

Derivatives of probability functions and some applications

Stanislav Uryasev*

International Institute for Applied Systems Analysis, A-2361 Laxenburg, Austria

Probability functions depending upon parameters are represented as integrals over sets given by inequalities. New derivative formulas for the integrals over a volume are considered. Derivatives are presented as sums of integrals over a volume and over a surface. Two examples are discussed: probability functions with linear constraints (random right-hand sides), and a dynamical shut-down problem with sensors.

Keywords: Probability functions, gradient of integral, sensitivity analysis, optimization, discrete event dynamic systems, shut-down problem, probabilistic risk analysis.

1. Introduction

Probability functions are important in many applications; they are widely used for probabilistic risk analysis (see, for example [1, 18, 23]), in optimizing of discrete event systems (see, for example [9, 17]), and other applications. Probability functions can be represented as integrals over sets given by inequalities. The sensitivity analysis and the optimization of these functions require the calculation of the derivatives with respect to parameters. To date, the theory for the differentiation of such integrals is not fully developed. Here, we discuss a general formula for the differentiation of an integral over a volume given by many inequalities. A gradient of the integral is represented as the sum of integrals taken over a volume and over a surface. We have used these formulas for different applications – for calculating the sensitivities of probability functions, and for chance-constrained optimization.

A full proof of the differentiation formula is presented in [22]. We give an idea of the alternative proof of the main theorem in the appendix. The differentiation formula is explained with two applications:

- The linear case – the probability functions with linear constraints and random right-hand sides. The probability function with a random matrix is considered in [22].
- A shutdown problem with sensors. The problem was studied by the author jointly with Prof. Yu. Ermoliev. The approach can be used, for example, to

* Present address: Brookhaven National Laboratory, Building 130, Upton, NY 11973, USA.

monitor passive components (the vessel of the nuclear power plant [12, 13]). This problem can be considered as a typical example of Discrete Event Dynamic Systems (DEDS). Sensitivity analysis and optimization techniques for similar problems can be found in [4, 5, 7, 17].

Let the function

$$F(x) = \int_{f(x,y) \leq 0} p(x,y) dy \quad (1)$$

be defined in the Euclidean space \mathbb{R}^n , where $f: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ and $p: \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ are some functions. The inequality $f(x,y) \leq 0$, actually, is a system of inequalities

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

Stochastic programming problems lead to such functions. For example, let

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

be a probability function, where $\zeta(\omega)$ is a random vector in \mathbb{R}^m . The random vector $\zeta(\omega)$ is assumed to have a probability density $p(x,y)$ that depends on a parameter $x \in \mathbb{R}^n$.

The differentiation formulas for function (1) in the case of only one inequality ($k = 1$) are described in papers by Raik [14] and Roenko [15]. More general results ($k \geq 1$) were given by Simon (see, for example, [19]). Special cases of probability function (2) with normal and gamma distributions were investigated by Prékopa [10], and Prékopa and Szantái [11]. In the forthcoming book by Pflug [9], the gradient of function (1) is represented in the form of a conditional expectation ($k = 1$). The gradient of the probability function can be approximated as the gradient of some other smooth function; see, for example, Ermoliev et al. [2].

The gradient expressions given in [14, 15, 19] have the form of surface integrals and are often inconvenient for computation, since the measure of a surface in \mathbb{R}^m equals zero.

In [20, 21], 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 [3] can be used for the minimization of function (1). Here, we consider the formula for the general case of $k \geq 1$; the formulas in [14] and [20] are special cases of this general result. Since the gradient of function (1) is presented in [20] and [21] as an integral over a volume, in the case of $k = 1$ it is clear that this integral can be reduced to an integral over a surface (see [14]). Furthermore, the gradient of function (1) can also be represented as the sum of integrals taken over a volume and over a surface (in the case of $k \geq 1$). This formula is especially convenient for the case

when the inequalities $f(x, y) \leq 0$ include the simple constraints $y_i \geq 0, i = 1, \dots, m$ (see also the examples in [22]).

It is also shown that the general differentiation formula covers the "change of variables" approach, considered under different names: "transformation of variables" method [8], and "push out" method [16, 17].

2. The general formula

Let us introduce the shorthand notations

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

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

A transposed matrix H is denoted by H^T , and the Jacobian of the function $f(x, y)$ is denoted by $\nabla_y^T f(x, y) = (\nabla_y f(x, y))^T$. Let H be some matrix

$$H = \begin{pmatrix} h_{11}, & \dots, & h_{1m} \\ \vdots \\ h_{n1}, & \dots, & h_{nm} \end{pmatrix};$$

further, we need a definition of divergence for the matrix H

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

We also define

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

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

$$\partial_i\mu(x) = \mu(x) \cap \{y \in \mathbb{R}^m : f_i(x, y) = 0\}.$$

Further, we consider that for a point x , all functions $f_i(x, y)$, $i = 1, \dots, k$, are active, i.e. $\partial_i\mu(x) \neq \emptyset$ for $i = 1, \dots, k$. For $y \in \partial\mu(x)$ we define

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

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

$$K_1 = \{1, \dots, l\} \quad \text{and} \quad K_2 = \{l+1, \dots, k\}.$$

We formulate a theorem about the derivatives of integral (1).

THEOREM 2.1

Let us assume that the following conditions are satisfied:

- (1) at the point x , all functions $f_i(x, y)$, $i = 1, \dots, k$, are active;
- (2) the set $\mu(z)$ is bounded in a neighborhood of the point x ;
- (3) the function $f : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^k$ has continuous partial derivatives $\nabla_x f(x, y)$, $\nabla_y f(x, y)$;
- (4) the function $p : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ has continuous partial derivatives $\nabla_x p(x, y)$, $\nabla_y p(x, y)$;
- (5) there exists a continuous matrix function $H_l : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^{n \times m}$ satisfying the equation

$$H_l(x, y)\nabla_y f_{1l}(x, y) + \nabla_x f_{1l}(x, y) = 0; \quad (3)$$

- (6) the matrix function $H_l(x, y)$ has a continuous partial derivative $\nabla_y H_l(x, y)$;
- (7) the gradient $\nabla_y f_i(x, y)$ is not equal to zero on $\partial_i\mu(x)$ for $i = 1, \dots, k$;
- (8) for each $y \in \partial\mu(x)$, the vectors $\nabla_y f_i(x, y)$, $i \in I(x, y)$, are linearly independent.

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

$$\begin{aligned} \nabla_x F(x) = & \int_{\mu(x)} [\nabla_x p(x, y) + \text{div}_y(p(x, y)H_l(x, y))] dy \\ & - \sum_{i=l+1}^k \int_{\partial_i\mu(x)} \frac{p(x, y)}{\|\nabla_y f_i(x, y)\|} [\nabla_x f_i(x, y) + H_l(x, y)\nabla_y f_i(x, y)] dS. \quad (4) \end{aligned}$$

Remark

In theorem 2.1 we consider the case when the subsets K_1 and K_2 are non-empty. If the set K_1 is empty, then matrix $H_l(x, y)$ is not included in the formula and

$$\nabla_x F(x) = \int_{\mu(x)} \nabla_x p(x, y) dy - \sum_{i=1}^k \int_{\partial_i \mu(x)} \frac{p(x, y)}{\|\nabla_y f_i(x, y)\|} \nabla_x f_i(x, y) dS. \quad (5)$$

If the set K_2 is empty, then the integral over the surface is absent and

$$\nabla_x F(x) = \int_{\mu(x)} [\nabla_x p(x, y) + \operatorname{div}_y(p(x, y)H_k(x, y))] dy. \quad (6)$$

A full proof of theorem 2.1 is presented in [22]. This proof contains all technical details which are difficult to understand. An alternative, much more transparent idea of how to prove the main formula of theorem 2.1 is shown in the appendix. The alternative proof has two major steps:

- (1) presenting the gradient of the probability function as an integral over the surface (extended variant of the Raik theorem);
- (2) using the Ostrogradski–Gauss theorem to link the integral over surface and volume.

