TRANSPORTATION RESEARCH

RECORD

No. 1620

Marine Transportation

Ports/Waterways Marine Transportation

A peer-reviewed publication of the Transportation Research Board

TRANSPORTATION RESEARCH BOARD
NATIONAL RESEARCH COUNCIL

NATIONAL ACADEMY PRESS
WASHINGTON, D.C. 1998

Optimization Through Simulation of Waterway Transportation Investments

CHING-JUNG TING AND PAUL SCHONFELD

The cost of tow delays is a serious problem in a waterway network. One way to reduce the delay cost is to increase capacity at waterway locks. Planners must determine how much additional capacity to provide at particular lock sites and when to implement the capacity expansion projects. Answers for such project sizing and timing problems are difficult to obtain analytically. The use of a new approach for optimizing through simulation, called simultaneous perturbation stochastic approximation (SPSA), is investigated. This approach, which seeks optimal values for all decision variables after each pair of simulation runs, is quite promising for optimizing large problems relatively fast. A small numerical example tests how this simulation and optimization algorithm may be used to optimize lock capacities and implementation times.

The inland waterway network includes ports and locks whose effects on traffic are often interdependent. Traffic departing from a port or lock affects the arrival distributions, and hence delays, at downstream ports and locks. Tow delays at locks are a major cost of inland waterway transportation, especially at congested locks (*1*). Ting and Schonfeld (*2*) showed how different traffic control policies at locks can reduce delays. Still, capacity expansion may often be the most cost-effective way to reduce delays, especially if traffic volumes approach the capacity of locks. Planners must decide not only how much capacity to add at given lock sites but also when to do it. The problem of optimizing both the size and time of capacity improvement projects is quite difficult to solve with analytical models, especially when the projects are interdependent. Thus, an efficient method of optimizing directly from simulation models is very valuable.

An inland waterway network is a complex stochastic system. Some decisions about it cannot be adequately modeled using analytical methods (i.e., functional derivations) but can be simulated. The objective function used to evaluate and optimize the system (which in this study is the sum of supplier and user costs) may be impossible to express as an explicit function of the controllable parameters. It may involve some response of the system that can be found only by running a simulation model or, in some cases, by observing the actual system.

The purpose of this paper is to introduce and describe a simultaneous perturbation stochastic approximation (SPSA)-based method for optimizing the selection, sizing, and sequencing of lock improvement projects. The next section (whose mathematics are not essential for reading the rest of this paper) briefly outlines the form and characteristics of the SPSA algorithm. A more detailed description of the algorithm may be found in Spall (*3*). The third section presents the assumptions in the simulation models. Three hypothesized examples are presented in the fourth section, before a summary of this paper.

Department of Civil Engineering, University of Maryland, College Park, MD 20742.

SIMULTANEOUS PERTURBATION STOCHASTIC APPROXIMATION (SPSA)

The Simultaneous Perturbation Stochastic Approximation (SPSA) algorithm was introduced and developed by Spall (*3*). SPSA is an iterative technique for finding local optimizers of linear or nonlinear objective functions from many types of systems. In each iteration, SPSA computes the positively and negatively perturbed objective function values. SPSA is like other Kiefer and Wolfowitz (*4*) stochastic approximation algorithms, such as finite difference stochastic approximation (FDSA), in that SPSA requires only measurements (possibly noisy) of an objective function to form gradient estimates and converge to a local optimum. However, SPSA differs significantly from FDSA in requiring only two objective function evaluations per gradient estimate, whereas FDSA requires $2p$ evaluations, where p is the number of system parameters being estimated. This gives SPSA a significant advantage in high-dimensional problems, especially when evaluating the objective function is expensive or time-consuming (*3*).

The algorithm uses objective function measurements to iteratively update system control parameters until parameter values are reached that locally optimize the objective function. Specifically, let $x \in R^p$ be a vector whose components represent system parameters to be controlled, for example, the capacity expansion ratios and implementation timing in waterway investments. Suppose $L(x)$ represents the objective function to be optimized. The goal is to find a root x^* of the gradient of this objective function. That is, x should be such that

$$g(x) \equiv \frac{\partial L(x)}{\partial x} = 0 \tag{1}$$

The SPSA algorithm attempts to find a minimizer x^* by starting at a fixed $\hat{x}_0$. Then a sequence of searches is done such that $\hat{x}_{k+1}$, the value of x in the $(k + 1)$th iteration, depends on the value of $\hat{x}_k$ according to the following scheme:

