean space R^n, where $f : R^n \times R^m \to R^k$ and $p : R^n \times R^m \to R$ are some functions. The inequality $f(x,y) \leq 0$ should be treated as a system of inequalities

$$f_i(x,y) \leq 0\ ,\quad i = 1, \ldots, k\ .$$

To solve optimization problems containing functions of the form (1), a differentiation formula for function (1) is needed. Stochastic programming problems lead to such functions. For example, let

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

be a probability function, where $\zeta(\omega)$ is a random vector in R^m. The random vector $\zeta(\omega)$ has a probability density $p(x,y)$ that depends on a parameter $x \in R$.

Differentiation formulae for function (1) in the case of only one inequality ($k = 1$) are described in the papers of E. Raik [7] and N. Roenko [8]. More general results ($k \geq 1$) were obtained in the papers of J. Simon (see, for example [9]). Special cases of probability function (2) with normal and gamma distributions have been investigated in the papers of A. Prékopa [3], and A. Prékopa and T. Szantái [4]. In the forthcoming book of G.Ch. Pflug [2] the gradient of the function (1) is represented in a form of a conditional expectation ($k = 1$).

The gradient expressions given in [7], [8] and [9] have the form of surface integrals, and are inconvenient from the computational point of view, since the measure of a surface in R^m is equal to zero.

In the papers [10], [11] of S. Uryas'ev, another type of formula was considered, where the gradient is an integral over a volume. For some applications this type of formula is more convenient. For example, stochastic quasi-gradient algorithms [1] can be used for the minimization of function (1). Here, we propose the formula for the general case of $k \geq 1$, and the formulae in the papers [7] and [10] are special cases of this general result. Since the gradient of the function (1) is represented in [10] and [11] as an integral over volume, in the case of $k = 1$, it is clear that this integral can be reduced to an integral over a surface (see [7]). Furthermore, it appears that the gradient of the function (1) can also be represented as a sum of integrals taken over volume and over surface (in the case of $k \geq 1$). This kind of formula is especially convenient for the case when the inequalities $f(x, y) \leq 0$ include simple constraints $y_i \geq 0$, $i = 1, \dots, m$ (see an example at the end of the paper).

2 The General Formula

Let us introduce the following shorthand notations

$$f(x,y) = \begin{pmatrix} f_1(x,y) \\ \vdots \\ f_k(x,y) \end{pmatrix}, \quad f_{1,l}(x,y) = \begin{pmatrix} f_1(x,y) \\ \vdots \\ f_l(x,y) \end{pmatrix},$$

$$\nabla_y f(x,y) = \begin{pmatrix} \frac{\partial f_1(x,y)}{\partial y_1}, & \dots, & \frac{\partial f_k(x,y)}{\partial y_1} \\ & \vdots & \\ \frac{\partial f_1(x,y)}{\partial y_m}, & \dots, & \frac{\partial f_k(x,y)}{\partial y_m} \end{pmatrix},$$

and

$$\nabla_y^N f_i(x,y) = \nabla_y f_i(x,y)/||\nabla_y f_i(x,y)|| .$$

A transposed matrix H is denoted by H^T, and the Jacobian of the function $f(x, y)$ is denoted by

$$\nabla_y^T f(x,y) = \begin{pmatrix} \frac{\partial f_1(x,y)}{\partial y_1}, & \dots, & \frac{\partial f_1(x,y)}{\partial y_m} \\ & \vdots & \\ \frac{\partial f_k(x,y)}{\partial y_1}, & \dots, & \frac{\partial f_k(x,y)}{\partial y_m} \end{pmatrix}.$$

Further, we need a definition of divergence

$$\operatorname{div}_y H = \begin{pmatrix} \sum_{i=1}^{m} \frac{\partial h_{1,i}}{\partial y_i} \\ \vdots \\ \sum_{i=1}^{m} \frac{\partial h_{n,i}}{\partial y_i} \end{pmatrix} , \quad \text{and} \quad \operatorname{div}_y^T H = \left(\sum_{i=1}^{m} \frac{\partial h_{1,i}}{\partial y_i} , \ldots , \sum_{i=1}^{m} \frac{\partial h_{n,i}}{\partial y_i} \right) ,$$

for the matrix

$$H = \begin{pmatrix} h_{1,1} , & \ldots & , h_{1,m} \\ & \vdots & \\ h_{n,1} , & \ldots & , h_{n,m} \end{pmatrix} .$$

We define

$$\mu(x) \; = \; \{ \, y \in R^m \; : \; f(x,y) \leq 0 \, \} \; \stackrel{\text{def}}{=} \; \{ \, y \in R^m \; : \; f_l(x,y) \leq 0 \, , \; 1 \leq l \leq k \, \} \, ,$$

$$\mu_\epsilon(x) \; = \; \{ \, y \in R^m \; : \; f(x,y) \leq \epsilon \} \; \stackrel{\text{def}}{=} \; \{ \, y \in R^m \; : \; f_l(x,y) \leq \epsilon_l \, , \; 1 \leq l \leq k \, \} \, ,$$

$$\epsilon = (\epsilon_1, \ldots, \epsilon_k) \, ,$$

and $\partial\mu(x)$ to be the surface of the set $\mu(x)$. We denote by $\partial_i\mu(x)$ a part of the surface which corresponds to the function $f_i(x,y)$

$$\partial_i\mu(x) \; = \; \mu(x) \bigcap \{ \, y \in R^m \; : \; f_i(x,y) = 0 \, \} \, ,$$

and

$$\partial_{l,k}\mu(x) \; = \; \bigcup_{l \leq i \leq k} \partial_i\mu(x) \, .$$

For $y \in \partial\mu(x)$, we define

$$I(x,y) = \{ \, i \; : \; f_i(x,y) = 0 \, \} \, ,$$

$$i(x,y) = \min_{j \in I(x,y)} j \, .$$

Let us denote by $x \times \mu_\epsilon(x)$, the Cartesian product of the point x and the set $\mu_\epsilon(x)$. Let $V(x_0)$ be some open neighborhood of the point x_0 and

$$G_{V,\epsilon}(x) \; = \; \bigcup_{x \in V(x)} \big(x \times \mu_\epsilon(x) \big) \, .$$