2.1. DISCUSSION OF THE FORMULA FOR THE GRADIENT OF THE PROBABILITY FUNCTIONS

The general formula (4) for calculating the derivatives of the probability functions shows that there are many equivalent expressions for these derivatives. The following components in this formula are not uniquely defined:

- two subsets K_1 and K_2 ,
- matrix $H_l(x, y)$,
- different vector functions $f(x, y)$ may present the same integration set $\mu(x)$.

The set K_2 defines an area integration over the surface. Usually, it is preferable to choose the set K_2 to be as small as possible, because the integral over the surface is often difficult to calculate numerically. In most cases, it is possible to set $K_2 = \emptyset$, so the gradient is presented as an integral over volume with formula (6).

The matrix $H_l(x, y)$ is a solution of the nonlinear system of equations (3) and, as a rule, is not uniquely defined. As indicated in [22], equation (3) can be solved explicitly. The matrix

$$-\nabla_x f_{li}(x, y) (\nabla_y^T f_{li}(x, y) \nabla_y f_{li}(x, y))^{-1} \nabla_y^T f_{li}(x, y) \quad (7)$$

is one possible solution, but it leads to complicated formulas, and, usually, is not used in practice.

In many cases, there is a simple way to solve equations (3) using a change of variables. Suppose that there is a change of variables

$$y = \gamma(x, z) \quad (8)$$

which eliminates the vector x from the function $f_l(x, y)$, i.e., the function $f_l(x, \gamma(x, z))$ does not depend upon the variable x . Denote by $\gamma^{-1}(x, y)$ the inverse function, defined by the equation

$$\gamma^{-1}(x, \gamma(x, z)) = z.$$

In this case, equation (3) has the following solution

$$H_l(x, y) = \nabla_x \gamma(x, z)|_{z=\gamma^{-1}(x, y)}. \quad (9)$$

Indeed, the gradient of the function $\gamma(x, y(x, z))$ with respect to x equals zero; therefore,

$$0 = \nabla_x f_l(x, \gamma(x, z)) = \nabla_x \gamma(x, z) \nabla_y f_l(x, y)|_{y=\gamma(x, z)} + \nabla_x f_l(x, y)|_{y=\gamma(x, z)},$$

i.e., function $\nabla_x \gamma(x, z)|_{z=\gamma^{-1}(x, y)}$ is a solution of equation (3). This special case covers the "change of variables" approach, considered previously under different names: the "transformation of variables" method [8] and "push out" method [16, 17]. This approach eliminates vector x from the integration set by changing variables in integral (1) with formula (8). Then, the well-known formula for the interchange of integral and gradient signs is used to calculate the gradient. Further, inverse transformation $z = \gamma^{-1}(x, y)$ is applied to return back to the original variables y . This multi-step procedure can be avoided by using the special case formula (6) directly with matrix (9), i.e.,

$$\nabla_x F(x) = \int_{\mu(x)} [\nabla_x p(x, y) + \text{div}_y(p(x, y) \nabla_x \gamma(x, z)|_{z=\gamma^{-1}(x, y)})] dy. \quad (10)$$

There are two advantages in using formula (4) with matrix (9) compared to the change of variables approach: First, it is not necessary to change variables twice, and to calculate Jacobians of transformations. Second, the change of variables approach is applicable only in a special case when the gradient can be presented as an integral over volume with formula (10), but it is not applicable when the gradient is presented with formula (4) as the sum of integrals over the volume and over the surface.

As mentioned above, different vector functions $f(x, y)$ may present the same integration set $\mu(x)$, leading to quite different equivalent formulas for the gradients. Moreover, with some vector functions $f(x, y)$, it is possible to set $K_2 = \emptyset$ and exclude integration over surface; with other functions, it is impossible. Different gradient formulas, in turn, generate quite different stochastic estimates of gradients (stochastic quasi-gradients [3]) with significantly different variance properties. Let us explain this with a trivial example:

$$F(x) = \int_{0 \leq y \leq x} p(y) dy = \int_{f(x,y) \leq 0} p(y) dy, \quad (11)$$

where

$$f(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} = \begin{pmatrix} y - x \\ -y \end{pmatrix}.$$

It is not possible to set

$$K_1 = \{1, 2\} \quad \text{and} \quad K_2 = \emptyset$$

in this case, because equation (3) does not have any solution with $l = 2$. Using formula (4) with

$$K_1 = \{1\} \quad \text{and} \quad K_2 = \{2\},$$

the gradient of the function $F(x)$ can be expressed by solving equation (3). This equation links the gradients of the function $f_1(x, y)$ with respect to x and y

$$H_1(x, y) \nabla_y f_1(x, y) + \nabla_x f_1(x, y) = 0.$$

The equation has an evident solution

$$H_1(x, y) = 1. \quad (12)$$

We also need the gradients of the function $f_2(x, y)$ w.r.t. parameters y and x

$$\nabla_y f_2(x, y) = -1, \quad \nabla_x f_2(x, y) = 0. \quad (13)$$

The gradient of the function $F(x)$ is calculated with formula (4)

$$\nabla_x F(x) = \int_{f(x,y) \leq 0} [\nabla_x p(y) + \text{div}_y(p(y)H_1(x, y))] dy$$

$$\begin{aligned}
& - \int_{y=0} \frac{p(y)}{\|\nabla_y f_2(x, y)\|} [\nabla_x f_2(x, y) + H_1(x, y) \nabla_y f_2(x, y)] dS \\
& = \int_{0 \leq y \leq x} \nabla_y p(y) dy + p(0).
\end{aligned} \tag{14}$$

Thus, the derivative $\nabla_x F(x)$ is expressed as the sum of an integral over a volume and an integral over a surface.

If $x \geq 0$, then function $F(x)$ defined by formula (11) can be equivalently presented with the vector function

$$f(x, y) = \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} = \begin{pmatrix} y/x - 1 \\ -y/x \end{pmatrix}.$$

Evidently, the change of variables

$$y = \gamma(x, z) = xz \tag{15}$$

eliminates the vector x from the function $f(x, y)$. Therefore, we can set

$$K_1 = \{1, 2\}, \quad K_2 = \emptyset, \quad l = k = 2,$$

and equation (3) has the solution defined by equation (9)

$$H_2(x, y) = \nabla_x \gamma(x, z)|_{z=\gamma^{-1}(x, y)} = \nabla_x xz|_{z=y/x} = y/x. \tag{16}$$

Finally, with formula (6) or (10)

$$\nabla_x F(x) = x^{-1} \int_{0 \leq y \leq x} \nabla_y (p(y)y) dy. \tag{17}$$

Expressions (14) and (17) for the gradient do not coincide; it can be shown that they are equivalent functions.

The next section describes two examples demonstrating possible applications of the formula for the derivatives of probability functions.