$$\hat{x}_{k+1} = \hat{x}_k - a_k \hat{g}_k(\hat{x}_k) \tag{2}$$

Here $\{a_k\}$ is a gain sequence of positive scalars satisfying certain conditions (in particular, $a_k \rightarrow 0$ and $\sum_{k=1}^{\infty} a_k = \infty$), and $\hat{g}_k$ is an estimate of the gradient g whose 1-th component is defined as

$$\hat{g}_{kl} = \frac{y_k^+ - y_k^-}{2c_k\Delta_{kl}} \tag{3}$$

Here, y_k^+ represents a measurement of $L(\hat{x} + c_k \Delta_k)$, and y_k^- is a similar measurement of $L(\hat{x} - c_k\Delta_k)$. The sequence $\{c_k\}$ is a sequence of positive scalars such that $c_k \to 0$, and $\Delta_k \in R^p$ is a vector of p mutually independent random variables satisfying conditions in Equation 3. For example, the components of Δ_k could be independent Bernoulli (± 1) distributed random variables.

Observe that the numerator in Equation 3 is the same for each component of g_k. Thus, only two measurements of the objective function are required to obtain the SPSA gradient estimate at each iteration. To illustrate, if $\hat{x}_k$ represents the current estimate of the best control policy x^*, then $\hat{x}_k + c_k\Delta_k$ and $\hat{x}_k - c_k\Delta_k$ are "perturbed" policies. Then y_k^+ and y_k^- could be total system cost values obtained from performing two simulations of the waterway network, one using the control policy $\hat{x}_k + c_k\Delta_k$ and one using $\hat{x}_k - c_k\Delta_k$. These values would then used in Equations 2 and 3 to obtain a new estimate $\hat{x}_{k+1}$ of the best control policy x^*. This process would be repeated until the objective function (total system cost) cannot be significantly improved any further. The general process is stated below and shown in Figure 1:

Step 0: Make initial "guess" at x ($= \hat{x}_k$).

Step 1: From current value of $\hat{x}_k$, change all elements of $\hat{x}_k$ simultaneously by a small (random) amount $c_k\Delta_k$ using SPSA guidelines. Collect measurement $L(\hat{x} + c_k\Delta_k)$.

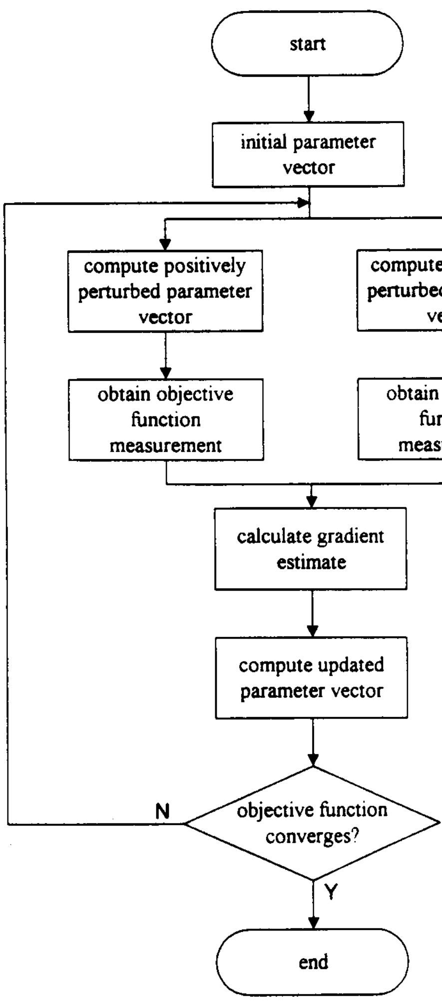

FIGURE 1 The SPSA algorithm.

Step 2: From same $\hat{x}_k$, change all elements in the opposite way according to SPSA. Collect measurement $L(\hat{x} - c_k\Delta_k)$.

Step 3: With information from Steps 1 and 2, estimate the gradient according to Equation 3 and update to new x value ($\hat{x}_{k+1}$) by recursion Equation 2.

Step 4: Repeat Steps 1–3 until the allowable number of iterations has been reached or x is effectively optimized (i.e., there is negligible change in the iterates for several successive iterations).

Note that the SPSA algorithm is very general and can be applied in many different situations to optimize many different kinds of objective functions. For instance, x could represent the optimal capacities at different locks in a waterway network for steady flow capacity optimization, while $L(x)$ could be the total system cost. The constraints can be imposed by adding penalty functions in the objective. The flexibility of the algorithm stems from the fact that only objective function measurements are required, instead of full objective function or gradient information.

The objective function measurements required by the SPSA algorithm can come from a real system as well as from a computer simulation of a real world probabilistic system, depending on the purpose of the optimization. Using SPSA to directly optimize a real system may work well for short-term traffic control but is impractical if the variables being optimized are characteristics of expensive constructed facilities.