If we split the set $K \stackrel{\text{def}}{=} \{1, \ldots, k\}$ into two subsets K_1 and K_2, without loss of generality we can consider

$$K_1 \; = \; \{1, \ldots, l\} \;\; \text{and} \;\; K_2 \; = \; \{l+1, \ldots, k\} \, .$$

Now we formulate a theorem about differentiation of integral (1).

Theorem 2.1 *Let us assume that the following conditions are satisfied:*

1. *the set* $G_{V,\epsilon}(x_0)$ *is bounded for some* $V(x_0)$ *and* $\epsilon = (\epsilon_1, \ldots, \epsilon_k)$; $\epsilon_i > 0$, $i = 1, \ldots, k$;

2. *the function* $f : R^n \times R^m \to R^k$ *has continuous partial derivatives* $\nabla_x f(x,y)$, $\nabla_y f(x,y)$, $\nabla_{xy} f(x,y)$, $\nabla_{yy} f(x,y)$ *on the set* $G_{V,\epsilon}(x_0)$;

3. *the function* $p : R^n \times R^m \to R$ *has continuous partial derivatives* $\nabla_x p(x,y)$, $\nabla_y p(x,y)$ *on the set* $G_{V,\epsilon}(x_0)$;

4. *there exists a matrix function* $H_{1,l} : R^n \times R^m \to R^n \times R^m$ *satisfying the equation:*

$$H_{1,l}(x,y)\, \nabla_y f_{1,l}(x,y) + \nabla_x f_{1,l}(x,y) = 0 \tag{3}$$

on the set $G_{V,\epsilon}(x_0)$;

5. *the matrix function* $H_{1,l}(x,y)$ *has a continuous partial derivative* $\nabla_y H_{1,l}(x,y)$ *on the set* $G_{V,\epsilon}(x_0)$;

6. *there exist continuous matrix functions* $H_i : R^n \times R^m \to R^n \times R^m$, $i = l+1, \ldots, k$ *satisfying the equations:*

$$H_i(x,y)\, \nabla_y f_i(x,y) + \nabla_x f_i(x,y) = 0 \ , \quad i = l+1, \ldots, k \ , \tag{4}$$

on the set $G_{V,\epsilon}(x_0)$;

7. *if the set* $\partial_i \mu(x_0)$ *is not empty then the gradient* $\nabla_y f_i(x_0, y)$ *is not equal to zero on* $\partial_i \mu(x_0)$ *i.e.*

$$\min_{y \in \partial\mu(x_0)} \ \min_{i \in I(x_0,y)} \|\nabla_y f_i(x_0,y)\| \geq \gamma > 0 \ ; \tag{5}$$

8. *for all* $y \in \partial\mu(x_0)$ *the vectors* $\nabla_y f_i(x_0, y)$, $i \in I(x_0, y)$ *are linearly independent.*

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

$$\nabla_x F(x_0) = \int_{\mu(x_0)} \left[\nabla_x p(x_0,y) + \mathrm{div}_y\big(p(x_0,y) H_{1,l}(x_0,y)\big) \right] \mathrm{d}y \ +$$

$$+ \int_{\partial_{l+1,k}\mu(x_0)} p(x_0,y) \left[H_{i(x_0,y)}(x_0,y) - H_{1,l}(x_0,y) \right] \nabla_y^N f_{i(x_0,y)}(x_0,y) \ \mathrm{d}S \ . \tag{6}$$

Proof. Prior to proving the theorem, we prove the following lemma.

Lemma 2.1 *If a function $r : R \to R^k$ satisfies the condition*

$$r(||\Delta x||)/||\Delta x|| \to 0 , \quad ||\Delta x|| \to 0 ,$$

then for sufficiently small $||\Delta x||$ the inequality

$$\left| \int\limits_{\mu_{r(||\Delta x||)}(x_0+\Delta x)} p(x_0 + \Delta x, y)\, \mathrm{d}y \ - \int\limits_{\mu(x_0+\Delta x)} p(x_0 + \Delta x, y)\, \mathrm{d}y \right| \le T \sum_{i=1}^{k} |r_i(||\Delta x||)| ,$$

holds, where T is a constant.

Proof. The sets

$$(x_0 + \Delta x) \times \mu(x_0 + \Delta x) \quad \text{and} \quad (x_0 + \Delta x) \times \mu_{r(||\Delta x||)}(x_0 + \Delta x)$$

belong to the set $G_{V,\epsilon}(x_0)$, if $||\Delta x||$ is small enough. Since the set $G_{V,\epsilon}(x_0)$ is bounded, then without loss of generality we can consider that the continuous function $p(x, y)$ is bounded on $G_{V,\epsilon}(x_0)$. Consequently, the value

$$p(x_0 + \Delta x, y), \ \text{for} \ y \in \mu(x_0 + \Delta x), \ y \in \mu_{r(||\Delta x||)}(x_0 + \Delta x)$$

is bounded for sufficiently small $||\Delta x||$.