3. Examples

3.1. EXAMPLE 1: LINEAR CASE – RANDOM RIGHT-HAND SIDES WITH SIMPLE CONSTRAINTS

Here, we consider a probability linear function with right-hand sides. The probability function with random matrix is considered in [22]. Let A be an

$m \times n$ matrix, (P, \mathcal{F}, Ω) be a probability space, and $b(\omega)$, $\omega \in \Omega$ be a random m -dimensional vector with the joint density $p(b)$. We define

$$F(x) = P\{Ax \leq b(\omega), b(\omega) \geq 0\}, \quad b = (b_1(\omega), \dots, b_m(\omega)) \in \mathbb{R}^m, \quad x \in \mathbb{R}^n, \quad (18)$$

i.e. $F(x)$ is the probability that the linear constraints $Ax \leq b(\omega)$, $b(\omega) \geq 0$ are satisfied. The constraint $b(\omega) \geq 0$ means non-negativity of all elements $b_j(\omega)$ of the vector $b(\omega)$. Let us denote by A_i the i th row of the matrix A

$$A = \begin{pmatrix} A_1 \\ \vdots \\ A_l \end{pmatrix} = \begin{pmatrix} (a_{11}, \dots, a_{1n}) \\ \vdots \\ (a_{l1}, \dots, a_{ln}) \end{pmatrix}.$$

Define the function $f(x, b)$ as

$$f(x, b) = \begin{pmatrix} f_1(x, b) \\ \vdots \\ f_k(x, b) \end{pmatrix} = \begin{pmatrix} A_1 x - b_1 \\ \vdots \\ A_l x - b_l \\ -b_1 \\ \vdots \\ -b_m \end{pmatrix}, \quad k = 2m.$$

The function $F(x)$ equals

$$F(x) = \int_{f(x, b) \leq 0} p(b) db = \int_{\substack{Ax \leq b \\ b \geq 0}} p(b) db. \quad (19)$$

PROPOSITION 3.1

The gradient of the function $F(x)$ can be presented as a sum of an integral over the volume and an integral over the surface

$$\nabla_x F(x) = \int_{\substack{Ax \leq b \\ b \geq 0}} A^T \nabla_b p(b) db + \sum_{i=1}^m \int_{\substack{Ax \leq b \\ b_i = 0}} A_i^T p(b) dS; \quad (20)$$

if the density function $p(b)$ equals zero on the boundary of the set $\{b \in \mathbb{R}^m : b \geq 0\}$,

then the integral over the surface equals zero, and

$$\nabla_x F(x) = \int_{\substack{Ax \leq b \\ b \geq 0}} A^T \nabla_b p(b) db = \int_{\substack{Ax \leq b \\ b \geq 0}} A^T \nabla_b (\ln p(b)) p(b) db. \quad (21)$$

Remark

The integral over the surface

$$\int_{\substack{Ax \leq b \\ b_i = 0}} A_i^T p(b) dS$$

is, evidently, an integral over an $(m-1)$ -dimensional volume, without variable b_i , which is fixed to zero.

We use formula (4) to calculate the gradient $\nabla_x F(x)$. Let us consider that $l = m$, $K_1 = \{1, \dots, m\}$ and $K_2 = \{m+1, \dots, k\}$. For this case, equation (3) is presented as

$$H_l(x, b) \nabla_b f_{1l}(x, b) + \nabla_x f_{1l}(x, b) = H_l(x, b)(-E) + A^T = 0. \quad (22)$$

Therefore,

$$H_l(x, b) = A^T.$$

Formula (4) and the last equality imply

$$\begin{aligned} \nabla_x F(x) &= \int_{\substack{Ax \leq b \\ b \geq 0}} \operatorname{div}_b (p(b) A^T) db \\ &\quad - \sum_{i=1}^k \int_{\substack{Ax \leq b \\ b_{i-1} = 0}} \frac{p(b)}{\|\nabla_b f_i(x, b)\|} [\nabla_x f_i(x, b) + A^T \nabla_b f_i(x, b)] dS. \end{aligned} \quad (23)$$

Since

$$\|\nabla_b f_i(x, b)\| = 1, \quad \nabla_x f_i(x, b) = 0; \quad i = m, \dots, 2m,$$

and

$$A^T \nabla_b f_i(x, b) = A_{i-m}^T; \quad i = m, \dots, 2m,$$

then (23) implies (20).

Formula (21) follows directly from (20), if the density function $p(b)$ equals zero on the boundary of the set $\{b \in \mathbb{R}^m : b \geq 0\}$.

3.2. EXAMPLE 2: A "SHUTDOWN" PROBLEM WITH SENSORS

In this section, we discuss a shutdown problem for a system with sensors. Usually, some measurements are made, and a decision to shut down the system is based on these measurements. In different situations, different information is available. For example, for monitoring the vessel of a nuclear power plant, some estimates of crack sizes can be used [12]; the time required to achieve full power, leakages, vibration, and corrosion are of interest for diesels, pumps, and other active components [6].

We consider that the dynamics of the system is described by discrete time equations

$$z^{t+1} = \psi^t(z^t, u^t, \zeta^t), \quad t = 1, \dots, T, \quad (24)$$

where z^t is a state vector in the Euclidean space \mathbb{R}^{j_t} . The functions

$$\psi^t : \mathbb{R}^{j_t} \times \mathbb{R}^{n_t} \times \mathbb{R}^{r_t} \rightarrow \mathbb{R}^{j_{t+1}}, \quad t = 1, \dots, T,$$

depend upon the state vector z^t , control vector u^t and random vector ζ^t . The system has been shut down at time t if an equality

$$\varphi_t(z^t, u^t, \eta^t) \leq 0 \quad (25)$$

is satisfied, where

$$\varphi_t : \mathbb{R}^{j_t} \times \mathbb{R}^{n_t} \times \mathbb{R}^{m_t} \rightarrow \mathbb{R}, \quad t = 1, \dots, T,$$

and η^t is a random vector. If the system has not been shut down at time t , $1 \leq t \leq T$, then it finishes operation at time $T + 1$. Thus, the shutdown time τ is given by the equation

$$\tau = \begin{cases} T + 1, & \text{if } \varphi_t(z^t, u^t, \eta^t) > 0, & \text{for } 1 \leq t \leq T; \\ \min\{t : 1 \leq t \leq T, \varphi_t(z^t, u^t, \eta^t) \leq 0\}, & \text{otherwise.} \end{cases}$$

If the system has been shut down at time t , $1 \leq t \leq T$, then the cost of this event equals $g_t(z^{1t}, u^{1t})$, where $z^{1t} = (z^1, \dots, z^t)$ and $u^{1t} = (u^1, \dots, u^t)$, and, at time $T + 1$, the cost equals $g_{T+1}(z^{1T}, u^{1T})$. As a criterion function, we take the expectation

$$G(u^{1T}) = \mathbb{E} g_\tau(z^{1\tau}, u^{1\tau}).$$

Denote

$$\eta^{1T} = (\eta^1, \dots, \eta^T),$$

$$f^1(z^1, u^1, \eta^1) = \varphi_1(z^1, u^1, \eta^1),$$

$$f^t(z^{1t}, u^{1t}, \eta^{1t}) = \begin{pmatrix} -\varphi_1(z^1, u^1, \eta^1) \\ \vdots \\ -\varphi_{t-1}(z^{t-1}, u^{t-1}, \eta^{t-1}) \\ \varphi_t(z^t, u^t, \eta^t) \end{pmatrix}, \quad t = 2, \dots, T,$$

$$f^{T+1}(z^{1T}, u^{1T}, \eta^{1T}) = - \begin{pmatrix} \varphi_1(z^1, u^1, \eta^1) \\ \vdots \\ \varphi_T(z^T, u^T, \eta^T) \end{pmatrix},$$

$$I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}} = \begin{cases} 1, & \text{if } f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0; \\ 0, & \text{otherwise.} \end{cases}$$