BASIC ASSUMPTIONS IN SIMULATION

The simulation model developed is based on the following assumptions:

1. The service time may be estimated as a function of the number of cuts (which are the sets of barges into which tows are subdivided for passage through locks) and the number of barges in a tow, that is,

$$s_i = \alpha + \beta b_i + \gamma(\text{cut}) \tag{4}$$

where

b_i = number of barges of tow i,
cut = 1, if tow i needs more than 1 cut; = 0, otherwise,
s_i = service time of tow i, and
α, β, γ = statistically estimated parameters of the service time function

2. Because the processing time for tows often depends on the direction of the preceding tow(s) in that chamber, a directional constant service time is added when a tow's direction is that of the previous tow. The constant service time does not consider the tow size and lock congestion level. It varies by lock and chamber and can be positive or negative. If tows following in the same direction have a short entry time, the constant service time is negative. Otherwise, the constant is positive.

3. Each tow passes through a lock separately from other tows according to a first-come, first-served discipline.

4. Tows generated for each origin/destination are exponentially distributed.

5. Tow's speed is normally distributed with known mean and standard deviation.

6. Tows may pass other tows between locks.
7. The travel speeds between locks and link capacities are not limited or otherwise affected by lock capacity changes.
8. The lock capacity increases instantaneously after a lock improvement project is implemented. A lock capacity increase means the service rate increases and, thus, the service time decreases. For instance, if the capacity increases by 50 percent (i.e., to 150 percent of its previous level), the capacity expansion ratio is 1.5 and the original service time per tow is divided by 1.5.

NUMERICAL EXAMPLES

The numerical examples described in this section demonstrate an application of the SPSA optimization approach. The gain sequences $\{a_k\}$ and $\{c_k\}$ from Equations 2 and 3 are defined as $a_k = a(k + 1 + 10)^{-\delta}$ and $c_k = c(k + 1)^{-\theta}$ with $\delta = 0.602$, and $\theta = 0.101$, as recommended by Spall (*3*). Figure 2 shows the tested inland waterway network that includes five ports and five locks having initially the same capacity. Each of the five locks has one chamber with the same dimensions, 33.528 × 182.88 m (110 × 600 feet). The service time models are based on 1987 data from the Mississippi River Lock 25 from the Lock Performance Monitoring System (LPMS) which is documented by Fleming and Goodwin (*5*).

The regression models for lock service times are estimated with the SAS (*6*) package. The models used are from Ting and Schonfeld (*2*):

$$\text{dir 1:} \quad s_i = \underset{(0.64)}{18.342} + \underset{(0.19)}{1.180} \times b_i + \underset{(2.73)}{57.343} \times \text{cut} \quad R^2 = 0.860 \tag{5}$$

$$\text{dir 2:} \quad s_i = \underset{(0.64)}{17.126} + \underset{(0.16)}{2.633} \times b_i + \underset{(2.14)}{51.762} \times \text{cut} \quad R^2 = 0.901 \tag{6}$$

For Equations 5 and 6, standard errors are shown in parentheses and R^2 is the coefficient of determination that indicates how well the estimated model accounts for the variation in the dependent variable.

Locks 2, 3, and 4 have higher arrival volumes than the other two locks. Given their equal capacities, the delays at these three locks are higher. The investment projects should provide optimal capacities at these three locks on the basis of different flow assumptions. The objective is to minimize the total system cost per hour, which includes supplier cost and user cost (delay and travel costs). The supplier cost includes a fixed cost and a variable cost that is proportional to the capacity

$$C_s^i = C_f + C_v \times \mu_i \tag{7}$$

where

C_s^i = supplier cost of lock i (\$/yr),
C_f = fixed cost (\$/yr),
C_v = variable cost (\$/yr), and
μ_i = capacity expansion ratio of lock i.

