OPTIMIZATION OF COMPLEX SYSTEMS FOR SIMULATION MODELING

N. K. Krivulin

Vestnik Leningradskogo Universiteta. Matematika,
Vol. 23, No. 2, pp. 100-102, 1990

UDC 519.87

Improvements of the efficiency of a complex system (optimization problem) are discussed, where the efficiency coefficient and its gradient are estimated by simulation. Conditions for unbiased estimates of the efficiency gradient in problems of optimization of a class of message switching networks are stated.

In studies of complex systems, one important problem is raising the system efficiency (the optimization problem): find

$$x^* = \arg\max_{x\in X} F(x), \tag{1}$$

where $F: X \to Y$ is a vector efficiency indicator, X is the set of admissible values of system parameters, Y is the set of values of F. If F(x) is a function differentiable on $X \subset R^n$, then for solving problem (1), we can employ optimization algorithms based on gradient $\frac{\partial}{\partial x} F(x)$. For complex systems, the efficiency characteristic is often defined as $F(x) = \mathrm{E}\mathcal{F}(x;\ \omega)$, where $\mathcal{F}$ is the function of a random argument ω and a parameter x (a random function). The explicit form of F(x) is unknown, but the values $\mathcal{F}(x;\ \omega)$ at any fixed x and a realization of ω can be obtained by means of a simulation model of the system.

Let us assume that with probability (1), there is a derivative $\frac{\partial}{\partial x}\mathcal{F}(x;\ \omega)$ on X and that we have an algorithm calculating its values for any x and ω. In that case, it is natural to use the estimate of the derivative

$$\frac{\partial}{\partial x} F(x) \approx \frac{1}{N}\sum_{i=1}^{N} \frac{\partial}{\partial x}\mathcal{F}(x;\ \omega^{(i)}), \tag{2}$$

where $\omega^{(1)}, \ldots, \omega^{(N)}$ are independent realizations of ω. In this paper we consider conditions where estimate (2) used in solution of the optimization problem for complex discrete systems is unbiased.

In order for estimate (2) to be unbiased, the sufficient condition is the equality

$$\frac{\partial}{\partial x}\mathrm{E}\mathcal{F}(x;\ \omega) = \mathrm{E}\frac{\partial}{\partial x}\mathcal{F}(x;\ \omega). \tag{3}$$

on X. The following statement takes place. It is a corollary to the Lebesgue theorem of convergence of majorized sequence [1] (other results can be found, e.g., in [2]).

Theorem 1. Let $(\Omega,\ \mathfrak{A},\ P)$ be a certain probability space $X \subset R^n$. For a random function $f: X\times\Omega \to R^1$ there exists a.c. derivative $\frac{\partial}{\partial x} f(x;\ \omega)$ on X. If a.c.

$$|f(x_1;\ \omega) - f(x_2;\ \omega)| < \lambda(\omega)\|x_1 - x_2\| \ \forall x_1,\ x_2 \in X \tag{4}$$

and $\mathrm{E}\lambda < \infty$, then on X expression (3) takes place.

In the case where a certain function f satisfies the conditions of Theorem 1, we will write $f \in \mathcal{D}$. The functions from $\mathcal{D}$ have the following properties.

Lemma. Let $f,\ g \in \mathcal{D},\ c_1,\ c_2,\ c_3,\ c_4$ be positive constants. In that case:

1) $f+g \in \mathcal{D}$;

2) if α is a random variable $u\ |\alpha| < c_1$, then $\alpha f \in \mathcal{D}$;

3) if $X\ |f| < c_2,\ |g| < c_3$ a.c. then $fg \in \mathcal{D}$;

4) if on $X\ |f| > c_4$ a.c. then $1/f \in \mathcal{D}$;

5) if $f \neq$ g a.c. for any $x \in X$, then $\max\{f,\ g\} \in \mathcal{D}$, $\min\{f,\ g\} \in \mathcal{D}$.

