Working Paper

ON OPTIMIZATION OF DYNAMICAL MATERIAL FLOW SYSTEMS USING SIMULATION

Yu. Ermoliev, S. Uryas'ev and J. Wessels

WP-92-76
October 1992

ON OPTIMIZATION OF DYNAMICAL MATERIAL FLOW SYSTEMS USING SIMULATION

Yu. Ermoliev, S. Uryas'ev and J. Wessels

WP-92-76
October 1992

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria

Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Summary

Up until now risk analysis, as a rule, ended with the estimation of the risks. Further improvements – optimal design, risk control, dynamic risk management – require many more efforts. Essential difficulties are connected with the discontinuous or nonsmooth behavior of performance functions with respect to the control and (or) random parameters due to possible failures of the system's parts. Usually, the systems also include discrete event elements - logical rules can change the structure of the system if some constraints are not satisfied, for example safety constraints. These problems require new formal analysis tools which will include dynamics, stochastics, nonsmoothness and discontinuity. In this paper, the authors consider a simple example of such a problem with the aim to explore the possibilities for its analysis. The problem is comprised of optimizing a material flow system based on an efficient use of simulation. The material flow system may be a production system, a distribution system or a pollutant-deposit/removal system. The important characteristic which is considered in this paper is that one of the components of the dynamic system is unreliable. This characteristic leads to simulation models in which criteria are discontinuous with respect to the optimization parameters. This makes it difficult to use the standard methods for the estimation of gradients of the expected criteria values. A method is introduced which overcomes the difficulty. From a formal point of view the problem can be viewed as a mixed integer stochastic optimization problem.

Contents

ON OPTIMIZATION OF DYNAMICAL MATERIAL FLOW SYSTEMS USING SIMULATION

Yu. Ermoliev, S. Uryas'ev and J. Wessels

1 Introduction

In several types of material flow systems there is at least one component which is unreliable. This feature makes it particularly important to design such a system carefully by taking into account the uncertainties introduced by the component unreliability. Such material flow situations occur in production and distribution systems, but also in environmental systems for the temporary deposit and removal or transformation of pollutants. In production systems one of the work-stations may be unreliable. In distribution systems one of the transport mechanisms may suffer from breakdowns. In environmental systems it might be the removal or transformation mechanism which is not always operational. Particularly for production systems, there is quite an extensive choice of literature on the modeling and analysis of material flows with unreliable work-stations.

Analytic approaches have primarily been developed for the case of two work-stations with an intermediate buffer. For the situation of discrete products with deterministic processing times we may refer to Buzacott [1] and Yeralan and Muth [27]. For the situation of continuous material flows with deterministic machine speeds important references are Wijngaard [26] and Mitra [14]. De Koster [11] gives a good overview of the literature and shows how to exploit Wijngaard's approach for the construction of a numerical procedure for the analysis of larger systems. Although these analytic approaches are very valuable for getting a better understanding of the characteristics of the relevant processes, they all suffer from the fact that they are based on severe assumptions. The usual requirement is that breakdown behavior as well as repair behavior is based on a negative-exponentially distributed time length or at least something very closely related to the negative-exponential distribution like a phase-type distribution with only very few phases. Therefore, for practical system design, simulation is the most important tool.

1

However, a serious drawback of simulation is that a guided search for a good design usually requires many simulations. Particularly, in the case of several design parameters, this can be prohibiting.

A general approach for the optimization of stochastic systems by using Monte Carlo simulations is based on the technique of stochastic optimization [2, 3, 12]. The technique relies on estimating the gradients (or their generalizations for nonsmooth functions) of a criterion and/or constraint functions. These estimates are called stochastic gradients, subgradients or quasigradients depending on whether the function is smooth or nonsmooth and whether we use unbiased or biased estimates. A lot of research effort has been concentrated on obtaining estimates which require only one observation of the random parameters - so called one–simulation–run estimates. The existing theory allows to calculate such estimates for various classes of applied problems, including problems involving stochastic inequalities or stochastic differential equations (ordinary or partial). Relatively recently a one–run gradient estimation technique was proposed for queuing type problems by Ho et al. in [9] under the name of perturbation analysis. For a good overview of its properties and possibilities, see Suri [20]. Subsequently, sensitivity analysis for discrete event processes with control parameters in the densities of the random values was developed by Pflug [15], Rubinstein and Shapiro [17] and others. A disadvantage of the standard form of perturbation analysis is that, generally speaking, it does not work when the sample path is discontinuous in the relevant parameter. In that case it may no longer be allowed to change the order of taking the expectation with taking the derivative:

$$\nabla_x \mathbb{E}\, g(x, \omega) \;\neq\; \mathbb{E}\, \nabla_x\, g(x, \omega),$$

where x indicates the decision vector and ω the random element, $g(x, \omega)$ is a sample performance function. In [10], Ho and Li present an approach for circumventing this difficulty by using finite difference approximation ideas. However, their method requires that time periods between events are negative-exponentially distributed and their method requires very long simulation runs for obtaining good estimates. Gaivoronski and Ermoliev [4] suggested to use artificial random variables to smooth the function $g(x, \omega)$ and combine it with concurrent approximation and optimization techniques. The general framework for the development of concurrent approximation and optimization techniques is studied by Ermoliev, Norkin, and Wets [5]. In [7], Gong and Ho suggested to smooth over the function $g(x, \omega)$ by taking conditional expectation with respect to a σ-algebra $\mathcal{F}$

2

$$\mathbb{E}\,g(x,\omega) \;=\; \mathbb{E}\mathbb{E}_{\mathcal{F}}\,g(x,\omega) \;=\; \mathbb{E}\tilde{g}(x,\omega)\,,$$

where the smoothed performance function $\tilde{g}(x,\omega)$ is supposed to be explicitly defined to calculate the gradient. As a rule the explicit formulae for these conditional expectations are not available. However in various control problems the function $\tilde{g}(x,\omega)$ can often be represented through multidimensional integrals over sets given by linear or nonlinear inequalities (with respect to decision and random variables). Although such integrals can not be presented as explicit functions of decision variables, it is possible to find implicit expressions of their gradients. We show that in an important case, when densities of the random variables and their derivatives are known, the gradient can be presented as an integral over a volume. We finally also show that the obtained expression of the gradient allows to estimate the gradient by one simulation run and that the stochastic quasi-gradient optimization procedures can be used for a directed simulation search.

In Section 2 the model and the mixed integer stochastic optimization problem are introduced. We briefly mention in this section the optimization technique to be applied. Sections 3 and 4 show how estimates for the derivatives of the performance function can be obtained. Section 5 treats the optimization search techniques. The formulas for derivatives of integrals over sets given by inequalities are given in the Appendix.

2   The Model Description and Statement of the Optimization Problem

2.1   The Model

In this section we introduce a very simple material flow problem, which is used for demonstration of our approach. There are two interacting processes (see Figure 1): a regular flow of material arriving at a "server" or "work station" and a service process of this material. Each batch of material arrives at the server at equidistant points in time $t = 0,\ x_1,\ 2x_1, \ldots$ The intensity of this process can be adjusted by the value $x_1 > 0$. The work-station empties the available batches one-by-one and x_2 is the time needed to process a batch by the work-station. The processing of a batch can be interrupted by the failure of the work-station, therefore, the work station has alternatively a period in which it is "operational" and in which it is in "repair". The lengths of these periods are independent random
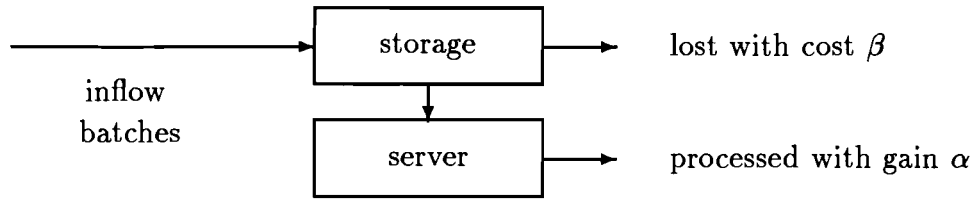
3

```
                              ┌──────────┐
  ───────────────────────────▶│ storage  │───────▶  lost with cost β
        inflow                └──────────┘
        batches                     │
                              ┌──────────┐
                              │  server  │───────▶  processed with gain α
                              └──────────┘
```

Figure 1: *System Flow Chart*

variables with density functions ν and ϱ respectively. We suppose that a batch requires a position in the storage (buffer) from the time of its arrival until the moment that processing has been finished. Denote by S the number of batches that may be stocked in the storage. If the storage is full, we suppose that a newly arriving batch is lost with cost β. The gain of each processed batch is equal to α. The cost function includes, on the one hand, the investments and maintenance costs as functions of the design of parameters x_1, x_2, S. On the other hand, the cost function provides a tradeoff between risk of profits and risk of losses. In particular, in the case when $\alpha \ll \beta$ main attention is paid to the losses due to exceeding storage capacity.

As a consequence of the interruptions, the real processing time of a batch may be essentially longer than x_2, since the processing might be interrupted several times. If T_1 and T_2 are the expected values for the life time (between two successive failures) and the repair time respectively, then the availability fraction is

$$\frac{T_1}{T_2 + T_1}$$

Consequently, the real processing time will be on the average

$$\frac{T_2 + T_1}{T_1} x_2 .$$

If x_1 would be chosen smaller than this latter value, then the work-station could not cope with the input even if the storage capacity would have infinite size.

The model as sketched above was inspired by the problem of designing a production system in which the batches would be delivered by a chain oven and the work-station treats the individual products of a batch one-by-one. In this way, the platter, which bears a batch of products through the oven, occupies a position in the buffer as long as it

4

contains some products. It is not possible to stop the oven when the buffer is full, since this would lead to the loss of several hours of production, namely all the batches which are in the oven would be lost in this case. The chain pulls the batches through the oven with a fixed speed and batches which leave the oven are mechanically delivered to the buffer. If the buffer has no position available, then the batch is set aside and lost for further processing, since outside the buffer the products cool down too much, which is not good for the quality.