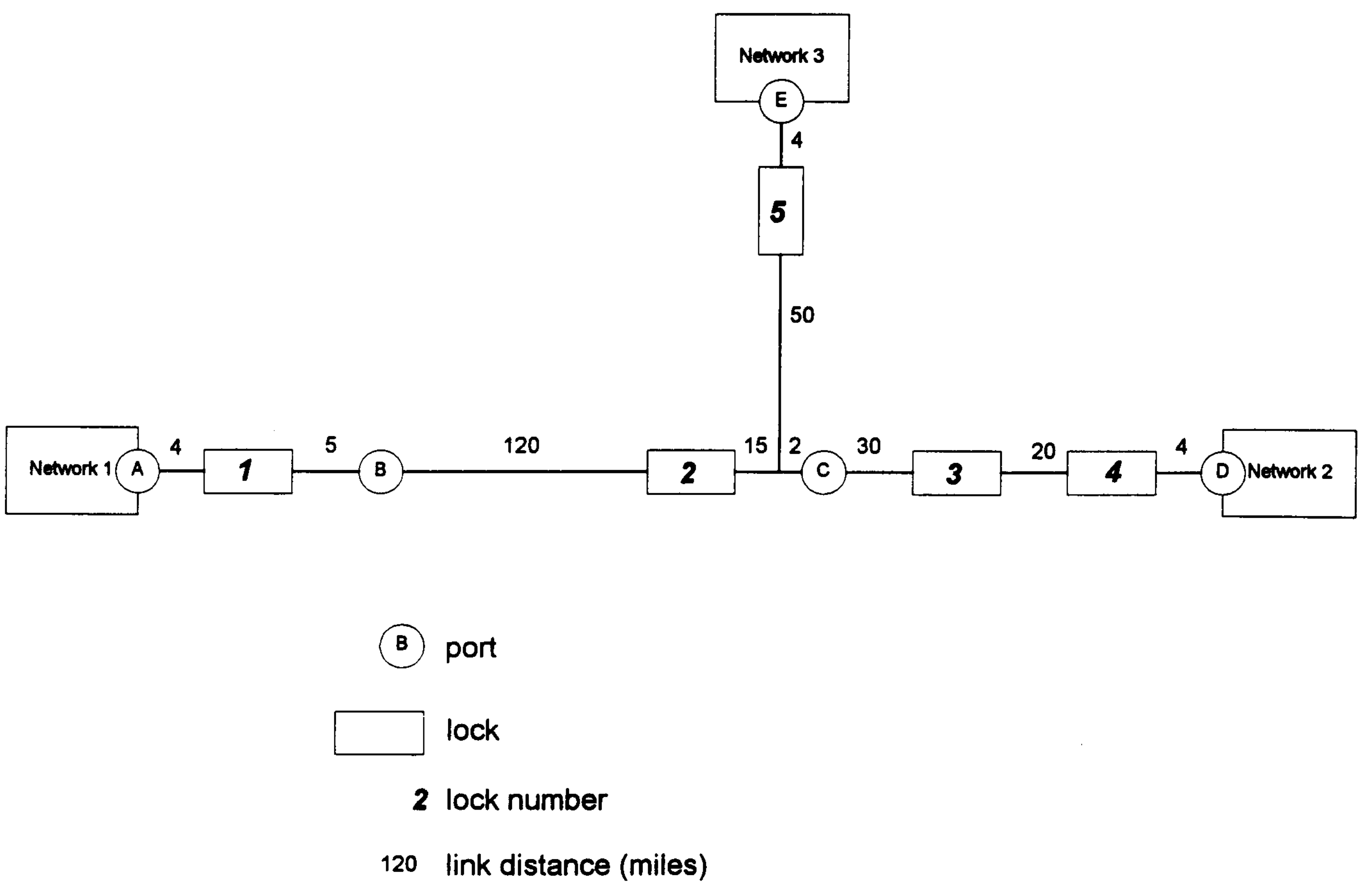

FIGURE 2 Tested waterway network.

TABLE 1 Origin/Destination Flows for Case 1 (tows/day)

	A	B	C	D	E
A	0.0	0.5	1.0	2.5	2.0
B	0.5	0.0	0.5	1.0	1.0
C	1.0	0.5	0.0	3.5	1.0
D	2.5	1.0	3.5	0.0	2.0
E	2.0	1.0	1.0	2.0	0.0

The user cost includes moving cost and delay cost, which are each assumed to be \$3,000 per tow per hr.

Three cases are tested for the same network with different origin/destination volumes. Each case simulates 3 years of traffic in the network and computes the total system cost per hour. A common random number is used for all the simulation iterations, which makes the objective function deterministic. Each iteration includes two simulation runs (one is for the positively perturbed parameter, and the other one for the negatively perturbed parameter). Case 1 assumes a steady-state flow between each pair of origin/destination (O/D) ports and optimizes the capacities individually for each lock. Cases 2 and 3 assume that the traffic flow increases continuously. Case 2 optimizes the implementation times at each lock of preset capacity expansions. Case 3 jointly optimizes the capacities and implementation times.

Case 1: Optimization of Capacities

This case assumes that the traffic flow is steady all year long for each O/D pair. The objective is to optimize the capacities for Locks 2, 3, and 4 at the beginning of simulation to minimize the total system cost. The O/D flow is shown in Table 1.

