Proceedings of the 1999 Winter Simulation Conference
P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, eds.

SOLVING STOCHASTIC OPTIMIZATION PROBLEMS WITH STOCHASTIC CONSTRAINTS: AN APPLICATION IN NETWORK DESIGN

Gül Gürkan

Center for Economic Research
Tilburg University
5000 LE Tilburg
The NETHERLANDS

A. Yonca Özge

GE Corporate R & D
One Research Circle
Niskayuna, NY 12309
U.S.A.

Stephen M. Robinson

Department of Industrial Engineering
University of Wisconsin-Madison
1513 University Avenue
Madison, WI 53706, U.S.A.

ABSTRACT

Recently sample-path methods have been successfully used in solving challenging simulation optimization and stochastic equilibrium problems. In this paper we deal with a variant of these methods to solve stochastic optimization problems with stochastic constraints. Using optimality conditions, we convert the problem to a stochastic variational inequality. We outline a set of sufficient conditions for the almost-sure convergence of the method. We also illustrate an application by using the method to solve a network design problem. We find optimal arc capacities for a stochastic network (in which the demand and supply at each node is random) that minimize the sum of the capacity allocation cost and a measure of the expected shortfall in capacity.

1 INTRODUCTION

This paper shows how to use the sample-path technique to solve stochastic constrained optimization problems, which can be seen as special cases of solving stochastic variational inequalities. This capability extends the range of application of sample-path methods, since in some important cases not only are the functions to be optimized stochastic in nature, but the constraints imposed on the problem could also be stochastic. In this section we review the existing forms of the sample-path method and illustrate a case for which it may not be clear how to apply the sample-path method in its usual forms. In the remainder of the paper, we provide a variant of the method and develop the necessary theory to deal with that case. We also report an application in network design, for which we compare the computational performance of sample-path optimization with that of stochastic approximation.

Roughly speaking, sample-path methods are concerned with solving a problem of optimization or equilibrium, involving a limit function f_∞ which we cannot observe. However, we can observe functions f_n that almost surely converge pointwise to f_∞ as $n \to \infty$. In the kind of applications we have in mind, f_∞ is typically a steady-state performance measure or an expected value and we use simulation to observe the f_n's. In systems that evolve over time, we simulate the operation of the system for, say, n time units and then compute an appropriate performance measure. In static systems we repeatedly observe instances of the system and compute an average. In both cases, to observe f_n at different parameter settings we use the method of common random numbers. Furthermore, in many cases derivatives or directional derivatives of the f_n can be obtained using well-established methods of gradient estimation such as infinitesimal perturbation analysis (IPA); see Ho and Cao (1991) and Glasserman (1991). The key point is the following fact: once we fix n and a sample point (using common random numbers), f_n becomes a deterministic function. The sample-path methods then solve the resulting deterministic problem (using f_n with the fixed sample path selected), and take the solution as an estimate of the true solution. Clearly, the availability of very powerful deterministic solvers (both for optimization and for equilibrium problems) makes this approach very attractive.

We distinguish between two types of problems. The first involves optimization; in this case the f_n are extended-real-valued functions: $f_n : \mathbf{R}^k \to \mathbf{R} \cup \{\pm\infty\}$ for $1 \le n \le \infty$, and we are interested in solving

$$\min_x f_\infty(x). \tag{1}$$

This setup also covers optimization problems with deterministic constraints since we can always set $f_\infty(x) = +\infty$ for x that do not satisfy the constraints.

The second problem type is a variational inequality; in this case the f_n are vector-valued functions: $f_n : \mathbf{R}^k \to \mathbf{R}^k$ for $1 \le n \le \infty$, and our aim is to find a point $x_0 \in C$, if any exists, satisfying

$$\text{for each } x \in C, \quad \langle x - x_0, f_\infty(x_0)\rangle \ge 0, \tag{2}$$

where $\langle y, z \rangle$ denotes the inner product of y and z, and C is a polyhedral convex subset of $\mathbf{R}^k$. An equivalent way of expressing (2) is via the *generalized equation*

$$0 \in f_\infty(x_0) + N_C(x_0) \tag{3}$$

where $N_C(x)$ is the normal cone of C at x, defined to be the set

$$\{y^* \mid \text{for each } c \in C, \quad \langle y^*, c - x \rangle \leq 0\} \tag{4}$$

provided that $x \in C$, and to be empty otherwise.