Further, it is convenient to use the following notations:

$$z^{1, T+1} = z^{1T}, \quad u^{1, T+1} = u^{1T}, \quad \eta^{1, T+1} = \eta^{1T}.$$

For example, $f^{T+1}(z^{1T}, u^{1T}, \eta^{1T})$ can also be denoted by $f^{T+1}(z^{1, T+1}, u^{1, T+1}, \eta^{1, T+1})$. Suppose that there are density functions for the random vectors ζ^t, η^t , $t = 1, \dots, T$. The criterion function can be presented as

$$\begin{aligned} G(u^{1T}) &= \mathbb{E} g_T(z^{1T}, u^{1T}) = \mathbb{E} \left[\sum_{t=1}^{T+1} g_t(z^{1t}, u^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}} \right] \\ &= \sum_{t=1}^{T+1} \mathbb{E} [g_t(z^{1t}, u^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}}]. \end{aligned} \quad (26)$$

With formula (6) we can express the gradient of the expectation of a product of nonlinear and indicator functions as the expectation of a product of another nonlinear and the same indicator function

$$\nabla_{u^{1T}} \mathbb{E} [g_t(z^{1t}, u^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}}] = \mathbb{E} [a_t(z^{1t}, u^{1t}, \zeta^{1t}, \eta^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}}]. \quad (27)$$

Thus, (26) and (27) imply

$$\begin{aligned}
 \nabla_{u^{1T}} G(u^{1T}) &= \nabla_{u^{1T}} \mathbb{E} g_T(z^{1T}, u^{1T}) \\
 &= \sum_{t=1}^{T+1} \nabla_{u^{1T}} \mathbb{E} [g_t(z^{1t}, u^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}}] \\
 &= \sum_{t=1}^{T+1} \mathbb{E} [a_t(z^{1t}, u^{1t}, \zeta^{1t}, \eta^{1t}) I_{\{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0\}}] \\
 &= \mathbb{E} a_T(z^{1T}, u^{1T}, \zeta^{1T}, \eta^{1T}). \tag{28}
 \end{aligned}$$

Formula (28) is valid in rather general cases, but to use it we have to find the functions

$$a_t(z^{1t}, u^{1t}, \zeta^{1t}, \eta^{1t}), \quad t = 1, \dots, T + 1.$$

We show with one special case how it can be done.

A special case

This is a special case of the problem of optimizing operational schedules for mechanical components [12]. The mechanical component has some defects (cracks) which evolve and increase the failure rate of the component. The component is inspected periodically to assure that the failure rate is within specified limits (safety constraints). If the estimate of the failure rate of the component does not satisfy the safety constraints, the operation of the component is terminated. The model originally considered in [12] also includes other actions and decision rules:

- additional intensive testing;
- repair of the component.

Here, we consider only part of the model, related to shutting down the component.

In this case, the variables described in the general shutdown model (24) and (25) have the following meaning:

- $t = 1, \dots, T$ – number of time points where inspections are performed;
- r_t – number of defects at time t ;
- $\zeta^t = (\zeta_1^t, \dots, \zeta_{r_t}^t)$ – vector of the sizes of the defects (cracks);
- $\eta^t = (\eta_1^t, \dots, \eta_{r_t}^t)$ – vector of uncertainties in the measurements of defects (errors of sensors);
- z_t – estimate of the failure rate of the component.

Two dynamic processes are modelled. The first stochastic process defines the

evolution of crack sizes

$$\zeta^{t+1} = h_t(\zeta^t), \quad t = 1, \dots, T. \quad (29)$$

The second stochastic process specifies the dynamics of the failure rate

$$z_{t+1} = \psi_t(z_t, \zeta^t), \quad t = 1, \dots, T. \quad (30)$$

Here, z_t is a scalar value and the function ψ_t does not depend on the control vector u^{1T} . The function φ_t defines a shutdown condition:

if $\varphi_t(z_t, u_t, \eta^t) > 0$, then continue operation of the component;
if $\varphi_t(z_t, u_t, \eta^t) \leq 0$, then shut down the component.

Crack sizes cannot be measured perfectly; therefore, uncertainties $\eta^t = (\eta_1^t, \dots, \eta_r^t)$ in measurements influence the decision. In this special case, the function φ_t is linear with respect to z_t and scalar control u_t , i.e.,

$$\varphi_t(z_t, u_t, \eta^t) = z_t - u_t + b(\eta^t), \quad (31)$$

where $b(\eta^t) = \sum_{i=1}^m (\eta_i^t)^\alpha$, $\alpha > 0$. Actually, the shutdown condition compares failure rate z_t with cutoff value u_t . The function $b(\eta^t)$ reflects the existence of uncertainties in the shutdown condition. This is explained in detail in [12].

If the component has been shut down at time t , $1 \leq t \leq T$, then the cost of this event $g_t(z_t)$ depends only upon z_t and the function $g_{T+1}(z_T)$ depends upon z_T . The random vectors, η^1, \dots, η^T , specifying uncertainties in measurements, are supposed to be independent and have density functions $\rho_1(\eta^1), \dots, \rho_T(\eta^T)$, respectively. Denote by \mathbb{E}_ζ the conditional expectation with respect to σ -algebra \mathcal{F}_ζ generated by the random values ζ^t , $t = 1, \dots, T$. In this case, the criterion function can be presented as

$$G(u^{1T}) = \mathbb{E} g_T(z_T) = \mathbb{E} \mathbb{E}_\zeta g_T(z_T) = \sum_{t=1}^{T+1} \mathbb{E} \mathbb{E}_\zeta [g_t(z_t) I_{\{f^t(z^t, u^t, \eta^t) \leq 0\}}].$$

Since the value $g_t(z_t)$ is measurable with respect to σ -algebra \mathcal{F}_ζ , then

$$\mathbb{E}_\zeta [g_t(z_t) I_{\{f^t(z^t, u^t, \eta^t) \leq 0\}}] = g_t(z_t) \mathbb{E}_\zeta I_{\{f^t(z^t, u^t, \eta^t) \leq 0\}}.$$

Therefore,

$$G(u^{1T}) = \sum_{t=1}^{T+1} \mathbb{E} [g_t(z_t) \mathbb{E}_\zeta I_{\{f^t(z^t, u^t, \eta^t) \leq 0\}}]. \quad (32)$$

Conditional expectations can be presented as follows:

$$\begin{aligned} \mathbb{E}_{\zeta} I\{f^{i(z^i, u^i, \eta^i)} \leq 0\} &= \int_{f^{i(z^i, u^i, \eta^i)} \leq 0} \prod_{\theta=1}^i \rho_{\theta}(\eta^{\theta}) d\eta^{1^i}, \quad i = 1, \dots, T, \\ \mathbb{E}_{\zeta} I\{f^{T+1(z^T, u^T, \eta^T)} \leq 0\} &= \int_{f^{T+1(z^T, u^T, \eta^T)} \leq 0} \prod_{\theta=1}^T \rho_{\theta}(\eta^{\theta}) d\eta^{1^T}. \end{aligned}$$