In fact, the situation described above is an example of the two-machine system with one unreliable machine and an intermediate buffer. All assumptions have been chosen as simple as possible in order to keep the discussion of the essential aspects clear.

A similar model might be reasonable for a pollutant removal/transformation situation, where batches of the pollutant are delivered by an uncontrollable process, whereas the transformation or removal only works under certain conditions. We will stick to the first interpretation for the introduction of a design criterion.

For the production problem the criterion is to design a system which works at the lowest possible costs. There is no constraint on the desired output, since it is already certain that several units will be needed, therefore the only goal is to find the most efficient design. The first components of the costs per time unit are caused by the investments and maintenance costs. We suppose that they are known functions of the design parameter for each of the parts of the system: $c_1(x_1) + c_2(x_2) + c_3(S)$. The most complicated cost/gain component is delivered by the performance of the system: processed batches bring a gain α and lost batches a cost β. So the criterion becomes

$$C(x, S) = c_1(x_1) + c_2(x_2) + c_3(S) + \alpha L(x, S) - \beta \Psi(x, S), \tag{1}$$

where $x = (x_1, x_2)$ and $L(x, S)$, $\Psi(x, S)$ are the expected numbers of lost and processed batches per time unit, respectively. We could minimize function (1) with respect to (x, S), by using the conventional optimization techniques, if we would know $L(x, S)$ and $\Psi(x, S)$. Unfortunately, it is impossible to find analytical expressions for $L(x, S)$ and $\Psi(x, S)$. With fixed (x, S), a simulation run could provide only estimates for $\Psi(x, S)$ and $L(x, S)$. However, we will show that one simulation run may also provide estimates for the gradient of function (1).

Let us make some useful rearrangements of the problem. We assume that all random variables are specified on the probability space $(P, \mathcal{F}, \Omega)$. Let us suppose that we simulate

5

the process for a given setting (x, S) of the design parameters until N batches are delivered (and lost or processed). Denote by $L_N(x, S, \omega)$ the random number of lost batches and by $\Psi_N(x, S, \omega)$ the random number of processed batches during the simulation run, while $\omega \in \Omega$ denotes the random element in the process. We have

$$L_N(x, S, \omega) + \Psi_N(x, S, \omega) = N ,$$

$$\lim_{N \to \infty} \frac{L_N(x, S, \omega)}{N x_1} = L(x, S), \quad \lim_{N \to \infty} \frac{\Psi_N(x, S, \omega)}{N x_1} = \Psi(x, S) \quad \text{(a.s.)}$$

Therefore, a sensible estimate of $\alpha L(x, S) - \beta \Psi(x, S)$, based on the simulation run, would be

$$\alpha \frac{L_N(x, S, \omega)}{N x_1} - \beta \frac{\Psi_N(x, S, \omega)}{N x_1} = \alpha \frac{L_N(x, S, \omega)}{N x_1} - \beta \frac{N - L_N(x, S, \omega)}{N x_1} =$$

$$(\alpha + \beta) \frac{L_N(x, S, \omega)}{N x_1} - \frac{\beta}{x_1} = \frac{\alpha + \beta}{x_1} \frac{L_N(x, S, \omega)}{L_N(x, S, \omega) + \Psi_N(x, S, \omega)} - \frac{\beta}{x_1} =$$

$$\frac{\alpha + \beta}{x_1} \frac{L_N(x, S, \omega) \Psi_N^{-1}(x, S, \omega)}{L_N(x, S, \omega) \Psi_N^{-1}(x, S, \omega) + 1} - \frac{\beta}{x_1} .$$

Let us denote

$$F(x, S) = \lim_{N \to \infty} L_N(x, S, \omega) \Psi_N^{-1}(x, S, \omega) \quad \text{(a.s.)},$$

then

$$\alpha L(x, S) - \beta \Psi(x, S) = \frac{\alpha + \beta}{x_1} \frac{F(x, S)}{F(x, S) + 1} - \frac{\beta}{x_1} .$$

Thus the criterion function is equal to

$$C(x, S) = c_1(x_1) + c_2(x_2) + c_3(S) + \frac{\alpha + \beta}{x_1} \frac{F(x, S)}{F(x, S) + 1} - \frac{\beta}{x_1} , \tag{2}$$

2.2 Optimization Problem

Usually feasible storage sizes $S \in \{S_1, \ldots, S_I\}$ are known a priori. The problem is to find values $x_1 \geq 0$, $x_2 \geq 0$ and S_i minimizing the function $C(x, S_i)$. It is easy to see that this problem is equivalent to the minimization problem

$$\Phi(x, y) \overset{\text{def}}{=} \sum_{i=1}^{I} C(x, S_i) y_i \rightarrow \min_{(x, y) \in \mathbb{R}^2 \times \mathbb{R}^I} \tag{3}$$

subject to constraints

$$\sum_{i=1}^{I} y_i = 1 ; \quad y_i \geq 0, \; i = 1, \ldots, I ; \tag{4}$$

6

$$x_1 \geq x_2 \,, \; x_2 \geq 0 \,. \tag{5}$$

Denote

$$X \;=\; \left\{\, x \in \mathbb{R}^2 \,:\, x_1 \geq x_2 \,, \; x_2 \geq 0 \,\right\} \,,$$

$$Y \;=\; \left\{\, y \in \mathbb{R}^I \,:\, \sum_{i=1}^{I} y_i \;=\; 1 \,, \; y_i \geq 0 \,, \; i = 1, \ldots, I \,\right\} \,,$$

then problem (3) can be formulated as

$$\Phi(x, y) \;\longrightarrow\; \min_{(x,y) \in X \times Y} \,. \tag{6}$$

This is a typical stochastic optimization problem with linear constraints. Despite the fact that the original problem is a mixed discrete-continuous stochastic optimization problem, we reduced it to the conventional problem with continuous variables by using the additional variables $y_1, \ldots, y_I$. As we see further, the function $\Phi(x, y)$ involves the calculation of multidimensional integrals, the exact evaluation of which is out of the question.

2.3 On Optimization Techniques

The function $\Psi(x, y)$ is smooth with respect to the variables x, y. We will apply stochastic quasigradient techniques [2], [3] (see Section 5) to optimize the function $\Phi(x, y)$ with respect to the variables x, y. The value of the criterion function $\Phi(x, y)$ cannot be calculated analytically since an analytical description of the function $F(x, S)$ is not available. Therefore, we simulate the process to estimate the values of the function $F(x, S_i)$, $i = 1, \ldots, I$. Because of technical (mathematical) reasons it is convenient (but not necessary) to change to the situation that the number of processed batches Ψ is fixed (rather then N) and the total number of processed batches $N_\Psi(x, S, \omega)$ and the number of lost batches $L_\Psi(x, S, \omega)$ are random functions of the variable Ψ and the control variables x, S. Let us denote by ψ the batch number processed by the work-station and by $\zeta_\psi(x, S, \omega)$ the number of lost batches during processing of the batch with number ψ. Since $L_\Psi(x, S, \omega)$ is the number of lost batches during the processing of Ψ batches by the work-station, we have

$$L_\Psi(x, S, \omega) \;=\; \sum_{\psi=1}^{\Psi} \zeta_\psi(x, S, \omega) \,. \tag{7}$$

As an approximation of $F(x, S)$ we consider the function

$$F_\Psi(x, S) \;=\; \Psi^{-1} \, \mathbb{E} \, L_\Psi(x, S, \omega) \,. \tag{8}$$

Partial derivatives $\frac{\partial}{\partial x_1}\Phi(x,y)$ and $\frac{\partial}{\partial x_2}\Phi(x,y)$ can be estimated through the derivatives $\frac{\partial}{\partial x_1}F_\Psi(x,S)$ and $\frac{\partial}{\partial x_2}F_\Psi(x,S)$. In the next two section, we develop unbiased estimates for the derivatives $\frac{\partial}{\partial x_1}F_\Psi(x,S)$ and $\frac{\partial}{\partial x_2}F_\Psi(x,S)$. The problem with the calculation of these derivatives is that the analytical description of the function $F_\Psi(x,S)$ is not available and the function $L_\Psi(x,S,\omega)$ is generated algorithmically. Moreover, the function $L_\Psi(x,S,\omega)$ is discontinuous with respect to x_1, x_2. For the discontinuous function $L_\Psi(x,S,\omega)$ (in spite of almost everywhere differentiability with respect to x_1, x_2) the interchange of taking the gradient and the mathematical expectation is not allowed. We differentiate the function $F_\Psi(x,S)$ with respect to x_1, x_2 by the direct formulae for derivatives of the integrals over the sets given by inequalities (see Appendix). Although the exact calculation of these derivatives is practically impossible we estimate these derivatives during one simulation run of the model. For these estimates we only need the following information: the sequence of busy periods of the work-station, the number of the batches which were supplied and processed from the beginning of the current busy period, operational and repair intervals. During one simulation run we estimate derivatives of the function $\Phi(x,y)$ at some point (x^s,y^s), then improve this point and get the next approximation (x^{s+1},y^{s+1}) and so on $s = 1,2,\ldots$ This procedure will further be described in Section 5.

3   The Derivative with Respect to the First Variable

In the present section and the next, we differentiate the mathematical expectation of the function $L_\Psi(x,S,\omega)$ with respect to variables x_1, and x_2 respectively. Since we do not variate variable S to simplify formulas we omit in this and the next section argument S. Further we write the random variables without the argument $\omega \in \Omega$, i.e. the random variable $\zeta(\omega)$ is denoted simply by ζ; we use bold face style for random variables. Thus the function $L_\Psi(x,S,\omega)$ is denoted as $\mathbf{L}_\Psi(x)$.