The problem (2) models a very large number of equilibrium phenomena in economics, physics, and operations research; for many examples, see Harker and Pang (1990) and Ferris and Pang (1997). One important area where variational inequalities are of use is nonlinear programming, because the first order necessary conditions for local optimality of a given point can be stated as a variational inequality. For example, in the particular special case of an unconstrained optimization problem, the associated variational inequality becomes a nonlinear equation and the method reduces to finding a zero of the gradient. Let f_∞ be this gradient and $C = \mathbf{R}^k$. It is easy to see that solving the first order necessary optimality conditions for this problem expressed in the form (2) is equivalent to finding $x_0 \in \mathbf{R}^k$ such that $f_\infty(x_0) = 0$. Another important special case of variational inequalities arises when we express the first-order necessary optimality conditions for a nonlinear-programming problem with continuously differentiable objective and constraint functions; in this paper we will deal with such a problem.

We are interested in solving the nonlinear optimization problem with a stochastic objective function and stochastic constraints. Consider

$$\begin{array}{ll} \min & F_\infty(x) \\ \text{s.t.} & g_\infty(x) \leq 0, \quad h_\infty(x) = 0 \\ & x \in C. \end{array} \tag{5}$$

In this formulation, C is a polyhedral convex set, F_∞ is a real-valued function, and g_∞ and h_∞ are possibly vector-valued functions. C is used to model deterministic constraints. In addition to the usual objective function F_∞ (or f_∞ in (1)) that we cannot observe, we also have constraints g_∞ and/or h_∞, that cannot be observed but have to be approximated/estimated using simulation. In Section 3, we bring (5) into the form (2) and show how to extend the theory developed for (2) to cover (5).

Let us demonstrate this setup with a simple example. Suppose that we are trying to find the optimal parameters of a control policy that minimizes the long-run expected cost of an inventory system, F_∞ (for example, the cost function may consist of inventory holding and backorder costs). In addition to this, we would also like to maintain an acceptable service level, so we impose a constraint that says that the long-run fill-rate (fraction of orders filled without backlogging) g_∞, should not fall below a preset service level, say α. By simulating this inventory system and using gradient estimation techniques, we can estimate these performance measures as well as their gradients. Using these we can further compute the optimal parameters of the inventory policy. In Section 3 we provide conditions that guarantee the existence and closeness of the estimate solutions to the exact solution. See also Gürkan and Karaesmen (1998) which addresses a similar problem in production and inventory control. In Section 4, we deal with another example that is related to the design of networks.

2 SAMPLE-PATH METHODS

Sample-path methods appeared in Plambeck *et al.* (1993, 1996) and were analyzed in Robinson (1996). That form, called sample-path optimization, concerned the solution of simulation optimization problems with deterministic constraints. The main condition imposed on f_n is their *epiconvergence* to the limit function f_∞. Roughly speaking, epiconvergence is the set convergence of the epigraphs of f_n to the epigraph of f_∞. See Kall (1986) for a treatment of various types of convergence and Rockafellar and Wets (1998) for a treatment of epiconvergence from the perspective of optimization. The proposals of Plambeck *et al.* (1993, 1996) used infinitesimal perturbation analysis (IPA) for gradient estimation. A closely related technique centered around likelihood-ratio methods appeared in Rubinstein and Shapiro (1993).

In Gürkan *et al.* (1996, 1999) we extended the basic idea of using sample-path information to solve stochastic equilibrium problems. There we presented a framework to model such equilibrium problems as stochastic variational inequalities and provided conditions under which equilibrium points of approximating problems (computed via simulation and deterministic variational inequality solvers) converge almost surely to the solution of the limiting problem which we cannot observe. Gürkan *et al.* (1999) also contains a numerical application of the derived theory for finding the equilibrium prices of natural gas as well as the equilibrium quantities to produce in the European natural gas market. Both forms of the sample-path method are explained in Gürkan *et al.* (1998), which also summarizes a number of applications.

Since we will consider solving (5) as a special case of solving (2), in the remainder of this section we briefly summarize the main convergence result for the sample-path method to solve stochastic variational inequalities.

To guarantee the closeness of the solution x_n of (2) with f_n in place of f_∞ to the true solution x_0, we need to impose certain functional convergence on the sequence

$\{f_n\}$. The specific property we require is called *continuous convergence* and is denoted by $\xrightarrow{\mathcal{C}}$; it is equivalent to uniform convergence to a continuous limit on compact sets; see e.g. Kall (1986). The motivating fact is the following: consider a sequence of functions f_n and of points x_n with the property that x_n solves (2) with f_n replacing f_∞ and $x_n \to x_0$ as $n \to \infty$. If $f_n \xrightarrow{\mathcal{C}} f_\infty$ then the limit point x_0 solves (2). Therefore we might reasonably use solutions x_n as estimates of a solution of (2). However, although useful, continuous convergence by itself guarantees neither the existence of solutions x_n nor their convergence.