Consequently,

$$G(u^{1^T}) = \sum_{i=1}^T \mathbb{E} \left[g_i(z_i) \int_{f^{i(z^i, u^i, \eta^i)} \leq 0} \prod_{\theta=1}^i \rho_{\theta}(\eta^{\theta}) d\eta^{1^i} \right] \tag{33}$$

$$+ \mathbb{E} \left[g_{T+1}(z_T) \int_{f^{T+1(z^T, u^T, \eta^T)} \leq 0} \prod_{\theta=1}^T \rho_{\theta}(\eta^{\theta}) d\eta^{1^T} \right]. \tag{34}$$

In the last equation, the functions under the expectation sign \mathbb{E} are smooth; therefore, under general conditions,

$$\begin{aligned} \nabla_{u^{1^T}} G(u^{1^T}) &= \nabla_{u^{1^T}} \sum_{i=1}^{T+1} \mathbb{E}[g_i(z_i) \mathbb{E}_{\zeta} I\{f^{i(z^i, u^i, \eta^i)} \leq 0\}] \\ &= \sum_{i=1}^{T+1} \mathbb{E}[g_i(z_i) \nabla_{u^{1^T}} \mathbb{E}_{\zeta} I\{f^{i(z^i, u^i, \eta^i)} \leq 0\}] \\ &= \sum_{i=1}^T \mathbb{E} \left[g_i(z_i) \nabla_{u^{1^T}} \int_{f^{i(z^i, u^i, \eta^i)} \leq 0} \prod_{\theta=1}^i \rho_{\theta}(\eta^{\theta}) d\eta^{1^i} \right] \\ &\quad + \mathbb{E} \left[g_{T+1}(z_T) \nabla_{u^{1^T}} \int_{f^{T+1(z^T, u^T, \eta^T)} \leq 0} \prod_{\theta=1}^T \rho_{\theta}(\eta^{\theta}) d\eta^{1^T} \right]. \end{aligned} \tag{35}$$

First, let us calculate the gradient of the functions

$$\phi_i(u^{1^i}) \stackrel{\text{def}}{=} \int_{f^{i(z^i, u^i, \eta^i)} \leq 0} \prod_{\theta=1}^i \rho_{\theta}(\eta^{\theta}) d\eta^{1^i}, \quad i = 1, \dots, T, \tag{36}$$

with respect to u^{1T} . Since the function $\phi_t(u^{1t})$ does not depend upon u_{t+1}, \dots, u_T , then

$$\nabla_{u^{1T}} \phi_t(u^{1t}) = \begin{pmatrix} \nabla_{u^{1t}} \phi_t(u^{1t}) \\ 0 \end{pmatrix}. \quad (37)$$

Further, let us calculate the gradient $\nabla_{u^{1t}} \phi_t(u^{1t})$ with formula (6). For this case, equation (3) is presented as follows

$$H_t \nabla_{\eta^{1t}} f^t(z^{1t}, u^{1t}, \eta^{1t}) + \nabla_{u^{1t}} f^t(z^{1t}, u^{1t}, \eta^{1t}) = 0. \quad (38)$$

Since

$$f^t(z^{1t}, u^{1t}, \eta^{1t}) = \begin{pmatrix} -z_1 + u_1 - b(\eta^1) \\ \vdots \\ -z_{t-1} + u_{t-1} - b(\eta^{t-1}) \\ z_t - u_t + b(\eta^t) \end{pmatrix}, \quad t = 2, \dots, T,$$

then gradients $\nabla_{\eta^{1t}} f^t(z^{1t}, u^{1t}, \eta^{1t})$ and $\nabla_{u^{1t}} f^t(z^{1t}, u^{1t}, \eta^{1t})$ can be easily calculated

$$\nabla_{\eta^{1t}} f^t(z^{1t}, u^{1t}, \eta^{1t}) = \alpha \begin{pmatrix} - \begin{pmatrix} (\eta_1^1)^{\alpha-1} \\ \vdots \\ (\eta_m^1)^{\alpha-1} \end{pmatrix} & & & 0 \\ & \ddots & & \\ & & - \begin{pmatrix} (\eta_1^{t-1})^{\alpha-1} \\ \vdots \\ (\eta_m^{t-1})^{\alpha-1} \end{pmatrix} & \\ & & & \begin{pmatrix} (\eta_1^t)^{\alpha-1} \\ \vdots \\ (\eta_m^t)^{\alpha-1} \end{pmatrix} \\ & 0 & & \end{pmatrix},$$

$$\nabla_{u^{1T}} f^t(z^{1t}, u^{1t}, \eta^{1t}) = \begin{pmatrix} 1 & & 0 \\ & \ddots & \\ & & 1 \\ 0 & & -1 \end{pmatrix}.$$

The matrix

$$H_t = \frac{1}{\alpha m} \begin{pmatrix} ((\eta_1^1)^{1-\alpha}, \dots, (\eta_m^1)^{1-\alpha}) & 0 \\ & \ddots \\ 0 & ((\eta_1^t)^{1-\alpha}, \dots, (\eta_m^t)^{1-\alpha}) \end{pmatrix} \tag{39}$$

is a solution of equation (38). Now formula (6) implies

$$\begin{aligned} \nabla_{u^{1t}} \phi_t(u^{1t}) &= \int_{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0} \operatorname{div}_{\eta^{1t}} \left(\prod_{\theta=1}^t \rho_\theta(\eta^\theta) H_t \right) d\eta^{1t} \\ &= \int_{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0} \frac{1}{\alpha m} \begin{pmatrix} \prod_{\theta=2}^t \rho_\theta(\eta^\theta) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^1} (\rho_1(\eta^1)(\eta_i^1)^{1-\alpha}) \\ \vdots \\ \prod_{\theta=1}^{t-1} \rho_\theta(\eta^\theta) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^t} (\rho_t(\eta^t)(\eta_i^t)^{1-\alpha}) \end{pmatrix} d\eta^{1t} \\ &= \int_{f^t(z^{1t}, u^{1t}, \eta^{1t}) \leq 0} \frac{1}{\alpha m} \begin{pmatrix} \rho_1^{-1}(\eta^1) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^1} (\rho_1(\eta^1)(\eta_i^1)^{1-\alpha}) \\ \vdots \\ \rho_t^{-1}(\eta^t) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^t} (\rho_t(\eta^t)(\eta_i^t)^{1-\alpha}) \end{pmatrix} \prod_{\theta=1}^t \rho_\theta(\eta^\theta) d\eta^{1t}. \end{aligned} \tag{40}$$

Denote

$$\begin{aligned} v^t(\eta^{1t}) &= \frac{1}{\alpha m} \begin{pmatrix} \rho_1^{-1}(\eta^1) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^1} (\rho_1(\eta^1)(\eta_i^1)^{1-\alpha}) \\ \vdots \\ \rho_t^{-1}(\eta^t) \sum_{i=1}^m \frac{\partial}{\partial \eta_i^t} (\rho_t(\eta^t)(\eta_i^t)^{1-\alpha}) \end{pmatrix} \\ &= \frac{1}{\alpha m} \begin{pmatrix} \sum_{i=1}^m ((1-\alpha)(\eta_i^1)^{-\alpha} + (\eta_i^1)^{1-\alpha} \frac{\partial}{\partial \eta_i^1} \ln \rho_1(\eta^1)) \\ \vdots \\ \sum_{i=1}^m ((1-\alpha)(\eta_i^t)^{-\alpha} + (\eta_i^t)^{1-\alpha} \frac{\partial}{\partial \eta_i^t} \ln \rho_t(\eta^t)) \end{pmatrix}. \end{aligned} \tag{41}$$