Let $A_1, A_2 \subset R^m$ and denote by $A_1 \Delta A_2$ a symmetric difference of the sets A_1 and A_2

$$A_1 \Delta A_2 \stackrel{\text{def}}{=} (A_1 \setminus A_2) \bigcup (A_2 \setminus A_1) .$$

For a set $A \in R^m$ denote by $q(A)$ the Lebesgue measure in R^m . Since the value

$$p(x_0 + \Delta x, y), \ \text{for} \ y \in \left(\mu(x_0 + \Delta x) \,\Delta\, \mu_{r(||\Delta x||)}(x_0 + \Delta x)\right)$$

is bounded for sufficiently small $||\Delta x||$, the statement of the lemma follows from the inequality

$$q\left(\mu(x_0 + \Delta x) \,\Delta\, \mu_{r(||\Delta x||)}(x_0 + \Delta x)\right) \le T_1 \sum_{i=1}^{k} |r_i(||\Delta x||)| , \tag{7}$$

$$T_1 = \text{const} .$$

Let us prove inequality (7). Denote

$$|r(||\Delta x||)| = \left(|r_1(||\Delta x||)|, \ldots, |r_k(||\Delta x||)|\right) .$$

Since

$$\mu_{-|r(||\Delta x||)|}(x_0 + \Delta x) \subset \mu(x_0 + \Delta x) , \quad \mu_{-|r(||\Delta x||)|}(x_0 + \Delta x) \subset \mu_{r(||\Delta x||)}(x_0 + \Delta x) ,$$

and

$$\mu(x_0 + \Delta x) \subset \mu_{|r(||\Delta x||)|}(x_0 + \Delta x) , \quad \mu_{r(||\Delta x||)}(x_0 + \Delta x) \subset \mu_{|r(||\Delta x||)|}(x_0 + \Delta x) ,$$

then

$$\left(\mu(x_0+\Delta x)\,\Delta\,\mu_{r(||\Delta x||)}(x_0+\Delta x)\right) \;\subset\; \left(\mu_{|r(||\Delta x||)|}(x_0+\Delta x) \setminus \mu_{-|r(||\Delta x||)|}(x_0+\Delta x)\right) . \quad (8)$$

We define a layer $D^i_\delta(x)$ as follows

$$D^i_\delta(x) = \{\, y \in R^m : -\delta \le f_i(x,y) \le \delta \,\} .$$

Further, we can write

$$\left(\mu_{|r(||\Delta x||)|}(x_0+\Delta x) \;\setminus\; \mu_{-|r(||\Delta x||)|}(x_0+\Delta x)\right) \;\subset \quad (9)$$

$$\bigcup_{i=1}^{k} \left(\mu_{|r(||\Delta x||)|}(x_0+\Delta x) \bigcap D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x) \right) .$$

Conditions 1, 2 and 7 of the theorem imply that the thickness of the layer $D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x)$ is less than $6\gamma^{-1}|r_i(||\Delta x||)|$. Indeed, it follows from conditions 1, 2 and 7 of the theorem that

$$||\nabla f_i(x_0+\Delta x, y)|| \ge 2^{-1}\gamma \quad \text{for} \quad y \in \partial\mu_i(x_0+\Delta x) ,$$

and sufficiently small $||\Delta x||$. Consequently, with Taylor's theorem

$$f_i\big(x_0+\Delta x, y + 3\gamma^{-1}|r_i(||\Delta x||)|\, \nabla^N_y f_i(x_0+\Delta x, y)\big) \;=$$

$$f_i(x_0+\Delta x, y) \;+\; 3\gamma^{-1}|r_i(||\Delta x||)|\, \langle \nabla^N_y f_i(x_0+\Delta x, y) \,,\, \nabla_y f_i(x_0+\Delta x, y)\rangle \;+$$

$$o\big(r_i(||\Delta x||)\big) \;\ge\; f_i(x_0+\Delta x, y) \;+\; 3\gamma^{-1}|r_i(||\Delta x||)|\, 2^{-1}\gamma \;+\; o\big(r_i(||\Delta x||)\big) \;>$$

$$f_i(x_0+\Delta x, y) \;+\; |r_i(||\Delta x||)| ,$$

for $y \in \partial\mu_i(x_0+\Delta x)$ and sufficiently small $||\Delta x||$. It means that point $y + 3\gamma^{-1}|r_i(||\Delta x||)|$ is out of the layer

$$D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x) \;\; \text{for} \;\; y \in \partial\mu_i(x_0+\Delta x) .$$

Analogously, the point $y - 3\gamma^{-1}|r_i(||\Delta x||)|$ is out of the layer

$$D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x) \;\; \text{for} \;\; y \in \partial\mu_i(x_0+\Delta x) .$$

Thus, thickness of the layer $D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x)$ is less than $6\gamma^{-1}|r_i(||\Delta x||)|$. Consequently, there exist a constant T_1 such that

$$q\left(\mu_{|r(||\Delta x||)|}(x_0+\Delta x) \bigcap D^i_{|r_i(||\Delta x||)|}(x_0+\Delta x) \right) \;\le\; T_1 |r_i(||\Delta x||)| . \quad (10)$$