3.1   The Conditional Expectation of the number of Lost Batches

The mathematical expectation of the function $\mathbf{L}_\Psi(x)$ is a highly dimensional integral. To smooth the function $\mathbf{L}_\Psi(x)$ it is sufficient to integrate it with respect to some random variables, which possess probability densities. This means that we can calculate some conditional expectation of the function $\mathbf{L}_\Psi(x)$ and afterwards interchange the gradient and the mathematical expectation sign.
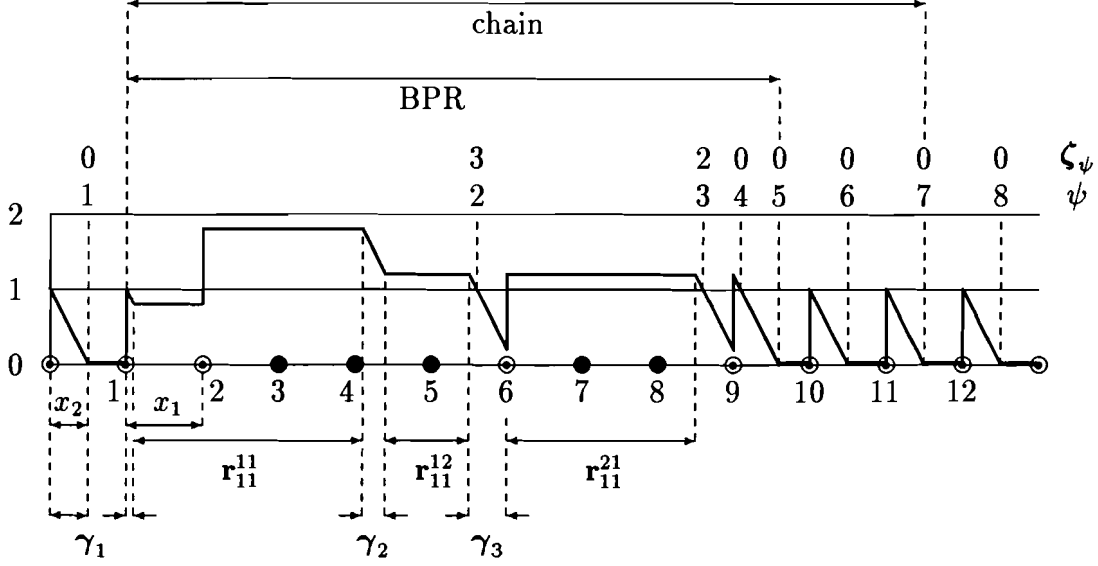
8

Figure 2: *Amount of material in the buffer. Buffer size equals to 2. Batches with numbers 3, 4, 5, 7, 8 are lost.*

The size S of the buffer is fixed. Denote by d the number of arriving batches such that after processing of these arrivals without repair the work-station is definitely idle

$$d = \min\{k \in \mathcal{N} : k \geq \frac{Sx_2}{x_1 - x_2}\} ,$$

where $\mathcal{N}$ is the set of natural numbers.

Let us split the sequence of batches $\{\psi\}$ processed by the work-station in "independent" chains such that each chain begins with a batch with repair and is finished at least by $S + d + 1$ batches without repairs (we exclude from consideration the batches from the beginning of the process up to the first batch with repairs). We might use the value $S + d$ for splitting the process in independent chains, but here we take $S + d + 1$ to guarantee that "small" changes of x do not influence the independence of the chains. Further, for sensitivity analysis with respect to x, we assume that the value d does not depend on x.

Let us numerate these chains with the index a. Denote the set of batches, belonging to the chain with the number a by $\delta(x_2, a)$. Further let us specify in each chain busy periods with repairs (subchains of batches between two idle periods of the work-station, at least for one batch in such a subchain the processing should be interrupted by repair). Let us call such a subchain (see Figure 2) Busy Period with Repair (BPR).

We numerate the sequence of BPR's by the index q, $q = 1, \ldots, Q(a)$ and numerate by the index θ, $\theta = 1, \ldots, \Theta(a, q)$ the batches which were processed with repairs in BPR

number q. Finally, let us numerate by the index j, $j = 1, \ldots, \mathbf{J}(a, q, \theta)$ the repairs during the processing of batch θ. Denote by $\mathbf{u}_{a,q}^{\theta}$ the number of batches which were processed by the work-station from the beginning of BPR (a, q) to batch θ (including batch θ) and denote by $\mathbf{v}_{a,q}^{\theta}$ the number of batches which were supplied to the buffer (some of them may be lost) from the beginning of BPR (a, q) to the finishing of the processing of batch θ. Each repair period can be identified by four indices a, q, θ, j, i.e. $\mathbf{r}_{a,q}^{\theta j}$ is a repair number j which happens during the processing of batch θ in BPR q from the chain a.

We suppose that lengths of repair periods are statistically independent and have a smooth probability density function ϱ.

Let us denote:

ℓ is the number of the operational period (this is a period in which the work-station is available);

γ_{ℓ} is the length of the operational period ℓ;

$\mathcal{F}_{\gamma}$ is the σ-algebra generated by the random lengths of the operational periods γ_{ℓ}, $\ell = 1, 2, \ldots$;

$\mathbb{E}_{\gamma}$ is the conditional expectation with respect to the σ-algebra $\mathcal{F}_{\gamma}$;

$\mathbb{P}_{\gamma}$ is the conditional probability with respect to the σ-algebra $\mathcal{F}_{\gamma}$;

$\mathbf{A}(x_2)$ is the number of chains for Ψ consumed batches.

The expectation of the function $\mathbf{L}_{\Psi}(x)$ can be represented as

$$
\mathbb{E}\,\mathbf{L}_{\Psi}(x) = \mathbb{E} \sum_{\psi=1}^{\Psi} \zeta_{\psi}(x) = \mathbb{E} \sum_{a=1}^{\mathbf{A}(x_2)} \sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_{\psi}(x)
$$

$$
= \mathbb{E}\,\mathbb{E}_{\gamma} \sum_{a=1}^{\mathbf{A}(x_2)} \sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_{\psi}(x) = \mathbb{E} \sum_{a=1}^{\mathbf{A}(x_2)} \mathbb{E}_{\gamma} \sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_{\psi}(x) . \tag{9}
$$

Let us denote the set of random values $\mathbf{u}_{a,q}^{\theta}$, $\mathbf{v}_{a,q}^{\theta}$, $q = 1, \ldots, \mathbf{Q}(a)$, $\theta = 1, \ldots, \Theta(a, q)$ by $\mathbf{U}_a(x)$. This random set $\mathbf{U}_a(x)$ can attain different values U_a, where the set U_a is the set of $u_{a,q}^{\theta}$, $v_{a,q}^{\theta}$, $q = 1, \ldots, Q(a)$, $\theta = 1, \ldots, \Theta(a, q)$. Denote the set of all sets U_a by $\mathbf{U}_a^{all}(x_2)$. The random value $\sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_{\psi}(x)$ is a function of the set $\mathbf{U}_a(x)$ i.e.

$$
\sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_{\psi}(x) = K\!\left(\mathbf{U}_a(x)\right) . \tag{10}
$$

10

Further, with the full probability formula,

$$\mathbb{E}_\gamma K\big(\mathbf{U}_a(x)\big) = \sum_{U_a \in \mathbf{U}_a^{all}(x_2)} \mathbb{E}_\gamma \Big\{ K\big(\mathbf{U}_a(x)\big) \,|\, \mathbf{U}_a(x) = U_a \Big\} \, \mathbb{P}_\gamma \{\mathbf{U}_a(x) = U_a\}$$

$$= \sum_{U_a \in \mathbf{U}_a^{all}(x_2)} K(U_a) \mathbb{P}_\gamma \{\mathbf{U}_a(x) = U_a\} \ . \tag{11}$$

Let us represent the value $\mathbb{P}_\gamma \{\mathbf{U}_a(x) = U_a\}$, which is the probability that the random set $\mathbf{U}_a(x)$ take the value U_a under the condition that the operational intervals are known. If the random values $\mathbf{u}_{a,q}^\theta$, $\mathbf{v}_{a,q}^\theta$ take the values $u_{a,q}^\theta$, $v_{a,q}^\theta$, it means that from the beginning of the subchain (a, q) to the end of the processing of the batch (a, q, θ) a time period with length

$$u_{aq}^\theta x_2 + \sum_{\tau=1}^\theta \sum_{j=1}^{J(a,q,\tau)} \mathbf{r}_{aq}^{\tau j}$$

passed and this length exceeds $v_{aq}^\theta x_1$ and is less than $(v_{aq}^\theta + 1)x_1$. Thus $\mathbf{u}_{a,q}^\theta = u_{a,q}^\theta$, $\mathbf{v}_{a,q}^\theta = v_{a,q}^\theta$ if and only if

$$v_{aq}^\theta x_1 \le u_{aq}^\theta x_2 + \sum_{\tau=1}^\theta \sum_{j=1}^{J(a,q,\tau)} \mathbf{r}_{aq}^{\tau j} \le (v_{aq}^\theta + 1)x_1 \ . \tag{12}$$

The probability that this inequality is satisfied for $\theta = 1, \ldots, \Theta(a, q)$; $j = 1, \ldots, J(a, q, \theta)$, under the condition that the operational intervals $\gamma_\ell, \ell = 1, 2, \ldots$ are known, is equal to

$$\phi_{aq}(x) =$$

$$\int_{\substack{v_{aq}^\theta x_1 \le u_{aq}^\theta x_2 + \sum_{\tau=1}^\theta \sum_{j=1}^{J(a,q,\tau)} r_{aq}^{\tau j} \le (v_{aq}^\theta + 1)x_1 \\ r_{aq}^{\theta j} \ge 0,\, \theta=1,\ldots,\Theta(a,q);\, j=1,\ldots,J(a,q,\theta)}} \prod_{\theta=1}^{\Theta(a,q)} \prod_{j=1}^{J(a,q,\theta)} \varrho\left(r_{aq}^{\theta j}\right) \mathrm{d}r_{aq}^{11} \ldots \mathrm{d}r_{aq}^{\Theta(a,q)J(a,q,\Theta(a,q))} \ . \tag{13}$$