To guarantee such existence and convergence we need to impose a certain nonsingularity condition called *coherent orientation*. Here we do not go into any detail about this concept, but refer the reader to Gürkan *et al.* (1996, 1999) where we give an extensive description with further references. In the simple case of nonlinear equations, i.e. when $C = \mathbf{R}^k$, this condition reduces to the usual nonsingularity requirement on the Hessian ∇f_∞. In its general form, the coherent orientation condition is a way of extending the idea of nonsingularity to the case of a nontrivial set C.

In Gürkan et al. (1996, 1999) we provide sufficient conditions for the existence of solutions of approximating variational inequalities and their convergence to the exact solution of the limit problem. Briefly, these conditions are

1. With probability one, $f_n \xrightarrow{\mathcal{C}} f_\infty$.
2. The limit variational inequality (2) has a solution x_0.
3. f_∞ has a strong Fréchet derivative $df_\infty(x_0)$ at x_0, and the normal map $df_\infty(x_0)_K$ associated with (2) is coherently oriented, where $K = K(x_0, -f_\infty(x_0))$ is the critical cone to C at $(x_0, -f_\infty(x_0))$:

   $$K = T_C(x_0) \cap \{y \in \mathbf{R}^k \mid \langle y, -f_\infty(x_0)\rangle = 0\},$$

   and $T_C(x_0)$, the tangent cone to C at x_0, is the polar of $N_C(x_0)$.

Condition (3) is rather technical and we will not elaborate on it here; we refer the reader to Gürkan *et al.* (1999). There, we also provide a bound on the distance between the approximate solutions and the limit solution in terms of the uniform norm of f_n and f_∞ on a compact set containing the exact solution.

Building on this framework, we can further extend the sample-path method to solve stochastic optimization problems with stochastic constraints. The next section contains a brief summary of the conditions derived to handle this case.

3 STOCHASTIC CONSTRAINED OPTIMIZATION

In this section we investigate the convergence properties of the method in the special case of constrained optimization with "probabilistic" constraints.

We work with an open set Θ and a polyhedral convex set C in $\mathbf{R}^k$ and functions $F : \Theta \to \mathbf{R}$ and $H : \Theta \to \mathbf{R}^m$. The problem we consider is

$$\begin{array}{ll} \min & F(x) \\ \text{s.t.} & x \in \Theta \cap C \\ & H(x) = 0 \end{array} \tag{6}$$

The representation (6) is more general than it might seem. Suppose we consider the problem of minimizing $Q(y)$ over the set $\{y \in Y \mid g(y) \le 0,\ h(y) = 0\}$ where Y is a polyhedral convex set and g and h are functions from an open subset of $\mathbf{R}^l$ into $\mathbf{R}^q$ and $\mathbf{R}^r$ respectively. We can convert this to the form (6) by introducing slack variables $s \in \mathbf{R}^q$, rewriting the feasible set and the constraint functions as

$$\begin{pmatrix} g(y) + s \\ h(y) \end{pmatrix}, \quad (y, s) \in Y \times \mathbf{R}^q_+, \tag{7}$$

and taking $k = l + q$, $x = (y, s)$, $C = Y \times \mathbf{R}^q_+$, $F(x) = Q(y)$, and

$$H(x) = \begin{pmatrix} g(y) + s \\ h(y) \end{pmatrix}.$$

Therefore the form (6) is quite general, covering any combination of inequalities and equations with a polyhedral convex constraining set. Hence we will concentrate on (6) in the rest of this section.

Let

$$\mathcal{L}(x, u) = F(x) + \langle u, H(x)\rangle, \tag{8}$$

and assume that the point x_0 is a local solution of (6). Then under a constraint qualification, e.g., transversality or non-degeneracy, as well as sufficient differentiability of the problem functions, there exists u_0 in $\mathbf{R}^m$ such that the pair (x_0, u_0) is a solution of the generalized equation:

$$0 \in d\mathcal{L}(x, u) + N_C(x) \times \{0\}^m. \tag{9}$$

Now we consider the situation where we cannot observe the functions F and H but sequences of functions $\{p_n\}$, $\{r_n\}$, and $\{H_n\}$ approximating dF, dH, and H respectively. In the rest of this section we will describe the conditions under which the solution of (9) is related to the solution of

$$0 \in \begin{pmatrix} p_n(x) + u^T r_n(x) \\ H_n(x) \end{pmatrix} + \begin{pmatrix} N_C(x) \\ \{0\}^m \end{pmatrix}. \tag{10}$$

Again, we need a generalized nonsingularity condition. This time it is more convenient to express this condition in terms of a property called *strong regularity* originally introduced by Robinson (1980). Assume that x_0 is a solution of (3) and f_∞ is Fréchet differentiable at x_0. The generalized equation (3) is strongly regular at x_0 if there are neighborhoods U of x_0 and W of the origin in $\mathbf{R}^k$ such that the generalized equation

$$y \in f_\infty(x_0) + df_\infty(x_0)(x - x_0) + N_C(x) \tag{11}$$

defines a single-valued, Lipschitzian map $x(y)$ from W to U, i.e., for each $y \in W$, $x(y)$ is the unique solution in U of (11). Now we are in position to state Theorem 1, the main convergence result.