Inclusions (8), (9) and inequality (10) imply

$$q\left(\mu(x_0+\Delta x)\,\Delta\,\mu_{r(||\Delta x||)}(x_0+\Delta x) \right) \;\le$$

$$q\left(\mu_{|r(\|\Delta x\|)|}(x_0+\Delta x) \setminus \mu_{-|r(\|\Delta x\|)|}(x_0+\Delta x)\right) \leq$$

$$q\left(\bigcup_{i=1}^{k}\left(\mu_{|r(\|\Delta x\|)|}(x_0+\Delta x) \bigcap D^i_{|r_i(\|\Delta x\|)|}(x_0+\Delta x)\right)\right) \leq$$

$$\sum_{i=1}^{k} q\left(\mu_{|r(\|\Delta x\|)|}(x_0+\Delta x) \bigcap D^i_{|r_i(\|\Delta x\|)|}(x_0+\Delta x)\right) \leq T_1 \sum_{i=1}^{k} |r_i(\|\Delta x\|)| .$$

The lemma is proved. □

Let us make an increment Δx in the argument of the function $F(x_0)$

$$F(x_0+\Delta x) = \int_{f(x_0+\Delta x,y)\leq 0} p(x_0+\Delta x, y)\,\mathrm{d}y ,$$

and make a change of variables

$$y = z + H^T_{1,l}(x_0,z)\Delta x . \tag{11}$$

Taylor's theorem implies that

$$f(x_0+\Delta x, y) = f(x_0+\Delta x, z + H^T_{1,l}(x_0,z)\Delta x) =$$

$$f(x_0,z) + \nabla^T_x f(x_0,z)\Delta x + \nabla^T_z f(x_0,z) H^T_{1,l}(x_0,z)\Delta x + o(\|\Delta x\|) , \tag{12}$$

where $o(\|\Delta x\|)/\|\Delta x\| \to 0$ as $\|\Delta x\| \to 0$. Since the set $G_{V,\epsilon}(x_0)$ is bounded, we can assume without loss of generality that functions

$$f(x,z) ,\ \nabla_x f(x,z) ,\ \nabla_z f(x,z) ,\ H_{1,l}(x,z)$$

are uniformly continuous on $G_{V,\epsilon}(x_0)$ (see conditions 1, 2 and 5 of Theorem 2.1). Consequently,

$$o(\|\Delta x\|)/\|\Delta x\| \to 0$$

uniformly with respect to $(x_0+\Delta x)\times z \in G_{V,\epsilon}(x_0)$.

Denote by $J(x_0,z)$ the determinant of the Jacobian for the mapping $y(z)$. With Lemma 2.1 and equations (11), (12) we have

$$F(x_0+\Delta x) = \int_{f(x_0+\Delta x,y)\leq 0} p(x_0+\Delta x, y)\,\mathrm{d}y =$$

$$\int_{f(x_0+\Delta x,y) - o(\|\Delta x\|)\leq 0} p(x_0+\Delta x, y)\,\mathrm{d}y + o(\|\Delta x\|) =$$

$$\int_{\substack{f(x_0,z)+\nabla^T_x f(x_0,z)\Delta x+ \\ +\nabla^T_z f(x_0,z)H^T_{1,l}(x_0,z)\Delta x\leq 0}} p\big(x_0+\Delta x, z + H^T_{1,l}(x_0,z)\Delta x\big)\, J(x_0,z)\,\mathrm{d}z + o(\|\Delta x\|) . \tag{13}$$

Let us compute the determinant $J(x_0, z)$

$$J(x_0,z) \;=\; \left|\nabla_z^T y(z)\right| \;=\; |E \,+\, \nabla_z^T H_{1,l}^T(x_0,z)\Delta x| \;=$$

$$1 \,+\, \mathrm{div}_z\big(\Delta x^T H_{1,l}(x_0,z)\big) \,+\, o(\,\|\Delta x\|\,)\,. \tag{14}$$

Furthermore,

$$p\big(x_0+\Delta x, z + H_{1,l}^T(x_0,z)\Delta x\big)\, J(x_0,z) \;=$$

$$\big[p(x_0,z) + \nabla_x^T p(x_0,z)\Delta x + \nabla_z^T p(x_0,z) H_{1,l}^T(x_0,z)\Delta x + o(\,\|\Delta x\|\,)\big] \times$$

$$\times\big[1 \,+\, \mathrm{div}_z\big(\Delta x^T H_{1,l}(x_0,z)\big) \,+\, o(\,\|\Delta x\|\,)\big] =$$

$$p(x_0,z) + \big[\,\nabla_x^T p(x_0,z) + \nabla_z^T p(x_0,z) H_{1,l}^T(x_0,z) + p(x_0,z)\mathrm{div}_z^T H_{1,l}(x_0,z)\big]\,\Delta x \,+$$

$$+\, o(\,\|\Delta x\|\,) \;=$$

$$p(x_0,z) + \big[\,\nabla_x^T p(x_0,z) + \mathrm{div}_z^T\big(p(x_0,z) H_{1,l}(x_0,z)\big)\big]\,\Delta x \,+ o(\,\|\Delta x\|\,)\,, \tag{15}$$

where $o(\,\|\Delta x\|\,)\,/\,\|\Delta x\| \to 0$ as $\|\Delta x\| \to 0$. Since the set $G_{V,\epsilon}(x_0)$ is bounded, we can assume without loss of generality that functions

$$p(x,z)\,,\ \nabla_x p(x,z)\,,\ \nabla_z p(x,z)\,,\ H_{1,l}(x,z)\,,\ \nabla_z H_{1,l}(x,z)$$

are uniformly continuous and bounded on $G_{V,\epsilon}(x_0)$ (see conditions 1, 3 and 5 of Theorem 2.1). Consequently, for the term $o(\,\|\Delta x\|\,)$ in formula (15) the convergence

$$o(\,\|\Delta x\|\,)\,/\,\|\Delta x\| \to 0 \tag{16}$$

is uniform with respect to $(x_0 + \Delta x) \times z \,\in\, G_{V,\epsilon}(x_0)$.