Since the repair intervals are independent random values, the conditional probability $\mathbb{P}_\gamma \{\mathbf{U}_a(x) = U_a\}$ is equal to

$$\mathbb{P}_\gamma \{\mathbf{U}_a(x) = U_a\} = \prod_{q=1}^{Q(a)} \phi_{aq}(x) \ . \tag{14}$$

Combining (9), (10), (11) and (14) we have

$$\mathbb{E}\,\mathbf{L}_\Psi(x) = \mathbb{E} \sum_{a=1}^{\mathbf{A}(x_2)} \sum_{U_a \in \mathbf{U}_a^{all}(x_2)} K(U_a) \prod_{q=1}^{Q(a)} \phi_{aq}(x) \ . \tag{15}$$

3.2 Derivative

The functions $\phi_{aq}(x)$, $a = 1,\ldots,A$, $q = 1,\ldots,Q(a)$ are smooth with respect to x_1. Lebesgue's theorem and equality (15) imply

$$\frac{\partial}{\partial x_1}\mathbb{E}\,\mathbf{L}_\Psi(x) = \mathbb{E}\sum_{a=1}^{\mathbf{A}(x_2)}\sum_{U_a\in\mathbf{U}_a^{all}(x_2)} K(U_a)\frac{\partial}{\partial x_1}\prod_{q=1}^{Q(a)}\phi_{aq}(x)\,. \tag{16}$$

Further we have

$$\frac{\partial}{\partial x_1}\prod_{q=1}^{Q(a)}\phi_{aq}(x) = \prod_{q=1}^{Q(a)}\phi_{aq}(x)\sum_{q=1}^{Q(a)}\left[\phi_{aq}^{-1}(x)\frac{\partial}{\partial x_1}\phi_{aq}(x)\right]\,. \tag{17}$$

At first we calculate the derivative of the function $\phi_{aq}(x)$ with respect to x_1. Here we omit the indexes a and q to simplify formulas

$$\phi(x) = \int_{\substack{v^\theta x_1 \le u^\theta x_2 + \sum\limits_{\tau=1}^\theta\sum\limits_{j=1}^{J(\tau)} r^{\tau j} \le (v^\theta+1)x_1 \\ r^{\theta j}\ge 0,\,\theta=1,\ldots,\Theta;\, j=1,\ldots,J(\theta)}}\prod_{\theta=1}^{\Theta}\prod_{j=1}^{J(\theta)}\varrho\left(r^{\theta j}\right)\,\mathrm{d}r^{11}\ldots\mathrm{d}r^{\Theta J(\Theta)} =$$

$$\int_{\substack{v^\theta \le x_1^{-1}\left(u^\theta x_2 + \sum\limits_{\tau=1}^\theta\sum\limits_{j=1}^{J(\tau)} r^{\tau j}\right) \le v^\theta+1 \\ r^{\theta j}\ge 0,\,\theta=1,\ldots,\Theta;\, j=1,\ldots,J(\theta)}}\prod_{\theta=1}^{\Theta}\prod_{j=1}^{J(\theta)}\varrho\left(r^{\theta j}\right)\,\mathrm{d}r^{11}\ldots\mathrm{d}r^{\Theta J(\Theta)}\,. \tag{18}$$

Denote

$$f(x,r) = \begin{pmatrix} f_1(x,r) \\ \vdots \\ f_k(x,r) \end{pmatrix} = \begin{pmatrix} v^1 - x_1^{-1}\left(u^1 x_2 + \sum\limits_{j=1}^{J(1)} r^{1j}\right) \\ -(v^1+1) + x_1^{-1}\left(u^1 x_2 + \sum\limits_{j=1}^{J(1)} r^{1j}\right) \\ \vdots \\ v^\Theta - x_1^{-1}\left(u^\Theta x_2 + \sum\limits_{\tau=1}^\Theta\sum\limits_{j=1}^{J(\tau)} r^{\tau j}\right) \\ -(v^\Theta+1) + x_1^{-1}\left(u^\Theta x_2 + \sum\limits_{\tau=1}^\Theta\sum\limits_{j=1}^{J(\tau)} r^{\tau j}\right) \\ -r^{11} \\ \vdots \\ -r^{\Theta J(\Theta)} \end{pmatrix}\,, \tag{19}$$

$$m = \sum_{\tau=1}^{\Theta} J(\tau)\,, \quad k = 2\Theta + m\,,$$

$$\mu(x) = \{r\in\mathbb{R}^m \,:\, f(x,r)\le 0\} \stackrel{\mathrm{def}}{=} \{r\in\mathbb{R}^m \,:\, f_l(x,r)\le 0,\, 1\le l\le k\}\,.$$

The vector function $f : \mathbb{R}^2 \times \mathbb{R}^m \to \mathbb{R}^k$ is smooth with respect to the variables x, r. With these notations

$$\phi(x) = \int\limits_{\mu(x)} \prod_{\theta=1}^{\Theta} \prod_{j=1}^{J(\theta)} \varrho\left(r^{\theta j}\right) \, \mathrm{d}r^{11} \ldots \mathrm{d}r^{\Theta J(\Theta)} . \tag{20}$$

We use formula (82) of the Appendix to differentiate function (20). Suppose $l = 2\Theta$. Initially we must solve the equation (see (83))

$$H_l(x,r) \, \nabla_r f_{1l}(x,r) + \frac{\partial}{\partial x_1} f_{1l}(x,r) = 0 . \tag{21}$$

Let us calculate $\nabla_r f_{1l}(x,r)$

$$\nabla_r f_{1l}(x,r) = \left(\nabla_r f_1(x,r), \ldots, \nabla_r f_l(x,r)\right) ,$$

$$- \nabla_r f_{2\theta-1}(x,r) = \nabla_r f_{2\theta}(x,r) = x_1^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} , \quad \theta = 1, \ldots, \Theta - 1 , \tag{22}$$

$$- \nabla_r f_{2\Theta-1}(x,r) = \nabla_r f_{2\Theta}(x,r) = x_1^{-1} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} . \tag{23}$$

Further let us calculate $\frac{\partial}{\partial x_1} f_{1l}(x,r)$

$$\frac{\partial}{\partial x_1} f_{1l}(x,r) = x_1^{-2} \left(u^1 x_2 + \sum_{j=1}^{J(1)} r^{1j}, \ -u^1 x_2 - \sum_{j=1}^{J(1)} r^{1j}, \ldots, \right.$$
$$\left. u^\Theta x_2 + \sum_{\tau=1}^{\Theta} \sum_{j=1}^{J(\tau)} r^{\tau j}, \ -u^\Theta x_2 - \sum_{\tau=1}^{\Theta} \sum_{j=1}^{J(\tau)} r^{\tau j} \right) .$$

Equation (21) has many smooth solutions. For example the vector $H_l^*(x,r)$

$$H_l^*(x,r) = \left(h^{11}, \ldots, h^{\Theta J(\Theta)}\right) , \tag{24}$$

$$h^{\theta j} = x_1^{-1}[r^{\theta j} + (u^\theta - u^{\theta-1})x_2 J^{-1}(\theta)] , \quad (u^0 = 0) .$$

is a solution of equation (21). Suppose that the function $p(r)$

$$p(r) = \prod_{\theta=1}^{\Theta} \prod_{j=1}^{J(\theta)} \varrho\left(r^{\theta j}\right)$$

13

is equal to zero on the boundary of the set

$$r^{\theta j} \geq 0, \quad \theta = 1, \ldots, \Theta, \quad j = 1, \ldots, J(\theta).$$

For example, this is valid if $\varrho(0) = 0$. In this case equation (82) implies

$$\frac{\partial}{\partial x_1} \phi(x) = \int\limits_{\mu(x)} \operatorname{div}_r\Big(p(r) H_l^*(x, r)\Big) \, dr, \tag{25}$$

where

$$\operatorname{div}_r\Big(p(r) H_l^*(x, r)\Big) = \sum_{\theta=1}^{\Theta} \sum_{j=1}^{J(\theta)} \frac{\partial\Big(p(r)\, h^{\theta j}\Big)}{\partial r^{\theta j}}. \tag{26}$$

By (25) and (26)

$$\frac{\partial}{\partial x_1} \phi(x) = \int\limits_{\mu(x)} \left[\sum_{\theta=1}^{\Theta} \sum_{j=1}^{J(\theta)} p^{-1}(r) \frac{\partial\Big(p(r)\, h^{\theta j}\Big)}{\partial r^{\theta j}} \right] p(r) \, dr. \tag{27}$$