In view of (40)

$$\begin{aligned} \nabla_{u^i} \phi_i(u^i) &= \int_{f^i(z^i, u^i, \eta^i) \leq 0} v^i(\eta^i) \prod_{\theta=1}^i \rho_\theta(\eta^\theta) d\eta^{1^i} = \mathbb{E}_\zeta[v^i(\eta^i) I_{\{f^i(z^i, u^i, \eta^i) \leq 0\}}]. \end{aligned} \quad (42)$$

Analogously, by applying formula (6), we obtain

$$\nabla_{u^i r} \int_{f^{T+1}(z^i r, u^i r, \eta^i r) \leq 0} \prod_{\theta=1}^T \rho_\theta(\eta^\theta) d\eta^{1^T} = \mathbb{E}_\zeta[v^T(\eta^i r) I_{\{f^{T+1}(z^i r, u^i r, \eta^i r) \leq 0\}}]. \quad (43)$$

By (36), (37), and (42)

$$\begin{aligned} \mathbb{E}[g_i(z_i) \nabla_{u^i r} \phi_i(u^i r)] &= \mathbb{E} \left[g_i(z_i) \mathbb{E}_\zeta \left[\begin{pmatrix} v^i(\eta^i r) \\ 0 \end{pmatrix} I_{\{f^i(z^i r, u^i r, \eta^i r) \leq 0\}} \right] \right] \\ &= \mathbb{E} \left[g_i(z_i) \begin{pmatrix} v^i(\eta^i r) \\ 0 \end{pmatrix} I_{\{f^i(z^i r, u^i r, \eta^i r) \leq 0\}} \right]. \end{aligned} \quad (44)$$

Analogously, by (43)

$$\begin{aligned} &\mathbb{E} \left[g_{T+1}(z_T) \nabla_{u^i r} \int_{f^{T+1}(z^i r, u^i r, \eta^i r) \leq 0} \prod_{\theta=1}^T \rho_\theta(\eta^\theta) d\eta^{1^T} \right] \\ &= \mathbb{E} [g_{T+1}(z_T) \mathbb{E}_\zeta[v^T(\eta^i r) I_{\{f^{T+1}(z^i r, u^i r, \eta^i r) \leq 0\}}]] \\ &= \mathbb{E} [g_{T+1}(z_T) v^T(\eta^i r) I_{\{f^{T+1}(z^i r, u^i r, \eta^i r) \leq 0\}}]. \end{aligned} \quad (45)$$

Let us denote

$$d^i(\eta^{1^i}) = \begin{cases} \begin{pmatrix} v^i(\eta^{1^i}) \\ 0 \end{pmatrix}, & \text{for } 1 \leq i \leq T-1; \\ v^T(\eta^{1^T}), & \text{for } i = T, T+1. \end{cases} \quad (46)$$

Finally, with (35), (44), (45), and (46)

$$\nabla_{u^i r} G(u^{1^T}) = \sum_{i=1}^T \mathbb{E} [g_i(z_i) d^i(\eta^{1^i}) I_{\{f^i(z^i r, u^i r, \eta^i r) \leq 0\}}]$$

$$\begin{aligned}
& + \mathbb{E}[g_{T+1}(z_T) a^T(\eta^{1T}) I_{\{f^{T+1}(z^{1T}, u^{1T}, \eta^{1T}) \leq 0\}}] \\
& = \mathbb{E}[g_\tau(z_\tau) a^T(\eta^{1\tau})], \tag{47}
\end{aligned}$$

where $a^T(\eta^{1\tau})$ is given by equations (41) and (46).

Numerical implementation of the gradient formula

We show that gradient formula (47) can be easily calculated numerically. By Monte Carlo simulation, each run generates a random trajectory with equations (29) and (30) and generates a sample $g_\tau(z_\tau)$ of the criterion function $G(u^{1T}) = \mathbb{E} g_\tau(z_\tau)$. The estimate of the criterion function can be obtained with N simulation runs as

$$\tilde{G}(u^{1T}) = N^{-1} \sum_{\nu=1}^N g_{\tau_\nu}(z_{\tau_\nu}). \tag{48}$$

Similarly, an estimate of the gradient $\nabla_{u^{1T}} G(u^{1T}) = \mathbb{E}[g_\tau(z_\tau) a^T(\eta^{1\tau})]$ can be obtained during the same runs as

$$\tilde{\nabla}_{u^{1T}} G(u^{1T}) = N^{-1} \sum_{\nu=1}^N g_{\tau_\nu}(z_{\tau_\nu}) a^{\tau_\nu}(\eta^{1\tau_\nu}). \tag{49}$$

Let us explain how *all* the components of the vector $a^T(\eta^{1\tau})$ can be calculated with one simulation run of the model. The vector $a^T(\eta^{1\tau})$ is defined by formulas (41) and (46) and is a function of the vector $v^T(\eta^{1\tau})$. By (41), the component number j of the vector $v^T(\eta^{1\tau})$ equals

$$v_j^T(\eta^{1\tau}) = \frac{1}{\alpha m} \sum_{i=1}^m ((1 - \alpha)(\eta_i^j)^{-\alpha} + (\eta_i^j)^{1-\alpha} \frac{\partial}{\partial \eta_i^j} \ln \rho_j(\eta^j)). \tag{50}$$

The density function $\rho_j(\eta^j)$, in the special case considered, is the product of density functions $\rho_{ji}(\eta_i^j)$ for each variable η_i^j

$$\rho_j(\eta^j) = \prod_{i=1}^{r_j} \rho_{ji}(\eta_i^j),$$

and random values η_i^j are normally distributed

$$\rho_{ji}(\eta_i^j) = \frac{1}{\sqrt{2\pi\sigma_i^j}} \exp\left\{-\frac{\eta_i^j - m_i^j}{2(\sigma_i^j)^2}\right\}.$$

Therefore,

$$\begin{aligned}
 \frac{\partial}{\partial \eta_i^j} \ln \rho_j(\eta^j) &= \frac{\partial}{\partial \eta_i^j} \sum_{i=1}^{r_j} \ln \rho_{ji}(\eta_i^j) \\
 &= \frac{\partial}{\partial \eta_i^j} \ln \rho_{ji}(\eta_i^j) \\
 &= \frac{\partial}{\partial \eta_i^j} \left(-\frac{\eta_i^j - m_i^j}{2(\sigma_i^j)^2} \right) \\
 &= -2^{-1}(\sigma_i^j)^{-2}.
 \end{aligned} \tag{51}$$

Combining (41), (50), and (51),

$$v^r(\eta^{1r}) = \frac{1}{\alpha m} \begin{pmatrix} \sum_{i=1}^m ((1-\alpha)(\eta_i^1)^{-\alpha} - 2^{-1}(\sigma_i^1)^{-2}(\eta_i^1)^{1-\alpha}) \\ \vdots \\ \sum_{i=1}^m ((1-\alpha)(\eta_i^r)^{-\alpha} - 2^{-1}(\sigma_i^r)^{-2}(\eta_i^r)^{1-\alpha}) \end{pmatrix}. \tag{52}$$

The simulation run generates random vector $\eta^{1r} = (\eta_1^r, \dots, \eta_r^r)$, therefore, vector $a^r(\eta^{1r})$ can be easily calculated with formulas (46) and (52).

Actually, the estimate of the gradient (49) with respect to *all* variables u_1, \dots, u_T is available “free of charge” since it involves far fewer calculations than generating N sample paths.