With (13) and (15) we have

$$F(x_0+\Delta x) \;=\; o(\,\|\Delta x\|\,) \;+ \int\limits_{\substack{f(x_0,z)+\nabla_x^T f(x_0,z)\Delta x+\\ +\nabla_z^T f(x_0,z)H_{1,l}^T(x_0,z)\Delta x \le 0}} o(\,\|\Delta x\|\,)\,\mathrm{d}z \;+$$

$$\int\limits_{\substack{f(x_0,z)+\nabla_x^T f(x_0,z)\Delta x+\\ +\nabla_z^T f(x_0,z)H_{1,l}^T(x_0,z)\Delta x \le 0}} \Big\{p(x_0,z) + \big[\,\nabla_x^T p(x_0,z) + \mathrm{div}_z^T\big(p(x_0,z)H_{1,l}(x_0,z)\big)\big]\,\Delta x\Big\}\,\mathrm{d}z\,. \tag{17}$$

Condition 1 of Theorem 2.1 and uniform convergence (16) imply

$$\int\limits_{\substack{f(x_0,z)+\nabla_x^T f(x_0,z)\Delta x+\\ +\nabla_z^T f(x_0,z)H_{1,l}^T(x_0,z)\Delta x \le 0}} o(\,\|\Delta x\|\,)\,\mathrm{d}z \;\;=\;\; o(\,\|\Delta x\|\,)\,. \tag{18}$$

Combining (17) and (18) we have

$$F(x_0+\Delta x) \;=\; o(\,\|\Delta x\|\,)+ \int\limits_{\substack{f(x_0,z)+\nabla_x^T f(x_0,z)\Delta x+\\ +\nabla_z^T f(x_0,z)H_{1,l}^T(x_0,z)\Delta x\le 0}} \Big\{p(x_0,z)+\big[\nabla_x^T p(x_0,z)+\mathrm{div}_z^T\big(p(x_0,z)H_{1,l}(x_0,z)\big)\big]\,\Delta x\Big\}\;\mathrm{d}z\;, \tag{19}$$

where $o(\,\|\Delta x\|\,)\,/\,\|\Delta x\|\to 0$ as $\|\Delta x\|\to 0$. Equations (3) and (4) imply

$$\{\,z\in R^m : f(x_0,z)+\nabla_x^T f(x_0,z)\Delta x+\nabla_z^T f(x_0,z)H_{1,l}^T(x_0,z)\Delta x\;\le\;0\}\;=$$

$$\{\,z\in R^m : f_{1,l}(x_0,z)\;\le\;0\;;$$

$$f_i(x_0,z)+\nabla_x^T f_i(x_0,z)\Delta x+\nabla_z^T f_i(x_0,z)H_{1,l}^T(x_0,z)\Delta x\;\le\;0\,,\,i=l+1,\ldots,k\}\;=$$

$$\{\,z\in R^m : f_{1,l}(x_0,z)\;\le\;0\;;$$

$$f_i(x_0,z)+\nabla_z^T f_i(x_0,z)\big(-H_i^T(x_0,z)+H_{1,l}^T(x_0,z)\big)\Delta x\;\le\;0\,,\,i=l+1,\ldots,k\}\,. \tag{20}$$

Combining (19) and (20), we obtain

$$F(x_0+\Delta x)\;-\;o(\,\|\Delta x\|\,)\;=$$

$$\int\limits_{\substack{f_{1,l}(x_0,z)\le 0\,,\\ f_i(x_0,z)+\\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z)-H_i^T(x_0,z))\Delta x\le 0\,,\\ i=l+1,\ldots,k}} \Big\{p(x_0,z)+\big[\nabla_x^T p(x_0,z)+\mathrm{div}_z^T\big(p(x_0,z)H_{1,l}(x_0,z)\big)\big]\,\Delta x\Big\}\;\mathrm{d}z\;=$$

$$\int\limits_{\substack{f_{1,l}(x_0,z)\le 0\,,\\ f_i(x_0,z)+\\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z)-H_i^T(x_0,z))\Delta x\le 0\,,\\ i=l+1,\ldots,k}} p(x_0,z)\;\mathrm{d}z\;+ \tag{21}$$

$$\int\limits_{\substack{f_{1,l}(x_0,z)\le 0\,,\\ f_i(x_0,z)+\\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z)-H_i^T(x_0,z))\Delta x\le 0\,,\\ i=l+1,\ldots,k}} \big[\nabla_x^T p(x_0,z)+\mathrm{div}_z^T\big(p(x_0,z)H_{1,l}(x_0,z)\big)\big]\,\Delta x\;\mathrm{d}z\;. \tag{22}$$

Now we need the following lemma.