Finally, we have from (17) and (27)

$$\frac{\partial}{\partial x_1} \prod_{q=1}^{Q(a)} \phi_{aq}(x) = \int\limits_{\mu_a(x)} K_1(U_a, r_a) \, p_a(r_a) \, dr_a, \tag{28}$$

where

$$r_{aq} = \Big(r_{aq}^{11}, \ldots, r_{aq}^{\Theta(a,q) J(a,q,\Theta(a,q))} \Big), \qquad r_a = \Big(r_{a1}, \ldots, r_{aQ(a)} \Big),$$

$$p_{aq}(r_{aq}) = \prod_{\theta=1}^{\Theta(a,q)} \prod_{j=1}^{J(a,q,\theta)} \varrho\big(r_{aq}^{\theta j}\big), \quad p_a(r_a) = \prod_{q=1}^{Q(a)} p_{aq}(r_{aq}), \tag{29}$$

$$\{ r_{aq} : f_{aq}(x, r_{aq}) \leq 0 \} \overset{\text{def}}{=} \{ r_{aq} : v_{aq}^{\theta} \leq x_1^{-1} \left(u_{aq}^{\theta} x_2 + \sum_{\tau=1}^{\theta} \sum_{j=1}^{J(a,q,\tau)} r_{aq}^{\tau j} \right) \leq (v_{aq}^{\theta} + 1),$$

$$r_{aq}^{\theta j} \geq 0, \, \theta = 1, \ldots, \Theta(a, q); \, j = 1, \ldots, J(a, q, \theta) \},$$

$$\mu_a(x) = \{ r_{aq} : f_{aq}(x, r_{aq}) \leq 0, \, q = 1, \ldots, Q(a) \},$$

$$h_{aq}^{\theta j} = x_1^{-1}[r_{aq}^{\theta j} + (u_{aq}^{\theta} - u_{aq}^{\theta-1}) x_2 J^{-1}(a, q, \theta)], \quad (u_{aq}^{0} = 0), \tag{30}$$

$$K_1(U_a, r_a) = \sum_{q=1}^{Q(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{J(a,q,\theta)} \left(p_{aq}^{-1}(r_{aq}) \frac{\partial\Big(p_{aq}(r_{aq})\, h_{aq}^{\theta j}\Big)}{\partial r_{aq}^{\theta j}} \right).$$

With (29) and (30)

$$p_{aq}^{-1}(r_{aq}) \frac{\partial\Big(p_{aq}(r_{aq})\, h_{aq}^{\theta j}\Big)}{\partial r_{aq}^{\theta j}} = p_{aq}^{-1}(r_{aq}) \left(h_{aq}^{\theta j} \frac{\partial p_{aq}(r_{aq})}{\partial r_{aq}^{\theta j}} + p_{aq}(r_{aq}) \frac{\partial h_{aq}^{\theta j}}{\partial r_{aq}^{\theta j}} \right)$$

14

$$= h_{aq}^{\theta j} \varrho^{-1}(r_{aq}^{\theta j}) \frac{\partial \varrho(r_{aq}^{\theta j})}{\partial r_{aq}^{\theta j}} + x_1^{-1} = h_{aq}^{\theta j} \frac{\partial \ln \varrho(r_{aq}^{\theta j})}{\partial r_{aq}^{\theta j}} + x_1^{-1} . \tag{31}$$

Therefore

$$K_1(U_a, r_a) = \sum_{q=1}^{Q(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{J(a,q,\theta)} \left(h_{aq}^{\theta j} \frac{\partial \ln \varrho(r_{aq}^{\theta j})}{\partial r_{aq}^{\theta j}} + x_1^{-1} \right) . \tag{32}$$

Equation (28) implies

$$K(U_a) \frac{\partial}{\partial x_1} \prod_{q=1}^{Q(a)} \phi_{aq}(x)$$

$$= \mathbb{E}_\gamma \left\{ K\big(\mathbf{U}_a(x)\big) K_1\big(\mathbf{U}_a(x), \mathbf{r}_a\big) \,\middle|\, \mathbf{U}_a(x) = U_a \right\} \mathbb{P}_\gamma \left\{ \mathbf{U}_a(x) = U_a \right\} . \tag{33}$$

By (16) and (33)

$$\frac{\partial}{\partial x_1} \mathbb{E} \, \mathbf{L}_\Psi(x) = \mathbb{E} \sum_{a=1}^{\mathbf{A}(x_2)} K\big(\mathbf{U}_a(x)\big) K_1\big(\mathbf{U}_a(x), \mathbf{r}_a\big) , \tag{34}$$

where (see (10) and (32))

$$K\big(\mathbf{U}_a(x)\big) = \sum_{\psi \in \boldsymbol{\delta}(x_2,a)} \zeta_\psi(x) , \tag{35}$$

$$K_1\big(\mathbf{U}_a(x), \mathbf{r}_a\big) = \sum_{q=1}^{\mathbf{Q}(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{\mathbf{J}(a,q,\theta)} \left(\mathbf{h}_{aq}^{\theta j} \frac{\partial \ln \varrho(\mathbf{r}_{aq}^{\theta j})}{\partial \mathbf{r}_{aq}^{\theta j}} + x_1^{-1} \right) . \tag{36}$$

For example, suppose that ϱ is a gamma distribution, i.e.

$$\varrho(r) = \frac{\lambda^\eta}{\Gamma(\eta)} r^{\eta-1} e^{-\lambda r} , \quad \lambda > 1 , \quad \eta > 1 ,$$

then

$$\frac{\partial \ln \varrho(r)}{\partial r} = \frac{\partial}{\partial r} \left(\ln \frac{\lambda^\eta}{\Gamma(\eta)} + \ln r^{\eta-1} + \ln e^{-\lambda r} \right) = \frac{\eta-1}{r} - \lambda .$$

Thus, for this special case

$$\frac{\partial}{\partial x_1} \mathbb{E} \, \mathbf{L}_\Psi(x) = \mathbb{E} \sum_{a=1}^{\mathbf{A}(x_2)} \left[\left(\sum_{\psi \in \boldsymbol{\delta}(x_2,a)} \zeta_\psi(x) \right) \sum_{q=1}^{\mathbf{Q}(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{\mathbf{J}(a,q,\theta)} \left(\mathbf{h}_{aq}^{\theta j} \left(\frac{\eta-1}{\mathbf{r}_{aq}^{\theta j}} - \lambda \right) + x_1^{-1} \right) \right] .$$

4 The Derivative with Respect to the Second Variable

4.1 The Expectation of the number of Lost Batches

In section 3.1 we presented (see (15)) the function $\mathbb{E}\,\mathbf{L}_\Psi(x)$ as expectation of the function

$$\tilde{g}(x,\omega) \stackrel{\text{def}}{=} \sum_{a=1}^{\mathbf{A}(x_2)} \sum_{U_a \in \mathbf{U}_a^{all}(x_2)} K(U_a) \prod_{q=1}^{Q(a)} \phi_{aq}(x) \,.$$

This function is smooth with respect to the variable x_1 for the fixed value of $\omega \in \Omega$ and the buffer size S. We differentiated the function $\tilde{g}(x,\omega)$ with respect to x_1 and got an estimate for the derivative of the function $\mathbb{E}\,\mathbf{L}_\Psi(x)$. Unfortunately, the function $\tilde{g}(x,\omega)$ is not smooth with respect to the second variable x_2. To smooth over the function $\tilde{g}(x,\omega)$ we have to integrate it additionally with respect to the random variables γ_ℓ, $\ell = 1,2,\ldots$. Let us define the random sequence $\mathbf{B}(x_2)$ which consists of $\mathbf{b}_\ell$, $\ell = 1,2,\ldots$ given by equation

$$\mathbf{b}_\ell \;=\; \min\{\kappa \in \mathcal{N} : \sum_{m=1}^\ell \gamma_m \leq \kappa x_2 \}\,,\;\; \ell = 1,2,\ldots$$

This random sequence $\mathbf{B}(x_2)$ can attain different values. Denote by B^{all} the set of such deterministic sequences of natural numbers. The random function $\tilde{g}(x,\omega)$ is a function of the control variable x and the random sequence $\mathbf{B}(x_2)$ i.e.

$$\tilde{g}(x,\omega) \;=\; \hat{g}\big(x,\mathbf{B}(x_2)\big) \stackrel{\text{def}}{=} \sum_{a=1}^{\mathbf{A}(\mathbf{B}(x_2))} \sum_{U_a \in \mathbf{U}_a^{all}(\mathbf{B}(x_2))} K(U_a) \prod_{q=1}^{Q(a)} \phi_{aq}(x) \,. \tag{37}$$

Then with the full probability formula

$$\mathbb{E}\,\mathbf{L}_\Psi(x) \;=\; \mathbb{E}\,\hat{g}\big(x,\mathbf{B}(x_2)\big) \;=\; \sum_{B \in B^{all}} \mathbb{P}\,\{\mathbf{B}(x_2) = B\}\,\hat{g}(x,B)\,. \tag{38}$$

4.2 Derivative

Since functions $\mathbb{P}\,\{\mathbf{B}(x_2) = B\}$, $\hat{g}\big(x,\mathbf{B}(x_2)\big)$ are smooth with respect to x_2, we can differentiate function (38)

$$\frac{\partial}{\partial x_2}\mathbb{E}\,\mathbf{L}_\Psi(x)$$

$$= \sum_{B \in B^{all}} \left[\hat{g}(x,B)\frac{\partial}{\partial x_2}\mathbb{P}\,\{\mathbf{B}(x_2) = B\} \;+\; \mathbb{P}\,\{\mathbf{B}(x_2) = B\}\frac{\partial}{\partial x_2}\hat{g}(x,B) \right]\,. \tag{39}$$

The derivative $\frac{\partial}{\partial x_2}\hat{g}(x, B)$ can be calculated as

$$\frac{\partial}{\partial x_2}\hat{g}(x, B) = \sum_{a=1}^{A(B)} \sum_{U_a \in U_a^{all}(B)} K(U_a)\frac{\partial}{\partial x_2} \prod_{q=1}^{Q(a)} \phi_{aq}(x) . \tag{40}$$

Analogously to the section 3.2 we use formula (82) (see Appendix) to calculate the derivative $\frac{\partial}{\partial x_2}\phi_{aq}(x)$. Skipping the subscripts a, q we have the function $\phi(x)$ (see (20)). Suppose $l = 2\Theta$. To differentiate the function $\phi(x)$, we must solve the equation (see (19), (83))

$$\tilde{H}_l(x, r)\, \nabla_r f_{1l}(x, r) + \frac{\partial}{\partial x_2} f_{1l}(x, r) = 0 . \tag{41}$$