Proof. The validity of statements 1)-4) is established easily. We will prove 5) for $h(x;\ \omega) = \max\{f(x;\ \omega),\ g(x;\ \omega)\}$. We take an arbitrary $x_0 \in X$. A derivative of function h at point x_0> may not exist if at least one of the derivatives $\frac{\partial}{\partial x} f(x;\ \omega)\Big|_{x=x_0}$, $\frac{\partial}{\partial x} g(x;\ \omega)\Big|_{x=x_0}$ does not exist. Besides, h may be undifferentiable in x_0 when $f(x_0;\ \omega) = g(x_0;\ \omega)$. By condition, these events take place with probability zero, and therefore, at almost all $\omega \in \Omega$ there is a derivative $\frac{\partial}{\partial x} h(x;\ \omega)\Big|_{x=x_0}$. Since $x_0 \in X$ is arbitrary, we conclude that there exists a.c. $\frac{\partial}{\partial x} h(x;\ \omega)$ on X.

Suppose that for any $x_1,\ x_2 \in X$ with probability (1)

$$|f(x_1;\ \omega) - f(x_2;\ \omega)| < \lambda_1(\omega)|x_1 - x_2|,\ E\lambda_1 < \infty,$$
$$|g(x_1;\ \omega) - g(x_2;\ \omega)| < \lambda_2(\omega)|x_1 - x_2|,\ E\lambda_2 < \infty.$$

We take an arbitrary f for which the above inequalities take place. We can readily verify the relation

$$|h(x_1;\ \omega) - h(x_2;\ \omega)| < \lambda(\omega)|x_1 - x_2|\ \forall x_1,\ x_2 \in X,$$

where $\lambda(\omega) = \max\{\lambda_1(\omega),\ \lambda_2(\omega)\}$.

Therefore, almost for all ω the function h satisfies (4) and $E\,\lambda = E\max\{\lambda_1,\ \lambda_2\} < E\,\lambda_1 + E\,\lambda_2 < \infty$.

Thus, $h \in \mathcal{D}$. The statement $\min\{f,\ g\} \in \mathcal{D}$ is proved similarly.

We will illustrate an application of these results in an analysis of a relatively universal model of a complex discrete system: a closed network of message switching. The network consists of L nodes that service messages. After a service at a certain network node is completed, the message is transferred to a different node according to a certain routing procedure. If the node is free, the message occupies it; otherwise, the message is added to the queue of messages waiting to be served at the given node. The messages are picked from the queue according to FCFS discipline.

We denote by τ_{ij} the service time at the node i of a message with serial number j τ_{ij} (i = 1, ..., L, j = 1, 2, ...) are mutually independent random variables. The routing procedure is organized as follows. For each node i the sequence $M_i = (m_{i1}, m_{i2}, \ldots)$ is given where m_{ij} is the number of the node where the message with serial number j serviced at node i should be transferred; $m_{ij} \in \{1, \ldots, L\}$, $i = 1, \ldots, L$, $j = 1, 2, \ldots$. At the initial (0) time point all the nodes are free. There are n ≥ 0 messages in the queue of the node i. Incidentally, an open network can be regarded as a special case of this model. For instance, for representation of the input message flows arriving into the network from outside, it is sufficient to introduce nodes for which the initial queues are set equal to ∞.

We can demonstrate that for this model the values of time points where the next service of each message is completed can be expressed in terms of τ_{ij} by using the operations of addition, max, and min. This can be illustrated by a simple example of a network consisting of three nodes, We set $M_1 = (2, 1, 1, 3, \ldots)$, $M_2 = (1, 3, 1, 1, \ldots)$, $M_3 = (2, 3, 1, 2, \ldots)$. At the initial time point $n_1 = n_2 = n_3 = 1$. We take a certain node, e.g., node 2. We introduce notations α_j, β_j, γ_j for the times of arrival at node 2, the beginning, and the end of the processing of message j. For this model we have

$$\begin{array}{lll} \alpha_1 = 0, & \beta_1 = 0, & \gamma_1 = \tau_{21}, \\ \alpha_2 = \min\{\tau_{11}, \tau_{31}\}, & \beta_2 = \max\{\gamma_1, \alpha_2\}, & \gamma_2 = \beta_2 + \tau_{22}, \\ \alpha_3 = \max\{\tau_{11}, \tau_{31}\}, & \beta_3 = \max\{\gamma_2, \alpha_3\}, & \gamma_3 = \beta_3 + \tau_{23} \text{ etc.} \end{array}$$

After substitutions we obtain

$$\gamma_3 = \max\{\max\{\tau_{21}, \min\{\tau_{11}, \tau_{31}\}\} + \tau_{22}, \max\{\tau_{11}, \tau_{31}\}\} + \tau_{23}.$$

Note that here, in expressions of the type $\max\{\xi, \eta\}$, either ξ and η are independent, or $\eta = \eta_1 + \eta_2$, where η_2 are pairwise independent with ξ and η_1 (likewise, for min).