Lemma 2.2 *If a function $r : R\to R^k$ satisfies the condition*

$$r(||\Delta x||)\to 0\,,\quad ||\Delta x||\to 0\,,$$

then for sufficiently small $||\Delta x||$ the inequality

$$\Bigg|\;\int\limits_{\mu_{r(||\Delta x||)}(x_0)} \big[\nabla_x^T p(x_0,z)+\mathrm{div}_z^T\big(p(x_0,z)H_{1,l}(x_0,z)\big)\big]\,\Delta x\;\mathrm{d}z\;-$$

$$\left| \int_{\mu(x_0)} \left[\nabla_x^T p(x_0, z) + \mathrm{div}_z^T\big(p(x_0, z) H_{1,l}(x_0, z)\big) \right] \Delta x \; \mathrm{d}z \right| \leq T \|\Delta x\| \sum_{i=1}^{k} |r_i(\|\Delta x\|)| ,$$

holds, where T is a constant.

Proof. This lemma can be proved analogously to Lemma 2.1. Since the set $G_{V,\epsilon}(x_0)$ is bounded, conditions 2, 3, 5 of Theorem 2.1 imply

$$\left[\nabla_x^T p(x_0, z) + \mathrm{div}_z^T\big(p(x_0, z) H_{1,l}(x_0, z)\big) \right] \Delta x \leq C \|\Delta x\| , \quad C = \text{const} , \tag{23}$$

Analogously to inequality (7) we can prove

$$q\left(\mu(x_0) \, \Delta \, \mu_{r(\|\Delta x\|)}(x_0)\right) \leq C_1 \sum_{i=1}^{k} |r_i(\|\Delta x\|)| , \quad C_1 = \text{const} . \tag{24}$$

The statement of the lemma follows from inequalities (23) and (24). □

Since the set $G_{V,\epsilon}(x_0)$ is bounded, the conditions 2, 5, 6 of Theorem 2.1 imply

$$\nabla_z^T f_i(x_0, z)\big(H_{1,l}^T(x_0, z) - H_i^T(x_0, z)\big)\Delta x \leq C_2 \|\Delta x\| , \quad i = l+1, \ldots, k , \tag{25}$$

$$C_2 = \text{const} ,$$

for $(x_0 + \Delta x) \times z \in G_{V,\epsilon}(x_0)$.

Thus, with (25) and Lemma 2.2 for integral (22) we have

$$\int_{\substack{f_{1,l}(x_0,z) \leq 0 , \\ f_i(x_0,z) + \\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z) - H_i^T(x_0,z))\Delta x \leq 0 , \\ i = l+1,\ldots,k}} \left[\nabla_x^T p(x_0, z) + \mathrm{div}_z^T\big(p(x_0, z) H_{1,l}(x_0, z)\big) \right] \Delta x \; \mathrm{d}z =$$

$$\int_{\mu(x_0)} \left[\nabla_x^T p(x_0, z) + \mathrm{div}_z^T\big(p(x_0, z) H_{1,l}(x_0, z)\big) \right] \Delta x \; \mathrm{d}z + o(\|\Delta x\|) , \tag{26}$$

where $o(\|\Delta x\|) / \|\Delta x\| \to 0$ as $\|\Delta x\| \to 0$. Let us rewrite integral (21) in the following way

$$\int_{\substack{f_{1,l}(x_0,z) \leq 0 , \\ f_i(x_0,z) + \\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z) - H_i^T(x_0,z))\Delta x \leq 0 , \\ i = l+1,\ldots,k}} p(x_0, z) \; \mathrm{d}z =$$

$$F(x_0) + \int_{\substack{f_{1,l}(x_0,z) \leq 0 , \\ f_i(x_0,z) + \\ \nabla_z^T f_i(x_0,z)(H_{1,l}^T(x_0,z) - H_i^T(x_0,z))\Delta x \leq 0 , \\ i = l+1,\ldots,k}} p(x_0, z) \; \mathrm{d}z - \int_{\substack{f_{1,l}(x_0,z) \leq 0 , \\ f_i(x_0,z) \leq 0 , \\ i = l+1,\ldots,k}} p(x_0, z) \; \mathrm{d}z \stackrel{\text{def}}{=}$$

$$F(x_0) + D . \tag{27}$$

The difference D can be represented as a surface integral. Indeed, if

$$z \in \partial_{l+1,k}\mu(x_0) \;=\; \bigcup_{l \le i \le k} \Big(\mu(x_0) \bigcap \{\, y \in R^m \;:\; f_i(x_0, y) = 0\,\}\Big) \;,$$

and the surface $\partial_{l,k}\mu(x_0)$ is smooth around a point z , then condition 8 of Theorem 2.1 imply that the index set $I(x_0, z)$ consists only from one index $i(x_0, z)$. The distance between point z and a nearest point on the surface of the set

$$\{\, z \in R^m : f_{1,l}(x_0, z) \;\le\; 0\;;$$

$$f_i(x_0, z) + \nabla_z^T f_i(x_0, z)\big(H_{1,l}^T(x_0, z) - H_i^T(x_0, z)\big)\Delta x \;\le\; 0\,,\, i = l + 1, \ldots, k\,\}$$

is equal to

$$\left| \frac{\nabla_z^T f_{i(x_0,z)}(x_0, z)}{\|\nabla_z^T f_{i(x_0,z)}(x_0, z)\|} \big(H_{1,l}^T(x_0, z) - H_i^T(x_0, z)\big)\Delta x \right| \;+\; o(\,\|\Delta x\|\,) \;=$$

$$\left| \langle \big(H_{1,l}(x_0, z) - H_i(x_0, z)\big)\nabla_z^N f_{i(x_0,z)}(x_0, z)\,,\, \Delta x\rangle \right| \;+\; o(\,\|\Delta x\|\,)\,.$$