Theorem 1 *Let Θ be an open subset of $\mathbf{R}^k$ and let C be a polyhedral convex set in $\mathbf{R}^k$. Let x_0 be a point of Θ, u_0 be a point of $\mathbf{R}^m$, and suppose F and H are functions from Θ to $\mathbf{R}$ and $\mathbf{R}^m$ respectively. Let $\{p_n \mid n = 1, 2, \ldots\}$ and $\{r_n \mid n = 1, 2, \ldots\}$ be random functions from Θ to $\mathbf{R}^k$, and $\{H_n \mid n = 1, 2, \ldots\}$ be a random function from Θ to $\mathbf{R}^m$, such that for all $x \in \Theta$ and all finite n the random variables $p_n(x)$, $r_n(x)$, and $H_n(x)$ are defined on a common probability space $(\Omega, \mathcal{F}, P)$. Let $\mathcal{L}(x, u)$ be defined as in (8), $f_n(x, u) = (p_n(x) + u^T r_n(x), H_n(x))$, and assume the following:*

- *a.* *With probability one, each p_n for $n = 1, 2, \ldots$ is continuous and $p_n \xrightarrow{\mathcal{C}} dF$.*
- *b.* *With probability one, each r_n for $n = 1, 2, \ldots$ is continuous and $r_n \xrightarrow{\mathcal{C}} dH$.*
- *c.* *With probability one, each H_n for $n = 1, 2, \ldots$ is continuous and $H_n \xrightarrow{\mathcal{C}} H$.*
- *d.* *(x_0, u_0) is a solution of (9).*
- *e.* *$d\mathcal{L}$ has a strong Fréchet derivative $d^2\mathcal{L}(x_0, u_0)$ at (x_0, u_0) and the generalized equation $0 \in d\mathcal{L}(x, u) + N_C(x) \times \{0\}^m$ is strongly regular at (x_0, u_0) with associated Lipschitz modulus μ.*

Then, there exist compact subsets $C_0 \subset C \cap \Theta$ containing x_0 and $U_0 \subset \mathbf{R}^m$ containing u_0, neighborhoods $X_1 \subset \Theta$ of x_0 and $U_1 \subset \mathbf{R}^m$ of u_0, a positive constant λ, and a set $\Delta \subset \Omega$ of measure zero, with the following properties: for $n = 1, 2, \ldots$ let

$$\xi_n = \sup_{(x,u) \in C_0 \times U_0} \| f_n(\omega, x, u) - d\mathcal{L}(x, u) \|,$$

and

$$X_n(\omega) := \{(x, u) \in (C \cap X_1) \times U_1 \mid 0 \in f_n(x, u) + N_C(x) \times \{0\}^m\}.$$

For each $\omega \notin \Delta$ there is then a finite integer N_ω such that for each $n \geq N_\omega$ the set $X_n(\omega)$ is a nonempty, compact subset of $B((x_0, u_0), \lambda\xi_n)$.

Again, condition (e) is highly technical and we refer the reader to Özge (1997) for a detailed discussion. In words, Theorem 1 says that under certain niceness conditions, for sufficiently large n (i.e., if we go out long enough on the sample-path), the solution set of (10) is nonempty and compact; furthermore, the distance of every such solution of (10) from the exact solution (x_0, u_0) of (9) is bounded by a constant multiple of the uniform norm of $f_n - d\mathcal{L}$ on a compact set. The proof of Theorem 1 is given in Özge (1997) and in Gürkan et al. (1999), along with a rigorous discussion of the relationship between coherent orientation and strong regularity, and some references where several equivalent forms of these generalized nonsingularity conditions are discussed.

For simplicity of exposition, Theorem 1 only dealt with exact solutions of the approximating problems. However, it is easy to verify that a similar result dealing with small perturbations of the approximating problems is valid as well.

Note that when the set $C = \mathbf{R}^k$, then one condition that guarantees strong regularity at (x_0, u_0) is the *strong second-order sufficient condition* together with the linear independence of the gradients of the constraints; see Theorem 4.1 in Robinson (1980). Precisely, the strong second-order sufficient condition in this case says the following: for each nonzero y with $dH(x_0)y = 0$ one has $yd^2\mathcal{L}(x_0, u_0, v_0)y > 0$. Robinson (1980, Theorem 3.1) shows that for general C a positive definiteness condition suffices for strong regularity.

A very relevant work is Shapiro (1993) which provides similar convergence results for stochastic programming problems. Shapiro does not use strong regularity; instead he assumes the convergence of the approximate solutions.