Suppose now that the durations of processing of messages at the nodes are defined by random functions $\tau_{ij}(x; \omega)$ which depend on the parameter $x \in X$. We take a node l and specify k as the number of messages to be processed by node l. As before, we denote by $\alpha_j(x; \omega)$, $\beta_j(x; \omega)$, $\gamma_j(x; \omega)$ the times of arrival beginning and completion of the processing of message j at node l.

For optimization problem (1), the following characteristics are usually adopted as measures of network efficiency:

1) $F_1(x) = \mathrm{E}\,\mathcal{F}_1(x; \omega)$ - the operation speed of the node l, $\mathcal{F}_1(x; \omega) = k/\gamma_k(x; \omega)$;

2) $F_2(x) = \mathrm{E}\,\mathcal{F}_2(x; \omega)$ - the utilization coefficient of the node l, $\mathcal{F}_2(x; \omega) = \sum_{j=1}^{k} \tau_{lj}(x; \omega)/\gamma_k(x; \omega)$;

3) $F_3(x) = \mathrm{E}\,\mathcal{F}_3(x; \omega)$ - the average waiting time at the node l, $\mathcal{F}_3(x; \omega) = \sum_{j=1}^{k} (\beta_j(x; \omega) - \alpha_j(x; \omega))/k$

(here the problem of finding a minimum is posed).

Theorem 2. Suppose that in this model $\tau_{ij} \in \mathcal{D}$ and at each $x \in X$ τ_{ij} $(i = 1, \ldots, L, j = 1, 2, \ldots)$ are mutually independent random variables with continuous distributions; c_1, c_2, and c_3 are positive constants. In that case

1) $\mathcal{F}_3 \in \mathcal{D}$;

2) if on X $\tau_{ij} \geq c_1$ a.c., then $\mathcal{F}_1 \in \mathcal{D}$;

3) if on X $c_2 \leq \tau_{ij} \leq c_1$ a.c., then $\mathcal{F}_2 \in \mathcal{D}$.

The proof of Theorem 2 can be constructed on the basis of the lemma. The objective is to demonstrate that α_j, β_j, $\gamma_j \in \mathcal{D}$ $j = 1, 2, \ldots$. This follows from the above-mentioned properties of the model and from the independence and continuity of τ_{ij}.

It can readily be verified that Theorem 2 holds also for a random routing procedure where the choice of the next node l for message j just processed at node i is accomplished with given probabilities:

$$P\{m_{ij} = l\} = p_{il}^{(j)};\ l \in \{1, \ldots, L\},\ i = 1, \ldots, L,\ j = 1, 2 \ldots$$

The conditions of Theorem 2 do not assume any special type of distribution (such as exponential) of the service time at nodes. Besides, it does not require that the durations of service of all messages at a given node follow the same distribution pattern.

In practical optimization efforts, obtaining the values of $\frac{\partial}{\partial x} \mathcal{F}(x;\ \omega)$ in simulation models is often difficult. For discrete systems, a new promising approach has been suggested based on analysis of small perturbations of the parameter [3,4]. With this approach, effective procedures have been developed for calculating the values of $\frac{\partial}{\partial x} \mathcal{F}(x;\ \omega)$ for optimization of message switching networks discussed in the paper. Our analysis has confirmed that estimates of the type of (2) in these problems are unbiased.

REFERENCES

1. M. Love, Theory of Probabilities [Russian translation], Moscow, 1962.
2. X. R. Cao, IEEE Trans. Autom. Control, vol. 30, no. 9, pp. 845-853, 1985.
3. Y. C. Ho and X. Cao, J. Optim. Theory Appl., vol. 40, no. 4, pp. 559-582, 1983.
4. R. Suri, J. ACM, vol. 34, no. 3, pp. 686-717, 1987.

26 September 1989