Thus, the difference D of the volume integrals can be represented as an integral over a surface

$$D \;=\; \int\limits_{\partial_{l+1,k}\mu(x_0)} \langle p(x_0, z)\left[H_{i(x_0,z)}(x_0, z) \;-\; H_{1,l}(x_0, z)\right]\nabla_z^N f_{i(x_0,z)}(x_0, z)\,,\, \Delta x\rangle\; \mathrm{d}z \;+ \qquad (28)$$

$$+\; o(\,\|\Delta x\|\,)\,,$$

where $o(\,\|\Delta x\|\,)\,/\,\|\Delta x\| \to 0$ as $\|\Delta x\| \to 0$. Combining (21), (22), (26), (27) and (28), we have

$$F(x_0 + \Delta x) \;-\; F(x_0) \;=\Big\langle \int\limits_{\mu(x_0)} \left[\nabla_x p(x_0, z) + \mathrm{div}_z\big(p(x_0, z)H_{1,l}(x_0, z)\big)\right] \mathrm{d}z \;+$$

$$+ \int\limits_{\partial_{l+1,k}\mu(x_0)} p(x_0, z)\left[H_{i(x_0,z)}(x_0, z) \;-\; H_{1,l}(x_0, z)\right]\nabla_z^N f_{i(x_0,z)}(x_0, z)\; \mathrm{d}z\,,\, \Delta x \Big\rangle \;+$$

$$+\; o(\,\|\Delta x\|\,)\,,$$

where $o(\,\|\Delta x\|\,)\,/\,\|\Delta x\| \to 0$ as $\|\Delta x\| \to 0$.

The last statement proves the theorem. □

3 An Example

The investigation of operational strategies for inspected components (see [6]) led us to the calculation of the gradient for the following integral function

$$F(x) = \int_{\substack{b(y)\le x\,, \\ y_i \ge \theta\,,\, i=1,\dots,m}} p(y)\,\mathrm{d}y\,, \tag{29}$$

where $x \in R^1$, $y \in R^m$, $p : R^m \to R^1$, $b(y) = \sum_{i=1}^{m} y_i^\alpha$, $\theta > 0$. In this case

$$f(x,y) = \begin{pmatrix} b(y) - x \\ \theta - y_1 \\ \vdots \\ \theta - y_m \end{pmatrix},$$

and

$$F(x) = \int_{f(x,y)\le 0} p(x,y)\,\mathrm{d}y = \int_{\mu(x)} p(x,y)\,\mathrm{d}y\,.$$

Here, we consider the case of $l = 1$ (see, condition 4 of Theorem 2.1). Equation (3) is represented as follows

$$H_{1,1}(x,y)\,\nabla_y f_1(x,y) + \nabla_x f_1(x,y) = 0\,. \tag{30}$$

Since

$$\nabla_y f_1(x,y) = \alpha \begin{pmatrix} y_1^{\alpha-1} \\ \vdots \\ y_m^{\alpha-1} \end{pmatrix}, \quad \nabla_x f_1(x,y) = -1\,,$$

then

$$H_{1,1} = h(y) \stackrel{\text{def}}{=} \big(h_1(y_1),\dots,h_m(y_m)\big) = \frac{1}{\alpha m}\left(y_1^{1-\alpha},\dots,y_m^{1-\alpha}\right) \tag{31}$$

is a solution to equation (30). We can write equations (4) as follows

$$H_i(x,y)\,\nabla_y f_i(y) + \nabla_x f_i(y) = 0\,, \quad i = 2,\dots,m+1\,. \tag{32}$$

The function $f_i(y)$ does not depend on x for $i = 2,\dots,m+1$ and $\nabla_x f_i(y) = 0$. Consequently,

$$H_i(x,y) = 0\,, \quad i = 2,\dots,m+1\,, \tag{33}$$

is a solution of the system of equations (32).

The set $\partial_{2,m+1}\mu(x)$ has a simple structure,

$$\partial_{2,m+1}\mu(x) = \bigcup_{2\le i\le m+1} \partial_i\mu(x) = \bigcup_{2\le i\le m+1} \left(\mu(x) \bigcap \{\, y \in R^m : y_i = \theta \,\}\right). \tag{34}$$

We can consider

$$\nabla_y^N f_{i(y)}(y) \;=\; e\big(i(y)\big)\,, \quad \text{for} \quad y \in \partial_{2,m+1}\mu(x)\,, \tag{35}$$

where

$$e\big(i(y)\big) \;=\; \begin{pmatrix} e_1 \\ \vdots \\ e_m \end{pmatrix},$$

$$e_j \;=\; 0 \;\; \text{for} \;\; j \neq i(y)\,; \quad e_j \;=\; -1 \;\; \text{for} \;\; j \;=\; i(y)\,,$$

$$i(y) \;=\; \min\{\, j \,:\, y_j = \theta \,\}\,.$$

In view of (6), (31), (33), (34) and (35), we have

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \mathrm{div}_y\big(p(y)h(y)\big)\,\mathrm{d}y \;+\; \int\limits_{\partial_{2,m+1}\mu(x)} p(y)\,\big(-h(y)\big)\,\nabla_y^N f_{i(y)}(y)\;\mathrm{d}S \;=$$

$$\int\limits_{\mu(x)} \mathrm{div}_y\big(p(y)h(y)\big)\,\mathrm{d}y \;+\; \sum_{i=1}^{m}\; \int\limits_{\partial_{i+1}\mu(x)} p(y)\,\big(-h(y)\big)e\big(i(y)\big)\;\mathrm{d}S \;=$$