4 APPLICATION: NETWORK DESIGN

As already mentioned, a computational illustration of the variational inequality formulation of the previous section appears in Gürkan et al. (1999). In this section we illustrate the application of the sample-path method to a problem that arises when designing networks, e.g., traffic or communication networks, transportation or distribution systems. This problem does not satisfy the conditions we imposed in the previous section, because some of the functions involved are nonsmooth. It is thus closer in structure to some of the problems investigated in Plambeck et al. (1996). We shall see in what follows that the sample-path approach worked well in spite of the lack of smoothness.

The problem we consider is that of allocating available capacity among the arcs of a network composed of a given (finite) number of supply and demand nodes and a set of arcs connecting these nodes. The demands and supplies at individual nodes are random. There is a cost, d_i, associated with assigning unit capacity to arc i. We would like to allocate the available capacity to the arcs so as to minimize the sum of the capacity allocation cost and a measure of the expected shortfall in capacity.

$$\text{(P)}\quad \begin{array}{ll} \min & E[f(u_1,\ldots,u_k)] + \sum_{i=1}^k d_i u_i \\ \text{s.t.} & \sum_{i=1}^k u_i \le C \\ & u_i \ge 0, \qquad i = 1,\ldots,k \end{array}$$

where C is the total available arc capacity and

$$f(u_1,\ldots,u_k) = \begin{array}{ll} \min & cx \\ \text{s.t.} & Ax = b \\ & 0 \le x_i \le u_i, \quad i = 1,\ldots,k \end{array}$$

where A is the node-arc incidence matrix, c is the cost vector for sending flow x through the network, and b is the random demand/supply vector. When we cannot satisfy the demand due to insufficient arc capacity, f, the cost of the network, increases because we have to use artificial arcs with unlimited arc capacity but large arc costs. Therefore, f is one measure of the shortfall, the incapability to satisfy demand that is due to insufficient arc capacity. For a similar problem (but with no random element) that arises in network synthesis, see Gomory and Hu (1964).

By trying to put a problem in a network format, one gains insight into the problem as well. For example, the arc capacities that we are trying to find may represent actual bounds on the flow on an arc, e.g., the number of trucks to assign on a route in a distribution system, or they may represent the ability of the destination node to handle the arriving flow, e.g., the number and/or the condition of rail tracks leading to a busy harbor. Such examples are numerous, and networks are powerful tools that can be used to model a wide variety of situations. Ahuja, Magnanti, and Orlin (1993) is an excellent treatment of network flows, extensively covering existing theory while pointing out possible real-world applications.

The network we considered is given in Figure 1; it has 12 nodes and 33 arcs. The numbers on the arcs denote the cost of sending unit flow from the origin to the destination of the corresponding arc. Demands and supplies of individual nodes (denoted by b_j) are uniformly distributed random variables with the data given in Table 1. A negative (positive) b_j indicates that node j is a demand (supply) point. We chose d to be a vector with each component equal to 5.

We would like to compare the numerical performance of the sample-path method with the performance of stochastic approximation (SA). In esssence, SA is a gradient-descent method and hence it tends to inherit some of the drawbacks of these methods. In addition, its empirical performance is highly dependent on the *a priori* choice of the initial

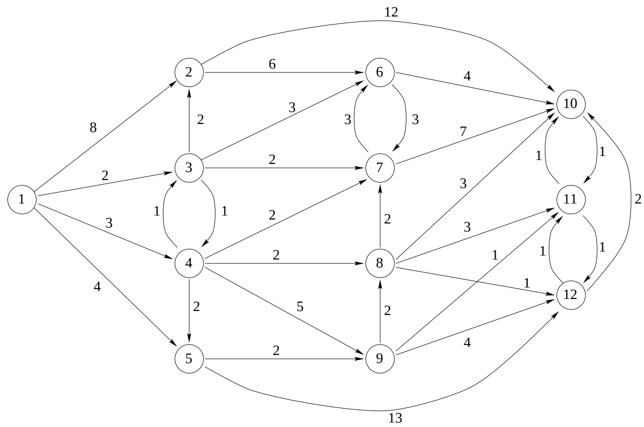

Figure 1: Network Problem

Table 1: Network Data

b_1	U[80,120]	b_5	U[12,18]	b_9	U[10,30]
b_2	U[-25,-5]	b_6	U[-45,-15]	b_{10}	U[-80,-20]
b_3	U[10,30]	b_7	U[-30,-10]	b_{11}	U[-50,-10]
b_4	U[0,20]	b_8	U[0,0]	b_{12}	U[-30,-10]