Let us represent $\frac{\partial}{\partial x_2} f_{1l}(x, r)$

$$\frac{\partial}{\partial x_2} f_{1l}(x, r) = x_1^{-1}\left(-u^1, u^1, \ldots, -u^\Theta, u^\Theta\right) . \tag{42}$$

Equation (41) with (22) and (42) has many smooth solutions. For example, the vector

$$\tilde{H}_l^*(x, r) = \left(\tilde{h}^{11}, \ldots, \tilde{h}^{\Theta J(\Theta)}\right), \quad \tilde{h}^{\theta j} = (-u^\theta + u^{\theta-1})J^{-1}(\theta), \quad (u^0 = 0) . \tag{43}$$

is a solution of equation (41). Suppose that the function $p(r)$

$$p(r) = \prod_{\theta=1}^\Theta \prod_{j=1}^{J(\theta)} \varrho\left(r^{\theta j}\right) \tag{44}$$

is equal to zero on the boundary of the set

$$r^{\theta j} \geq 0, \quad \theta = 1, \ldots, \Theta, \quad j = 1, \ldots, J(\theta) .$$

In this case equation (82) implies

$$\frac{\partial}{\partial x_2}\phi(x) = \int_{\mu(x)} \text{div}_r\left(p(r)\tilde{H}_l^*(x, r)\right) dr , \tag{45}$$

where

$$\text{div}_r\left(p(r)\tilde{H}_l^*(x, r)\right) = \sum_{\theta=1}^\Theta \sum_{j=1}^{J(\theta)} \frac{\partial\left(p(r)\tilde{h}^{\theta j}\right)}{\partial r^{\theta j}} = \sum_{\theta=1}^\Theta \sum_{j=1}^{J(\theta)} \tilde{h}^{\theta j}\frac{\partial p(r)}{\partial r^{\theta j}} . \tag{46}$$

By (45) and (46)

$$\frac{\partial}{\partial x_2}\phi(x) = \int_{\mu(x)} \left[\sum_{\theta=1}^\Theta \sum_{j=1}^{J(\theta)} \tilde{h}^{\theta j} p^{-1}(r)\frac{\partial p(r)}{\partial r^{\theta j}}\right] p(r)\, dr . \tag{47}$$

Hence, we have from (47) and (44)

$$\frac{\partial}{\partial x_2} \prod_{q=1}^{Q(a)} \phi_{aq}(x) = \int_{\mu_a(x)} K_2(U_a, r_a)\, p_a(r_a)\, dr_a , \tag{48}$$

17

where

$$\tilde{h}_{aq}^{\theta j} = (-u_{aq}^{\theta} + u_{aq}^{\theta-1})J^{-1}(a,q,\theta), \quad (u_{aq}^0 = 0), \tag{49}$$

$$K_2(U_a, r_a) = \sum_{q=1}^{Q(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{J(a,q,\theta)} \left(\tilde{h}_{aq}^{\theta j} \, p_{aq}^{-1}(r_{aq}) \frac{\partial p_{aq}(r_{aq})}{\partial r_{aq}^{\theta j}} \right)$$

$$= \sum_{q=1}^{Q(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{J(a,q,\theta)} \left(\tilde{h}_{aq}^{\theta j} \frac{\partial \ln \varrho(r_{aq}^{\theta j})}{\partial r_{aq}^{\theta j}} \right). \tag{50}$$

With (40) and (48)

$$\frac{\partial}{\partial x_2} \hat{g}(x, B) = \sum_{a=1}^{A(B)} \sum_{U_a \in U_a^{all}(B)} \int_{\mu_a(x)} K(U_a) K_2(U_a, r_a) \, p_a(r_a) \, dr_a, \tag{51}$$

Now let us represent the function $\mathbb{P}\{\mathbf{B}(x_2) = B\}$ as probability that the following constraints are satisfied

$$(b_\ell - 1)x_2 \leq \sum_{m=1}^{\ell} \gamma_m \leq b_\ell x_2, \quad \gamma_\ell \geq 0, \quad \ell = 1, \ldots, \ell^{max}.$$

We consider that the random variables γ_ℓ, $\ell = 1, 2, \ldots$ are independent and have density function $\nu(\cdot)$. Denote

$$\gamma = (\gamma_\ell \ldots, \gamma_{\ell^{max}}), \quad \hat{p}(\gamma) = \prod_{\ell=1}^{\ell^{max}} \nu(\gamma_\ell).$$

Thus

$$\mathbb{P}\{\mathbf{B}(x_2) = B\} \stackrel{\text{def}}{=} \varphi(x_2) = \int_{\substack{(b_\ell - 1)x_2 \leq \sum_{m=1}^{\ell} \gamma_m \leq b_\ell x_2 \\ \gamma_\ell \geq 0, \, \ell=1,\ldots,\ell^{max}}} \hat{p}(\gamma) \, d\gamma$$

$$= \int_{\substack{(b_\ell - 1) \leq x_2^{-1} \sum_{m=1}^{\ell} \gamma_m \leq b_\ell \\ \gamma_\ell \geq 0, \, \ell=1,\ldots,\ell^{max}}} \hat{p}(\gamma) \, d\gamma. \tag{52}$$

Denote

$$\hat{f}(x_2, \gamma) = \begin{pmatrix} \hat{f}_1(x_2, \gamma) \\ \vdots \\ \hat{f}_{\hat{k}}(x_2, \gamma) \end{pmatrix} = \begin{pmatrix} b_1 - 1 - x_2^{-1}\gamma_1 \\ -b_1 + x_2^{-1}\gamma_1 \\ \vdots \\ b_{\ell^{max}} - 1 - x_2^{-1} \sum_{\ell=1}^{\ell^{max}} \gamma_\ell \\ -b_{\ell^{max}} + x_2^{-1} \sum_{\ell=1}^{\ell^{max}} \gamma_\ell \\ \\ -\gamma_1 \\ \vdots \\ -\gamma_{\ell^{max}} \end{pmatrix}, \tag{53}$$

18

$$\hat{m} = \ell^{max}, \quad \hat{k} = 3\hat{m},$$

$$\hat{\mu}(x_2) = \{\gamma \in \mathbb{R}^{\hat{m}} : \hat{f}(x_2, \gamma) \le 0\} \stackrel{\text{def}}{=} \{\gamma \in \mathbb{R}^{\hat{m}} : \hat{f}_l(x_2, \gamma) \le 0, \ 1 \le l \le \hat{k}\}.$$

Let us calculate $\frac{\partial}{\partial x_2}\varphi(x_2)$ with the formula (82). Suppose $\hat{l} = 2\ell^{max}$. We should solve the equation

$$\hat{H}_{\hat{l}}(x_2, \gamma)\nabla_\gamma \hat{f}_{1\hat{l}}(x_2, \gamma) + \frac{\partial}{\partial x_2}\hat{f}_{1\hat{l}}(x_2, \gamma) = 0. \tag{54}$$

We obtain

$$\nabla_\gamma \hat{f}_{1\hat{l}}(x_2, \gamma) = \left(\nabla_\gamma \hat{f}_1(x_2, \gamma), \ldots, \nabla_\gamma \hat{f}_{\hat{l}}(x_2, \gamma)\right),$$

$$-\nabla_\gamma f_{2\ell-1}(x_2, \gamma) = \nabla_\gamma f_{2\ell}(x_2, \gamma) = x_2^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ell = 1, \ldots, \ell^{max} - 1, \tag{55}$$

$$-\nabla_\gamma f_{2\ell^{max}-1}(x_2, \gamma) = \nabla_\gamma f_{2\ell^{max}}(x_2, \gamma) = x_2^{-1}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}. \tag{56}$$

Further let us calculate $\frac{\partial}{\partial x_2}\hat{f}_{1\hat{l}}(x_2, \gamma)$

$$\frac{\partial}{\partial x_2}\hat{f}_{1\hat{l}}(x_2, \gamma) = x_2^{-2}\left(\gamma_1, -\gamma_1, \ldots, \sum_{\ell=1}^{\ell^{max}}\gamma_\ell, -\sum_{\ell=1}^{\ell^{max}}\gamma_\ell\right). \tag{57}$$

Equation (54) has many smooth solutions. For example, the vector $\hat{H}_{\hat{l}}^*(x_2, \gamma)$

$$\hat{H}_{\hat{l}}^*(x_2, \gamma) = x_2^{-1}\gamma = x_2^{-1}(\gamma_1, \ldots, \gamma_{\ell^{max}}) \tag{58}$$

is a solution of equation (54). The equation (82) implies

$$\frac{\partial}{\partial x_2}\varphi(x_2) = \int\limits_{\hat{\mu}(x_2)} \text{div}_\gamma\left(\hat{p}(\gamma)\hat{H}_{\hat{l}}^*(x_2, \gamma)\right) d\gamma$$

$$- \sum_{i=\hat{l}+1}^{\hat{k}} \int\limits_{\partial_i\hat{\mu}(x_2)} \frac{\hat{p}(x_2, \gamma)}{\|\nabla_\gamma \hat{f}_i(x_2, \gamma)\|}\left[\frac{\partial}{\partial x_2}\hat{f}_i(x_2, \gamma) + \hat{H}_{\hat{l}}^*(x_2, \gamma)\nabla_\gamma f_i(x_2, \gamma)\right] dS. \tag{59}$$

Since the functions $\hat{f}_i(x_2, \gamma)$, $i = \hat{l}+1, \ldots, \hat{k}$ do not depend upon x_2, we have

$$\frac{\partial}{\partial x_2} \hat{f}_i(x_2, \gamma) = 0 , \quad i = \hat{l}+1, \ldots, \hat{k} . \tag{60}$$

It is easy to see that for $i = \hat{l}+1, \ldots, \hat{k}$ all components of the vector $\nabla_\gamma f_i(x_2, \gamma)$ are equal to zero except

$$\frac{\partial f_i(x_2, \gamma)}{\partial \gamma_{i-2\ell^{max}}} = -1 . \tag{61}$$