$$\int\limits_{\mu(x)} \mathrm{div}_y\big(p(y)h(y)\big)\,\mathrm{d}y \;+\; \sum_{i=1}^{m} h_i(\theta) \int\limits_{\partial_{i+1}\mu(x)} p(y)\,\mathrm{d}S \;=$$

$$\int\limits_{\substack{b(y)\le x\,, \\ y_i \ge \theta\,,\; i=1,\dots,m}} \mathrm{div}_y\big(p(y)h(y)\big)\,\mathrm{d}y \;+\; \sum_{i=1}^{m} \frac{\theta^{1-\alpha}}{\alpha m} \int\limits_{\substack{b(\theta_i \mid y)\le x\,, \\ y^{-i}\ge\theta}} p(\theta_i \mid y)\;\mathrm{d}y^{-i}\,,$$

where

$$(\theta_i \mid y) \;=\; (y_1, \dots, y_{i-1}, \theta, y_{i+1}, \dots, y_m)\,,$$

$$y^{-i} \;=\; (y_1, \dots, y_{i-1}, y_{i+1}, \dots, y_m)\,,$$

$$b(\theta_i \mid y) \;=\; \theta^{\alpha} \;+\; \sum_{\substack{j=1 \\ j\neq i}}^{m} y_j^{\alpha}\,.$$

The inequality $y^{-i} \ge \theta$ should be treated as the set of inequalities

$$y_j \ge \theta\,, \quad j = 1, \dots, i-1, i+1, \dots, m\,.$$

Since

$$\mathrm{div}_y\big(p(y)h(y)\big) \;=\; h(y)\nabla_y p(y) \;+\; p(y)\mathrm{div}_y h(y) \;=$$

$$\frac{1}{\alpha m}\sum_{i=1}^{m} \frac{\partial p(y)}{\partial y_i} y_i^{1-\alpha} \;+\; p(y)\frac{1-\alpha}{\alpha m}\sum_{i=1}^{m} y_i^{-\alpha}\,,$$

then we finally have

$$\nabla_x F(x) = \int\limits_{\substack{b(y)\le x\,,\\ y_i\ge\theta\,,\,i=1,\dots,m}} \frac{1}{\alpha m}\sum_{i=1}^{m} y_i^{-\alpha}\big[y_i\frac{\partial p(y)}{\partial y_i} + (1-\alpha)p(y)\big]\,\mathrm{d}y\ +$$

$$+\ \frac{\theta^{1-\alpha}}{\alpha m}\sum_{i=1}^{m}\ \int\limits_{\substack{b(\theta_i|y)\le x\,,\\ y^{-i}\ge\theta}} p(\theta_i\mid y)\,\mathrm{d}y^{-i}\,.$$

The formula for the $\nabla_x F(x)$ is valid for an arbitrary sufficiently smooth function $p(y)$.

4 Summary

We have proved a general formula for the differentiation of an integral over the volume given by many inequalities. This formula can be used in different applied areas. One important area is the optimal control of systems with very high failure costs. In this case optimization should be made under the condition that the probability of failure is sufficiently small. A probability function can be represented as an integral over a set depending upon parameters. A gradient of the integral is expressed as a sum of integrals taken over volume and over surface. These results are very useful for the calculation of parameter sensitivities and the optimization of probability functions.

References

[1] Ermoliev, Yu. (1983): Stochastic Quasi-Gradient Methods and Their Applications to System Optimization. *Stochastics*, **4**. pp. 1-36.

[2] Pflug G.Ch. (1992): *Simulation and Optimization – the Interface*, Kluwer Academic Publishers (forthcoming).

[3] Prékopa, A. (1970): *On probabilistic constrained programming*, in: Proceedings of the Princeton Symposium on Mathematical Programming (Princeton University Press, Princeton, N.J., 1970) 113138.

[4] Prékopa, A. and T. Szántai (1978): *A new multivariate gamma distribution and its fitting to empirical streamflow data*, Water Resources Research, 14 (1978) 1924.

[5] Pulkkinen, A. and S. Urya'sev (1990): *Optimal Operational Strategies for an Inspected Component - Statement of the Problem.* Working Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, WP-90-62, 22 p.

[6] Pulkkinen, A. and S. Urya'sev (1991): *Optimal Operational Strategies for an Inspected Component - Solution Techniques.* Collaborative Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, CP-91-13, 16 p.

[7] Raik, E. (1975): *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** pp. 3-6 (in Russian).

[8] Roenko, N. (1983): *Stochastic Programming Problems with Integral Functionals over Multivalued Mappings.* Synopsis of Ph.D. Thesis, USSR, Kiev, (in Russian).

[9] Simon, J. (1989): *Second Variation in Domain Optimization Problems,* In "International Series of Numerical Mathematics", **91**, Eds. F.Kappel, K.Kunish and W.Schappacher, Birkhauser Verlag, pp. 361–378

[10] Uryas'ev, S. (1988): *Differentiability of an Integral over a Set Defined by Inclusion.* Kibernetika (Kiev) **5**, (1988), 83–86 (in Russian). Cybernetics, v.24, **5**, (1988), 638–642 (in English).

[11] Uryas'ev, S. (1989): *A Differentiation Formula for Integrals over Sets Given by Inclusion.* Numerical Functional Analysis and Optimization. **10(7 & 8)**, 827–841.