Appendix: Proof of theorem 2.1

Here, we do not prove all statements strictly. Our aim is only to demonstrate that the differentiation formula for the probability functions can be obtained relatively easily.

Let $x \in \mathbb{R}^n$ and

$$F(x) = \int_{f(x,y) \leq 0} p(x,y) dy = \int_{\mu(x)} p(x,y) dy.$$

We increment argument x with the vector Δx . The difference $F(x + \Delta x) - F(x)$ can

be presented as

$$\begin{aligned}
 F(x + \Delta x) - F(x) &= \int_{\mu(x+\Delta x)}^{p(x + \Delta x, y)} p(x + \Delta x, y) dy - \int_{\mu(x)}^{p(x, y)} p(x, y) dy \\
 &= \int_{\mu(x+\Delta x)}^{p(x + \Delta x, y)} (p(x + \Delta x, y) - p(x, y)) dy \\
 &\quad + \int_{\mu(x+\Delta x)}^{p(x, y)} p(x, y) dy - \int_{\mu(x)}^{p(x, y)} p(x, y) dy.
 \end{aligned}
 \tag{53}$$

By the Taylor theorem

$$\int_{\mu(x+\Delta x)}^{p(x + \Delta x, y)} (p(x + \Delta x, y) - p(x, y)) dy \approx \int_{\mu(x+\Delta x)}^{p(x, y)} \langle \nabla_x p(x, y), \Delta x \rangle dy.
 \tag{54}$$

Conditions 2, 7, and 8 of theorem 2.1 imply (this statement is strictly proven in [22])

$$\int_{\mu(x+\Delta x)} \langle \nabla_x p(x, y), \Delta x \rangle dy \approx \left\langle \int_{\mu(x)} \nabla_x p(x, y) dy, \Delta x \right\rangle.
 \tag{55}$$

Denote (see figure 1)

$$\bar{\mu}(x, i, \Delta x) = \{y \in \mathbb{R}^m : f_{i,i-1}(x, y) \leq 0; f_{i,k}(x + \Delta x, y) \leq 0\}, \quad i = 2, \dots, k;$$

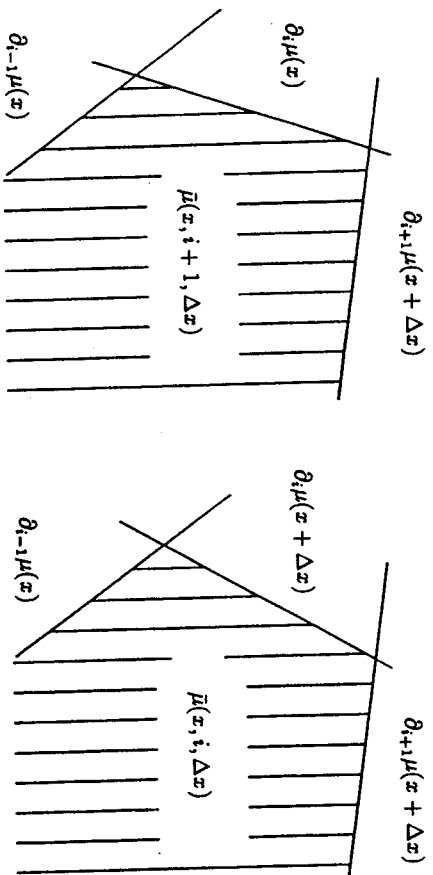


Figure 1. The sets $\bar{\mu}(x, i, \Delta x)$ and $\bar{\mu}(x, i + 1, \Delta x)$.

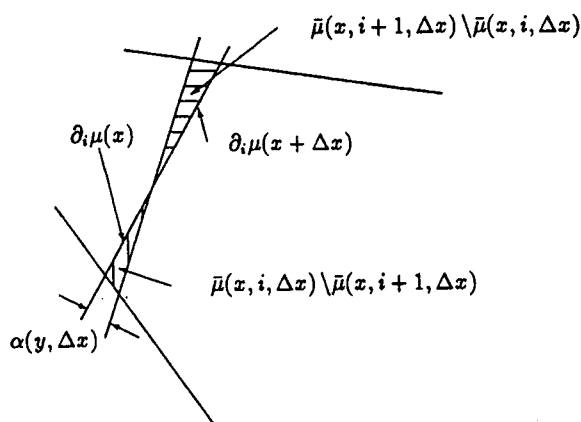


Figure 2. The sets $\bar{\mu}(x, i, \Delta x) \setminus \bar{\mu}(x, i+1, \Delta x)$ and $\bar{\mu}(x, i+1, \Delta x) \setminus \bar{\mu}(x, i, \Delta x)$.

$$\bar{\mu}(x, 1, \Delta x) = \mu(x + \Delta x),$$

$$\bar{\mu}(x, k+1, \Delta x) = \mu(x).$$

With these definitions

$$\begin{aligned} & \int_{\mu(x+\Delta x)} p(x, y) dy - \int_{\mu(x)} p(x, y) dy \\ &= \sum_{i=1}^k \left[\int_{\bar{\mu}(x, i, \Delta x)} p(x, y) dy - \int_{\bar{\mu}(x, i+1, \Delta x)} p(x, y) dy \right]. \end{aligned} \quad (56)$$

The difference

$$\begin{aligned} & \int_{\bar{\mu}(x, i, \Delta x)} p(x, z) dz - \int_{\bar{\mu}(x, i+1, \Delta x)} p(x, z) dz \\ &= \int_{\bar{\mu}(x, i, \Delta x) \setminus \bar{\mu}(x, i+1, \Delta x)} p(x, y) dy - \int_{\bar{\mu}(x, i+1, \Delta x) \setminus \bar{\mu}(x, i, \Delta x)} p(x, y) dy \stackrel{\text{def}}{=} U_i \end{aligned} \quad (57)$$

can be represented as a surface integral. Denote by $\alpha_i(y, \Delta x)$ the thickness of the layer $\bar{\mu}(x, i, \Delta x) \setminus \bar{\mu}(x, i+1, \Delta x)$ (see figure 2). This thickness $\alpha_i(y, \Delta x)$ can be found from the equation

$$f_i \left(x + \Delta x, y + \alpha_i(y, \Delta x) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \right) = f_i(x, y).$$

The Taylor expansion theorem and condition 7 of theorem 2.1 imply

$$f_i(x, y) + \nabla_x^T f_i(x, y) \Delta x + \alpha_i(y, \Delta x) \nabla_y^T f_i(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \approx f_i(x, y).$$

Therefore

$$\alpha_i(y, \Delta x) \approx \frac{\nabla_x^T f_i(x, y) \Delta x}{\|\nabla_y f_i(x, y)\|}.$$

Analogously, the thickness of the layer $\bar{\mu}(x, i+1, \Delta x) \setminus \bar{\mu}(x, i, \Delta x)$ approximately equals $-\alpha_i(y, \Delta x)$. The integrals over the layers

$$\bar{\mu}(x, i, \Delta x) \setminus \bar{\mu}(x, i+1, \Delta x) \quad \text{and} \quad \bar{\mu}(x, i+1, \Delta x) \setminus \bar{\mu}(x, i, \Delta x)$$

can be presented approximately as an integral over surface

$$\begin{aligned} U_i &\approx \int_{\partial_i \mu(x)} \alpha_i(y, \Delta x) p(x, y) \, dS = \int_{\partial_i \mu(x)} -\frac{\nabla_x^T f_i(x, y) \Delta x}{\|\nabla_y f_i(x, y)\|} p(x, y) \, dS \\ &= \int_{\partial_i \mu(x)} -\frac{\langle \nabla_x f_i(x, y), \Delta x \rangle}{\|\nabla_y f_i(x, y)\|} p(x, y) \, dS \\ &= \left\langle - \int_{\partial_i \mu(x)} \frac{\nabla_x f_i(x, y)}{\|\nabla_y f_i(x, y)\|} p(x, y) \, dS, \Delta x \right\rangle. \end{aligned}$$