Thus, by (58) and (61)

$$\hat{H}_{\hat{l}}^*(x_2, \gamma) \nabla_\gamma \hat{f}_i(x_2, \gamma) = -x_2^{-1} \gamma_{i-2\ell^{max}} . \tag{62}$$

Since on the set $\partial_i \hat{\mu}(x_2)$ the value $\gamma_{i-2\ell^{max}}$ is equal to zero for $i = \hat{l}+1, \ldots, \hat{k}$, we obtain (60) and (62)

$$\sum_{i=\hat{l}+1}^{\hat{k}} \int_{\partial_i \hat{\mu}(x_2)} \frac{\hat{p}(x_2, \gamma)}{\|\nabla_\gamma \hat{f}_i(x_2, \gamma)\|} \left[\frac{\partial}{\partial x_2} \hat{f}_i(x_2, \gamma) + \hat{H}_{\hat{l}}(x_2, \gamma) \nabla_\gamma f_i(x_2, \gamma) \right] \mathrm{d}S = 0 . \tag{63}$$

Hence, by (58) and (59)

$$\begin{aligned}
\frac{\partial}{\partial x_2} \varphi(x_2) &= \int_{\hat{\mu}(x_2)} \mathrm{div}_\gamma \left(\hat{p}(\gamma) \hat{H}_{\hat{l}}^*(x_2, \gamma) \right) \mathrm{d}\gamma \\
&= \int_{\hat{\mu}(x_2)} \sum_{\ell=1}^{\ell^{max}} \frac{\partial \left(\hat{p}(\gamma) x_2^{-1} \gamma_\ell \right)}{\partial \gamma_\ell} \mathrm{d}\gamma \\
&= \int_{\hat{\mu}(x_2)} \left[\hat{p}^{-1}(\gamma) \sum_{\ell=1}^{\ell^{max}} \frac{\partial \left(\hat{p}(\gamma) x_2^{-1} \gamma_\ell \right)}{\partial \gamma_\ell} \right] \hat{p}(\gamma) \, \mathrm{d}\gamma \\
&= \int_{\hat{\mu}(x_2)} \sum_{\ell=1}^{\ell^{max}} \left[\nu^{-1}(\gamma_\ell) \frac{\partial \nu(\gamma_\ell)}{\partial \gamma_\ell} x_2^{-1} \gamma_\ell + \frac{\partial (x_2^{-1} \gamma_\ell)}{\partial \gamma_\ell} \right] \hat{p}(\gamma) \, \mathrm{d}\gamma \\
&= \int_{\hat{\mu}(x_2)} \sum_{\ell=1}^{\ell^{max}} \left[\frac{\partial \ln \nu(\gamma_\ell)}{\partial \gamma_\ell} x_2^{-1} \gamma_\ell + x_2^{-1} \right] \hat{p}(\gamma) \, \mathrm{d}\gamma \\
&= \int_{\hat{\mu}(x_2)} K_3(B, \gamma) \, \hat{p}(\gamma) \, \mathrm{d}\gamma , \tag{64}
\end{aligned}$$

where

$$K_3(B, \gamma) = x_2^{-1} \sum_{\ell=1}^{\ell^{max}} \frac{\partial \ln \nu(\gamma_\ell)}{\partial \gamma_\ell} \gamma_\ell + x_2^{-1} \ell^{max} . \tag{65}$$

With (37), (39), (52) and (64)

$$\frac{\partial}{\partial x_2} \mathbb{E} \, \mathbf{L}_\Psi(x)$$

20

$$= \sum_{B \in B^{all}} \left[\sum_{a=1}^{A(B)} \sum_{U_a \in U_a^{all}(B)} \int_{\mu_a(x)} K(U_a) p_a(r_a) \, dr_a \times \int_{\hat{\mu}(x_2)} K_3(B, \gamma) \hat{p}(\gamma) \, d\gamma \right.$$

$$\left. + \int_{\hat{\mu}(x_2)} \hat{p}(\gamma) \, d\gamma \times \sum_{a=1}^{A(B)} \sum_{U_a \in U_a^{all}(B)} \int_{\mu_a(x)} K(U_a) K_2(U_a, r_a) p_a(r_a) \, dr_a \right]$$

$$= \mathbb{E}\left[K\big(\mathbf{U}_a(x)\big) K_3\big(\mathbf{B}(x_2), \gamma\big) + K\big(\mathbf{U}_a(x)\big) K_2\big(\mathbf{U}_a(x), \mathbf{r}_a\big) \right]$$

$$= \mathbb{E}\, K\big(\mathbf{U}_a(x)\big)\big[K_3\big(\mathbf{B}(x_2), \gamma\big) + K_2\big(\mathbf{U}_a(x), \mathbf{r}_a\big) \big] , \tag{66}$$

where (see (10), (50) and (65)),

$$K\big(\mathbf{U}_a(x)\big) = \sum_{\psi \in \boldsymbol{\delta}(x_2, a)} \zeta_\psi(x) , \tag{67}$$

$$K_2\big(\mathbf{U}_a(x), \mathbf{r}_a\big) = \sum_{q=1}^{Q(a)} \sum_{\theta=1}^{\Theta(a,q)} \sum_{j=1}^{J(a,q,\theta)} \left(\tilde{h}_{aq}^{\theta j} \frac{\partial \ln \varrho(r_{aq}^{\theta j})}{\partial r_{aq}^{\theta j}} \right) . \tag{68}$$

$$K_3\big(\mathbf{B}(x_2), \gamma\big) = x_2^{-1} \sum_{\ell=1}^{\ell^{max}} \frac{\partial \ln \nu(\gamma_\ell)}{\partial \gamma_\ell} \gamma_\ell + x_2^{-1} \ell^{max} . \tag{69}$$

5 Stochastic Quasigradient Algorithm

Let us consider the optimization problem (6)

$$\Phi(x, y) \rightarrow \min_{(x,y) \in X \times Y} \tag{70}$$

By $X \times Y \in \mathbb{R}^2 \times \mathbb{R}^I$ we denote a feasible set for the decision vector (x, y). The standard nonlinear programming techniques are not efficient for this problem since computation of gradients requires computation of multi-dimensional integrals. In order to avoid this, stochastic quasi-gradient algorithms can be used (see, for example, [2] and [3]). One of the most simple stochastic quasi-gradient algorithms is

$$(x^{s+1}, y^{s+1}) = \Pi_{X \times Y}\big((x^s, y^s) - \rho_s \xi^s\big) , \tag{71}$$

where s is the number of the algorithm iteration; (x^s, y^s) is the approximation point of the extremum on the sth iteration; $\Pi_{X \times Y}(\cdot)$ is the orthoprojection operation on the convex set $X \times Y$; $\rho_s > 0$ is a step size; and ξ^s is a stochastic quasi-gradient satisfying the following property

$$E\left[\xi^s \mid (x^0, y^0), \ldots, (x^s, y^s) \right] = \nabla_x \Phi(x^s, y^s) ,$$

i.e. the conditional expectation of the vector ξ^s is equal to the gradient of the function $\Phi(x,y)$ at the point (x^s, y^s). Results of computation experiments show that the algorithm (71) rapidly leads to the point of the extremum if the objective function is not ill-conditioned, i.e., for non-"ravine" functions. In cases where the function $\Phi(x,y)$ is "ravine", the algorithm gets stuck "at the bottom of the ravine". This difficulty may be overcome by using more complicated stochastic quasigradient algorithms with averaging or a variable metrics algorithm [25] with metric transformation.

The algorithm with the averaging of stochastic quasi-gradients was considered by many authors (see, for example [8], [13], [21]).

For improving the convergence rate of the algorithm (71), we use here the stochastic quasi-gradient algorithm with adaptively controlled step sizes (see [22]) and the scaling procedure, suggested by Saridis [18]

$$(x^{s+1}, y^{s+1}) = \Pi_{X \times Y}\big((x^s, y^s) - \rho_s H^s \xi^s\big), \tag{72}$$

where H^s is a scaling matrix.

Although the exact calculation of the derivatives of the function $\Phi(x,y)$ is practically impossible, the stochastic quasigradient ξ^s at the point (x^s, y^s) can be calculated during I runs of the simulation model. Indeed, formula (2) implies that

$$C(x,S) = \Lambda\big(x, S, F(x,S)\big)$$

and

$$\Phi(x,y) = \sum_{i=1}^{I} \Lambda\big(x, S_i, F(x,S_i)\big) y_i \tag{73}$$

Hence,

$$\nabla_x \Phi(x,y) = \sum_{i=1}^{I} y_i \left[\nabla_x \Lambda(x, S_i, z)|_{z=F(x,S_i)} \right.$$
$$\left. + \frac{\partial \Lambda(x, S_i, z)}{\partial z}\bigg|_{z=F(x,S_i)} \nabla_x F(x,S_i) \right]. \tag{74}$$
$$\tag{75}$$

$$\nabla_y \Phi(x,y) = \begin{pmatrix} \Lambda\big(x, S_1, F(x,S_1)\big) \\ \vdots \\ \Lambda\big(x, S_I, F(x,S_I)\big) \end{pmatrix}. \tag{76}$$

Instead of values $F(x,S_i)$, $\nabla_x F(x,S_i)$, $i = 1, \ldots, I$ in formulas (75) and (76), we can use the estimates obtained in the previous sections.