step size a_0. To be able to compare the sample-path solution with the solution of SA, we considered a problem with only simple bounds on u. It is well known that in the absence of bounds, SA may suffer from unboundedness problems whereas in the presence of other linear inequality constraints, SA experiences difficulties in enforcing feasibility. Therefore we chose a C large enough so that the capacity constraint of problem (P) was inactive. We first solved the problem with the sample-path method and determined the required number of function evaluations, K (this was 50 in our case). Then we allowed the SA algorithm to run for K iterations and generated a sequence of points according to the following rule:

$$u^{k+1} = \Pi_\Theta(u^k - \frac{a_0}{k} g^k)$$

where g^k is an estimate of the (sub)gradient or the directional derivative (whichever is available) at u^k, a_0 is the predetermined step size constant, and $\Pi_\Theta(\cdot)$ is the projection onto the feasible set Θ determined by the bound constraints. In both methods we used a simulation run of length 10,000. To determine the "Optimal" solution we solved the problem using the Sample-Path method and a long simulation run of length 100,000 (the number of iterations required was 41 in this case). The results are reported in Table 2. The "Error" column is the Euclidean distance between the corresponding point and the "Optimal" solution. To make an additional test of quality of these solutions, we randomly generated 20,000 instances of the supply/demand vector b (using the same network topology) and computed the average shortfall (measured by the cost of the network which incorporates the cost of using artificial arcs due to insufficient arc capacity and the cost of assigning the required capacity) and its variance; these are reported under "Average badness" and "Variance of badness" respectively. All the solutions are obtained starting from a vector of ones.

Note that although the SA solution using $a_0 = 30$ is the one most distant from the "Optimal" solution, its variance is much lower. This could be explained as follows. The particular SA solution largely overestimates the true solution which leads to a large cost as shown by the associated objective function value and the average badness; but at the same time it leaves great leeway and hence the cost is not largely affected by the fluctuations in demand and supply. In that sense the measures "Objective function" and "Variance of badness" are conflicting.

Table 2: Solutions generated by SA and the Sample-Path methods

	Final points			"Optimal" solution
	Sample-path solution	SA solution		
		$a_0 = 20$	$a_0 = 30$	
u_1	28.1	25.0	78.1	28.4
u_2	45.0	53.5	60.1	44.5
u_3	36.1	16.4	18.7	36.3
u_4	11.8	25.5	15.8	12.3
u_5	0.0	3.7	0.0	0.0
u_6	19.1	14.7	27.6	19.5
u_7	7.0	9.3	78.1	6.7
u_8	0.0	0.0	74.1	0.0
u_9	45.5	39.3	0.0	45.7
u_{10}	19.5	32.1	68.3	18.7
u_{11}	0.0	3.4	0.0	0.0
u_{12}	0.0	0.8	0.0	0.0
u_{13}	11.0	2.0	0.3	11.3
u_{14}	37.9	19.8	16.4	37.4
u_{15}	4.3	3.1	15.0	3.8
u_{16}	22.3	33.2	22.7	22.9
u_{17}	6.2	7.9	30.8	5.7
u_{18}	0.0	0.0	0.0	0.0
u_{19}	17.2	16.5	23.3	17.4
u_{20}	0.0	3.0	0.0	0.0
u_{21}	7.8	11.6	8.5	7.0
u_{22}	0.0	0.0	0.0	0.0
u_{23}	22.7	11.3	9.8	21.6
u_{24}	7.5	1.3	0.0	7.5
u_{25}	17.5	12.3	10.6	17.7
u_{26}	0.0	5.0	49.5	0.0
u_{27}	42.0	45.1	78.1	42.2
u_{28}	6.4	7.7	45.0	6.8
u_{29}	0.0	0.0	0.0	0.0
u_{30}	19.4	22.0	20.9	19.4
u_{31}	0.0	0.0	0.0	0.0
u_{32}	0.0	1.3	0.0	0.0
u_{33}	0.1	7.4	2.3	0.8
Objective function	7673	7901	9390	7682
"Error"	2.32	41.63	135.20	–
Average badness	3003	3123	4585	3003
Variance of badness	1742E2	4509E2	589E1	1909E2

We also considered problem (P) when $C = 350$ and the capacity constraint was active. In this case, we did not apply the SA method due to the difficulties the method experiences in enforcing feasibility. The results are given in Table 3. The number of function evaluations required is denoted by $nfeval$, and n denotes the different simulation lengths used to compute f and its subgradient at any given u. "Error" is the Euclidean distance of the corresponding solution to the solution given in the last column.