Therefore (see (56) and (57)),

$$\begin{aligned} \int_{\mu(x+\Delta x)} p(x, y) \, dy - \int_{\mu(x)} p(x, y) \, dy &= \sum_{i=1}^k U_i \\ &\approx \left\langle - \sum_{i=1}^k \int_{\partial_i \mu(x)} \frac{\nabla_x f_i(x, y)}{\|\nabla_y f_i(x, y)\|} p(x, y) \, dS, \Delta x \right\rangle. \end{aligned} \quad (58)$$

Combining (53), (54), (55), and (58), we obtain

$$\nabla_x F(x) = \int_{\mu(x)} \nabla_x p(x, y) \, dy - \sum_{i=1}^k \int_{\partial_i \mu(x)} \frac{p(x, y)}{\|\nabla_y f_i(x, y)\|} \nabla_x f_i(x, y) \, dS. \quad (59)$$

Thus, formula (5) is valid.

The Ostrogradski–Gauss theorem links the integral over a volume and the integral over a surface (see definition of the matrix $H_l(x, y)$ in conditions 5 and 6 of theorem 2.1):

$$\begin{aligned} \int_{\mu(x)} \operatorname{div}_y(p(x, y)H_l(x, y)) \, dy &= \sum_{i=1}^k \int_{\partial_i\mu(x)} p(x, y)H_l(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS \\ &= \sum_{i=1}^l \int_{\partial_i\mu(x)} p(x, y)H_l(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS \\ &\quad + \sum_{i=l+1}^k \int_{\partial_i\mu(x)} p(x, y)H_l(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS. \end{aligned} \quad (60)$$

Since the matrix $H_l(x, y)$ satisfies equation (3), then

$$\begin{aligned} \sum_{i=1}^l \int_{\partial_i\mu(x)} p(x, y)H_l(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS \\ = - \sum_{i=1}^l \int_{\partial_i\mu(x)} p(x, y) \frac{\nabla_x f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS. \end{aligned} \quad (61)$$

With (60), and (61)

$$\begin{aligned} - \sum_{i=1}^l \int_{\partial_i\mu(x)} p(x, y) \frac{\nabla_x f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS &= \int_{\mu(x)} \operatorname{div}_y(p(x, y)H_l(x, y)) \, dy \\ &\quad - \sum_{i=l+1}^k \int_{\partial_i\mu(x)} p(x, y)H_l(x, y) \frac{\nabla_y f_i(x, y)}{\|\nabla_y f_i(x, y)\|} \, dS. \end{aligned} \quad (62)$$

Combining (59) and (62) we obtain (4). This concludes the proof. \square

References

- [1] P. Bjerager, *Methods for Structural Reliability Computations*, Course on Reliability Problems: General Principles and Applications in Mechanics of Solids and Structures, International Center for Mechanical Sciences, Udine, Italy, CISM Lecture Notes (Springer, 1991).
- [2] Yu. Ermoliev, V. Norkin and R. Wets, The minimization of discontinuous functions: mollifier subgradients, Working Paper WP-92-73, International Institute for Applied Systems Analysis, Laxenburg, Austria (1992).

- [3] Yu. Ermoliev, Stochastic quasi-gradient methods and their applications to system optimization, *Stochastics* 4 (1983) 1–36.
- [4] A.A. Gaivoronski, L.Y. Shi and R.S. Sreenivas, Augmented infinitesimal perturbation analysis: an alternative explanation, *Discr. Event Dyn. Syst.* 2 (1992) 121–138.
- [5] P. Glasserman, *Gradient Estimation Via Perturbation Analysis* (Kluwer, Boston, 1991).
- [6] D.P. Graver and P.K. Samanta, Detecting component failure potential using proportional hazard model, US Nuclear Regulatory Commission, NUREG/CR-5324, BNL-NUREG-52188 (1989).
- [7] Y.C. Ho and X.R. Cao, *Perturbation Analysis of Discrete Event Dynamic Systems* (Kluwer, Boston, 1991).
- [8] K. Marti, Stochastic optimization methods in structural design, *ZAMM* 4 (1990) T742–T745.
- [9] G.Ch. Pflug, *Simulation and Optimization – the Interface* (Kluwer, 1995), in press.
- [10] A. Prékopa, On probabilistic constrained programming, in: *Proc. Princeton Symp. on Mathematical Programming* (Princeton University Press, Princeton, NJ, 1970) pp. 113–138.
- [11] A. Prékopa and T. Szántai, A new multivariate gamma distribution and its fitting to empirical streamflow data, *Water Resources Res.* 14 (1978) 19–24.
- [12] A. Pulkkinen and S. Uryasev, Optimal operational strategies for an inspected component – statement of the problem, Working Paper WP-90-62, International Institute for Applied Systems Analysis, Laxenburg, Austria (1990).
- [13] A. Pulkkinen and S. Uryas'ev, Optimal operational strategies for an inspected component – solution techniques, Collaborative Paper CP-91-13, International Institute for Applied Systems Analysis, Laxenburg, Austria (1991).
- [14] E. Raik, The differentiability in the parameter of the probability function and optimization of the probability function via the stochastic pseudogradient method, *Izv. Akad. Nayk Est. SSR, Phis. Math.* 24 (1975) 3–6 (in Russian).
- [15] N. Roenko, Stochastic programming problems with integral functionals over multivalued mappings, Synopsis of Ph.D. Thesis, Kiev, Ukraine (in Russian) (1983).
- [16] R. Rubinstein, Sensitivity analysis of discrete event systems by the “push out” method, *Ann. Oper. Res.* 39 (1992) 229–250.
- [17] R. Rubinstein and A. Shapiro, *Discrete Event Systems: Sensitivity Analysis and Stochastic Optimization via the Score Function Method* (Wiley, Chichester, 1993).
- [18] P.K. Samanta, W.E. Vesely and I.S. Kim, Study of operational risk-based configuration control, US Nuclear Regulatory Commission, NUREG/CR-5641, BNL-NUREG-52261 (1991).
- [19] J. Simon, Second variation in domain optimization problems, in: *Int. Series of Numerical Mathematics*, vol. 91, eds. F. Kappel, K. Kunish and W. Schappacher (Birkhäuser, 1985) pp. 361–378.
- [20] S. Uryas'ev, Differentiability of an integral over a set defined by inclusion, *Kibernetika (Kiev)* 5 (1988) 83–86 (in Russian) [Transl.: *Cybernetics* 24 (1988) 638–642].
- [21] S. Uryas'ev, A differentiation formula for integrals over sets given by inclusion, *Numer. Funct. Anal. Optim.* 10 (1989) 827–841.
- [22] S. Uryas'ev, Derivatives of probability functions and integrals over sets given by inequalities, *J. Comp. Appl. Math.* 56 (1995).
- [23] W.E. Vesely, Approaches for age-dependent probabilistic safety assessments with emphasis on prioritization and sensitivity studies, US Nuclear Regulatory Commission, NUREG/CR-5587, SAIC-92/1137 (1992).