22

Let us suppose that at the sth iteration of algorithm (72) the current approximation (x^s, y^s) is available. Let us sample the simulation model and as an estimate of $F(x^s, S_i)$ we use the value (see (8))

$$\Psi^{-1} L_\Psi(x^s, S_i, \omega^s) . \tag{77}$$

An estimate

$$\Psi^{-1} \sum_{a=1}^{\mathbf{A}(x_2)} K\big(\mathbf{U}_a(x)\big) K_1\big(\mathbf{U}_a(x), \mathbf{r}_a\big) , \tag{78}$$

for the derivative $\frac{\partial}{\partial x_1} F(x^s, S_i)$ follows from formula (34), where $K\big(\mathbf{U}_a(x)\big)$, $K_1\big(\mathbf{U}_a(x), \mathbf{r}_a\big)$ are given by formulas (35), (36). An estimate

$$\Psi^{-1} K\big(\mathbf{U}_a(x)\big)\big[K_3\big(\mathbf{B}, \boldsymbol{\gamma}\big) + K_2\big(\mathbf{U}_a(x), \mathbf{r}_a\big)\big] , \tag{79}$$

for the derivative $\frac{\partial}{\partial x_2} F(x^s, S_i)$ follows from formula (66), where

$$K\big(\mathbf{U}_a(x)\big), \quad K_2\big(\mathbf{U}_a(x), \mathbf{r}_a\big), \quad K_3\big(\mathbf{B}(x_2), \boldsymbol{\gamma}\big)$$

are given by formulas (35), (68) and (69). Since for each (x^s, y^s) and S_i the estimates (77), (78) and (79) can be obtained during one run of the model, we should sample the original simulation model I times in order to obtain the stochastic quasigradient ξ^s with formula (75) and (76).

6 Conclusions and Remarks

The preceding analysis shows that it is indeed possible to develop efficient search algorithms based on performance estimates obtained via simulation. It even appears possible in the case of non-smooth behavior of the sample path as a function of the system parameters to estimate "gradients" which are useful for an optimality search.

The gradients are obtained easily from the simulation run, however, the analysis for finding the right formula is rather complicated.

In order to make this approach practically applicable, it is essential that progress is made in two directions of research: in the first place, formulas should be developed for one-simulation-run estimation of gradients in large classes of realistic models; in the second place, the analysis leading to such formulas should be standardized in such a way that the formulas can be produced (semi-) automatically without requiring mathematical skill of the user. A promising approach seems to be a combination of the ideas in this paper with those presented in [5].

23

7 Appendix. The Analytical Derivatives of the Integrals over Sets Given by Inequalities

Let the function

$$F(x) = \int\limits_{f(x,y) \leq 0} p(x,y) \, dy \tag{80}$$

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

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

For example, let

$$F(x) = P\{f(x, \zeta(\omega)) \leq 0\} \tag{81}$$

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

We present a general formula [24] for the differentiation of integral (80). A gradient of the integral is represented as a sum of integrals taken over a volume and over a surface.

Let us introduce the following shorthand notations

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

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

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

$$\mathrm{div}_y H = \begin{pmatrix} \sum\limits_{i=1}^{m} \frac{\partial h_{1i}}{\partial y_i} \\ \vdots \\ \sum\limits_{i=1}^{m} \frac{\partial h_{ni}}{\partial y_i} \end{pmatrix} , \text{where} \quad H = \begin{pmatrix} h_{11} , & \cdots , & h_{1m} \\ & \vdots & \\ h_{n1} , & \cdots , & h_{nm} \end{pmatrix} .$$

We define

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

24

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

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

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

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

There is freedom in the choice of the sets K_1 and K_2 and the representation of the gradient of the function (80). At first we consider the case when the subsets K_1 and K_2 are not empty. In this case the derivative of integral (80) is given by the formula

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \Big[\nabla_x p(x, y) \;+\; \mathrm{div}_y\big(p(x, y) H_l(x, y)\big) \Big] \, \mathrm{d}y \;-$$

$$-\; \sum_{i=l+1}^{k} \int\limits_{\partial_i \mu(x)} \frac{p(x, y)}{\|\nabla_y f_i(x, y)\|} \Big[\nabla_x f_i(x, y) \;+\; H_l(x, y) \, \nabla_y f_i(x, y) \Big] \, \mathrm{d}S \;, \tag{82}$$

where the matrix function $H_l : R^n \times R^m \to R^{n \times m}$ satisfies the equation

$$H_l(x, y) \, \nabla_y f_{1l}(x, y) \;+\; \nabla_x f_{1l}(x, y) \;=\; 0 \;. \tag{83}$$

The last equation can have a lot of solutions and we can choose an arbitrary solution, differentiable with respect to the variable y .

Further, let us present the derivative of function (80) for the case with the empty set K_1. In this case the matrix function H_l is absent and

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \nabla_x p(x, y) \, \mathrm{d}y \;-\; \sum_{i=1}^{k} \int\limits_{\partial_i \mu(x)} \frac{p(x, y)}{\|\nabla_y f_i(x, y)\|} \nabla_x f_i(x, y) \, \mathrm{d}S \;.$$

Finally, let us consider a formula for the derivative of function (80) for the case with the empty set K_2. In this case the integral over the surface is absent and the derivative is represented as an integral over the volume

$$\nabla_x F(x) \;=\; \int\limits_{\mu(x)} \Big[\nabla_x p(x, y) \;+\; \mathrm{div}_y\big(p(x, y) H(x, y)\big) \Big] \, \mathrm{d}y \;,$$

where the matrix function $H : R^n \times R^m \to R^{n \times m}$ satisfies the equation

$$H(x, y) \, \nabla_y f(x, y) \;+\; \nabla_x f(x, y) \;=\; 0 \;.$$

References

[1] J.A. Buzacott (1982) "Optimal" operating rules for automatied manufacturing systems. *IEEE - Transactions on Automatic Control* **27**, pp. 80-86.

[2] Ermoliev, Yu. (1976): *Stochastic Programming Methods*. Nauka, Moscow. (In Russian)

[3] Ermoliev, Yu. (1988): *Stochastic Quasi-Gradient Methods*. In: "Numerical Techniques for Stochastic Optimization" Eds. Yu. Ermoliev and R. Wets, Springer-Verlag, 393–401.

[4] Ermoliev, Yu. and A. Gaivoronski (1991): *On Optimization of Discontinuous Systems*. Working Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, WP-91-49, 46 p.

[5] Ermoliev Yu., V. Norkin, and R. Wets (1992): *The Minimization of Discontinuous Functions: Mollifier Subgradients*. Working Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, WP-92-73.

[6] Freedman, D. (1971): *Markov Chains*, Holden Day, San Francisco.

[7] W.B. Gong, Y.C. Ho (1987) Smoothed (conditional) perturbation analysis of discrete event dynamic systems. *IEEE - Transactions on Automatic Control* **32**, pp. 858-866.

[8] A.M. Gupal, L.T. Bazhenov. (1972): *A Stochastic Analog of the Methods of Conjugate Gradients*. Kibernetika, 1, 124–126, (in Russian).

[9] Y.C. Ho, M.A. Eyler, T.T. Chien (1979) A gradient technique for general buffer storage design in a serial production line. *Intern. J. Production Res.* **17**, pp. 557-580.

[10] Y.C. Ho, S. Li (1988) Extensions of infinitesimal perturbation analysis. *IEEE - Transactions on Automatic Control* **33**, 427-438.

[11] M.B.M. de Koster (1989) *Capacity oriented analysis and design of production systems*. Springer-Verlag, Berlin (LNEMS 323).

[12] H.J. Kushner, D.S. Clark (1978): *Stochastic Approximation for Constrained and Unconstrained Systems*. Appl. Math. 26, Springer.

[13] H.J. Kushner, Hai-Huang. (1981): *Asymptotic Properties of Stochastic Approximation with Constant Coefficients.* SIAM Journal on Control and Optimization, 19, 87–105.

[14] D. Mitra (1988) Stochastic fluid models. in: P.-J. Courtois, G. Latouche (eds.) *Performance '87.* Elsevier, Amsterdam. pp. 39-51.

[15] G. Pflug (1992): Applicational Aspects of Stochastic Approximation. MV Seminar, Band 17, Birkhäuser Verlag.

[16] Rockafellar R.T., and R. Wets (1982): On the Interchange of Subdifferentiation and Conditional Expectation for Convex Functionals. *Stochastics,* **7**, pp. 173-182.

[17] R. Rubinstein and A. Shapiro (1990): *Mathematics and Computers in Simulation.* **32**, 373-392.

[18] Saridis, G.M. (1970): Learning applied to successive approximation algorithms, *IEEE Trans. Syst. Sci. Cybern.,* 1970, SSC-6, Apr., pp. 97–103.

[19] Seneta, E. (1981): *Non-Negative Matrices and Markov-Chains,* Springer Verlag, New York.

[20] R. Suri (1989) Perturbation analysis: the state of the art and research issues explained via the *GI/G/*1 queue. *Proceedings of the IEEE* **77**, pp. 114-137.

[21] Syski W. (1988): A Method of Stochastic Subgradients with Complete Feedback Stepsize Rule for Convex Stochastic Approximation Problems. *J. of Optim. Theory and Applic.* Vol. **39**, No. **2**, pp. 487–505.

[22] Uryas'ev, S.P. (1988): *Adaptive Stochastic Quasi–Gradient Methods.* In "Numerical Techniques for Stochastic Optimization". Eds. Yu. Ermoliev and R. Wets. Springer Series in Computational Mathematics **10**, (1988), 373–384.

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

[24] Uryas'ev, S.P. (1991): *Differentiation Formula for Integrals over Sets Given by Inequalities.* Working Paper, International Institute for Applied Systems Analysis, Laxenburg, Austria, WP-91-38, 15 p.

[25] Uryas'ev, S. (1990): *A Stochastic Quasi–Gradient Algorithm with Variable Metric*, in Annals of Operations Research, *Sensitivity Analysis and Optimization of Discrete Event Systems*, (accepted for publication).

[26] J. Wijngaard (1979) The effect of interstage buffer storage on the output of two unreliable production units in series with different production rates. *AIIE - Transactions* **11**, pp. 42-47.

[27] S. Yeralan, E.J. Muth (1987) A general model of a production line with intermediate buffer and station breakdown. *AIIE - Transactions* **19**, pp. 130-139.