Note that in problem (P), f is a convex but nonsmooth function of u. In principle, we could have used a cost function different from $\sum_{i=1}^{k} d_i u_i$ in the objective function, in (P). However, since we used a nonsmooth convex optimizer, the Bundle-Trust method of Schramm and Zowe (1990), we

Table 3: Sample-Path solutions for the constrained problem

n	10,000	30,000	100,000	500,000
u_1	67.3	67.0	66.6	66.5
u_2	27.1	27.4	28.3	28.2
u_3	13.0	12.9	12.0	12.2
u_4	9.8	9.6	10.0	10.2
u_5	5.7	5.1	5.2	5.1
u_6	49.5	49.5	49.1	48.8
u_7	0.0	0.0	0.0	0.0
u_8	0.0	0.0	0.0	0.0
u_9	37.2	37.9	37.7	37.8
u_{10}	12.9	12.9	13.8	13.7
u_{11}	0.0	0.0	0.0	0.0
u_{12}	0.0	0.0	0.0	0.0
u_{13}	16.4	16.4	15.7	15.5
u_{14}	2.7	2.6	2.8	3.2
u_{15}	7.0	6.8	6.9	6.6
u_{16}	2.9	3.0	3.0	3.0
u_{17}	22.9	22.6	23.1	23.2
u_{18}	0.0	0.0	0.0	0.0
u_{19}	10.1	10.4	10.5	10.5
u_{20}	0.0	0.0	0.0	0.0
u_{21}	7.1	6.9	6.6	6.4
u_{22}	0.0	0.0	0.0	0.0
u_{23}	0.0	0.0	0.0	0.0
u_{24}	2.6	2.6	2.8	3.2
u_{25}	0.1	0.0	0.0	0.0
u_{26}	0.0	0.0	0.0	0.0
u_{27}	26.1	26.4	26.5	26.2
u_{28}	5.6	6.0	5.9	5.8
u_{29}	12.7	12.9	12.3	12.1
u_{30}	3.2	3.1	3.5	3.7
u_{31}	0.0	0.0	0.0	0.0
u_{32}	3.3	3.0	3.2	3.4
u_{33}	4.6	4.8	4.7	4.7
$nfeval$	51	56	50	50
Objective function	8283	8273	8276	8284
"Error"	2.70	2.50	0.93	–

only considered convex cost functions. Similarly, we could easily have added linear equality or inequality constraints to problem (P).

In both problems (i.e. capacity constraint active or inactive), to compute the function value f and its subgradient g at any given vector u of arc capacities, we had to solve a minimum cost network flow (MCNF) problem. For this we used an earlier version (copy number: 361026, version 3.61-12/1979) of the code RNET developed at the Department of Computer Science of Rutgers University by M.D. Grigoriadis and T. Hsu. RNET is a network specialization of the revised simplex method for bounded variables, primarily designed for solving MCNF but it also has the capability of solving other problems such as assignment, transshipment, maximum flow, and shortest path problems, see Grigoriadis and Hsu (1979) for a discussion of the algorithm. We thank M.D. Grigoriadis for providing us a copy of this code for use at the University of Wisconsin-Madison. As output, RNET reports the values of both the primal variables and the node potentials. Using this information we computed the ith component of the subgradient by the following rule, see Ahuja, Magnanti, and Orlin (1993):

$$g_i(u) = -\max\{0, \pi(O_i) - \pi(D_i) - c_i\}$$

where O_i and D_i are the origin and the destination of the ith arc, c_i is the cost of sending flow on the ith arc, and $\pi(j)$ is the potential of node j.

As mentioned earlier, we used the Bundle-Trust method of Schramm and Zowe (1990). We thank Dr. Helga Schramm for providing us her code. The code finds an ϵ-subgradient whose norm is at most ϵ and this controls the number of function evaluations required by the method. In all the problems we used $\epsilon = 1$.

As illustrated in Table 2, the sample-path method outperforms stochastic approximation, even without considering the effort required to find a suitable value for a_0. Since in the second problem, SA was not applicable (without some *ad hoc* techniques to enforce feasibility) we considered different simulation lengths. Table 3 shows that even with a simulation run of length 10,000 the sample-path method produces reasonable results.

5 CONCLUSIONS

In this paper we have shown how to use a variant of sample-path method to solve optimization problems with stochastic objective function and stochastic constraints. We presented a set of sufficient conditions for the convergence of a new form of the method. We also gave a computational example of sample-path optimization applied to a medium-sized network design problem with random supply and demand nodes.

The sufficient conditions outlined here extend the range of application of sample-path methods. We hope that in turn this may help in providing solutions for problems that have been difficult to handle with current techniques.

ACKNOWLEDGMENT

Part of the research of A. Y. Özge reported here was performed at the University of Wisconsin–Madison.

The research reported here was sponsored by the Air Force Office of Scientific Research, Air Force Materiel Command, USAF, under grant numbers F49620-95-1-0222, F49620-97-1-0283, and F49620-98-1-0417. The U. S. Government has certain rights in this material, and is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the sponsoring agency or the U. S. Government.

REFERENCES

Ahuja, K.K., Magnati, T.L., and Orlin, J.B. 1993. *Network flows: theory, algorithms, and applications* (Prentice Hall, Englewood Cliffs, NJ).

Ferris, M. C. and Pang, J. S. 1997. Engineering and economic applications of complementarity problems. *SIAM Review* 39(4): 669–713.

Glasserman, P. 1991. *Gradient estimation via perturbation analysis* (Kluwer, Norwell, MA).

Gomory, R.E. and Hu, T.C. 1964. Synthesis of a communication network. *SIAM Journal* 12(2): 348–369.

Grigoriadis, M.D. and Hsu, T. 1979. The Rutgers minimum cost network flow subroutines. *SIGMAP Bulletin of the ACM* 26: 17–18.

Gürkan, G. and Karaesmen F. 1998. Computation of optimal flow control policies of a manufacturing system with multiple production rates. Working paper.

Gürkan, G., Özge, A. Y. and Robinson, S. M. 1996. Sample-path solution of stochastic variational inequalities, with applications to option pricing. In *Proceedings of the 1996 Winter Simulation Conference,* ed. J. M. Charnes, D. M. Morrice, D. T. Brunner and J. J. Swain, 337–344.

Gürkan, G., Özge, A. Y. and Robinson, S. M. 1998. Sample-path solutions for simulation optimization problems and stochastic variational inequalities. In: *Interfaces in Computer Science and Operations Research,* ed. D. L. Woodruff (Kluwer, Boston), 169–188.

Gürkan, G., Özge, A. Y. and Robinson, S. M. 1999. Sample-path solution of stochastic variational inequalities. *Mathematical Programming* 84: 313–333.

Harker, P. T. and Pang, J.-S. 1990. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming* 48: 161–220.

Ho, Y.-C. and Cao, X.-R. 1991. *Perturbation analysis of discrete event dynamical systems* (Kluwer, Norwell, MA).

Kall, P. 1986. Approximation to optimization problems: an elementary review. *Mathematics of Operations Research* 11: 9–18.

Özge, A. Y. 1997. Sample-path solution of stochastic variational inequalities and simulation optimization problems. Ph.D. Dissertation, Department of Industrial Engineering, University of Wisconsin-Madison, Madison, Wisconsin, USA.

Plambeck, E. L., B.–R. Fu, S. M. Robinson, and R. Suri. 1993. Throughput optimization in tandem production lines via nonsmooth programming. In *Proceedings of 1993 Summer Computer Simulation Conference,* ed. J. Schoen, 70–75. San Diego, CA: Society for Computer Simulation.

Plambeck, E. L., B.–R. Fu, S. M. Robinson, and R. Suri. 1996. Sample-path optimization of convex stochastic performance functions. *Mathematical Programming* 75: 137–176.

Robinson, S. M. 1980. Strongly regular generalized equations. *Mathematics of Operations Research* 5: 43–62.

Robinson, S. M. 1996. Analysis of sample-path optimization. *Mathematics of Operations Research* 21: 513–528.

Rockafellar, R. T. and Wets, R. J-B. 1998. *Variational analysis* (Springer-Verlag, Berlin).

Rubinstein, R. Y. and Shapiro, A. 1993. *Discrete event systems: sensitivity analysis and stochastic optimization by the score function method* (Wiley, Chichester and New York).

Schramm, H., and Zowe, J. 1990. A version of the bundle idea for minimizing a nonsmooth function: Conceptual idea, convergence analysis, numerical results. Technical Report, 209. Mathematisches Institut, Universität Bayreuth, Bayreuth, Germany.

Shapiro, A. 1993. Asymptotic behavior of optimal solutions in stochastic programming. *Mathematics of Operations Research* 18: 829–845.

AUTHOR BIOGRAPHIES

GÜL GÜRKAN is an Assistant Professor of Operations Research and Management Science at CentER for Economic Research, within the Faculty of Economics and Business Administration at Tilburg University, in the Netherlands. She received her Ph.D. in Industrial Engineering from the University of Wisconsin–Madison in 1996. Her research interests include simulation, mathematical programming, stochastic optimization and equilibrium models, with applications in logistics, production, telecommunications, economics, and finance. She is a member of INFORMS.

A. YONCA ÖZGE received her Ph.D. in Industrial Engineering from the University of Wisconsin-Madison in 1997. Since then she has been working at GE Corporate Research & Development Center as operations researcher. Her research interests include analysis of complex stochastic systems by combining simulation and optimization with applications in production systems, option pricing, network design, and economic equilibrium problems. She is a member of INFORMS.

STEPHEN M. ROBINSON is Professor of Industrial Engineering and Computer Sciences at the University of Wisconsin–Madison. His research interests are in quantitative methods in managerial economics, especially deterministic and stochastic optimization and methods to support decision under uncertainty. He is a member of INFORMS, the Mathematical Programming Society, SIAM, and IIE.