Starting at a feasible initial solution (i.e., starting point) $x = (1.5, 1.5, 1.5)$, Figure 3 shows the total system cost changes through successive iterations for same random seed. The total cost fluctuates in the first few iterations. It stays very stable after around 40 iterations. Figure 4 shows the optimal capacity ratios for Locks 2, 3, and 4. The capacity ratios of Locks 3 and 4 are similar because they are relatively closely spaced (20 miles between them). The optimal capacity expansion ratios after 200 iterations are 1.729 for Lock 2, 1.844 for Lock 3, and 1.846 for Lock 4. Table 2 shows the results for five different random seeds with the same starting point. The likely reason for the small differences is that the objective function curve is very flat near the optimum.

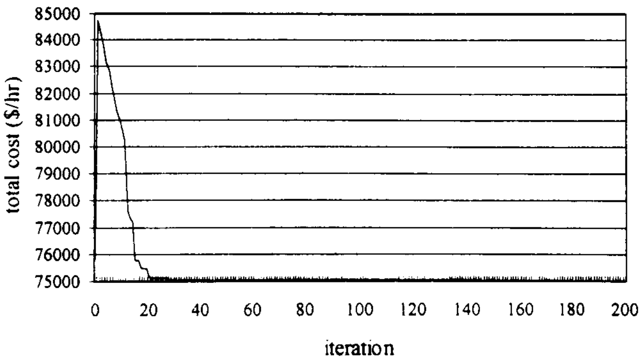

FIGURE 3 Objective function value for steady flow capacity optimization.

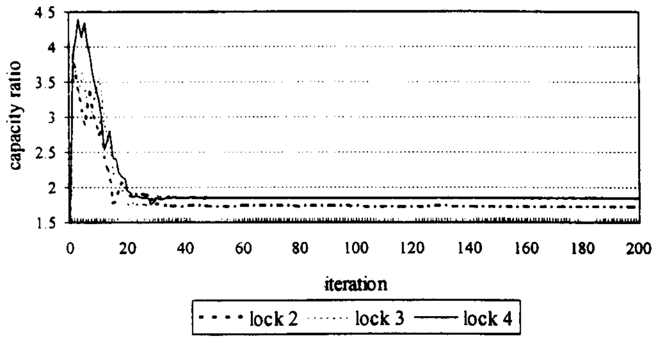

FIGURE 4 Optimal capacity ratios at locks 2, 3, and 4 for steady flow capacity optimization.

Case 2: Optimization of Implementation Times for a Given Capacity Expansion Ratio

In this case, the O/D flow starts as half of Case 1, as shown in Table 3, and grows over time according to Equation 8:

$$\lambda_t = \lambda_0 G_t = \lambda_0 (1 + r)^t \tag{8}$$

where

λ_0 = O/D flow at time 0,
λ_t = O/D flow at time t,
G_t = growth factor at time t,
r = growth rate, and
t = time.

To reduce the simulation time here, it is assumed that the flow will grow quickly, reaching the same level as in Case 1 by end of the third year. Based on Equation 8, $\lambda_3 = 2\lambda_0 = \lambda_0 (1 + r)^3$, $r = 2^{1/3} - 1 = 0.2599$ Thus, the annual growth rate is almost 26 percent.

Figure 5 shows the objective function value of optimizing the implementation time for preset capacity ratio 1.5 (i.e., 50 percent capacity expansion). The optimal implementation times for locks 2, 3, and 4 are shown in Figure 6. Locks 3 and 4 should be expanded at slightly different times even though they are near each other and their traffic characteristics are the same. The probable reason is that the arrival distributions at these two locks are somewhat different. The final optimized values for implementation times with three starting points are shown in Table 4. After 200 iterations, the solutions converge to the optimum even though the starting points are different.

Table 5 shows the results with same starting points but different gain sequence values $\{c_k\}$ ($c = 0.5, 1.0, 1.5$). After 200 iterations, the implementation times are quite different. This indicates that 200 iterations may not achieve the convergence or that SPSA will

TABLE 2 Results of Optimal Capacities for Different Random Numbers

Random Number	lock 2	lock 3	lock 4
RN1	1.729	1.844	1.846
RN2	1.729	1.842	1.837
RN3	1.728	1.849	1.841
RN4	1.715	1.859	1.840
RN5	1.711	1.835	1.825

TABLE 3 Origin/Destination Flows for Cases 2 and 3 (tows/day)

	A	B	C	D	E
A	0.00	0.25	0.50	1.25	1.00
B	0.25	0.00	0.25	0.50	0.50
C	0.50	0.25	0.00	1.75	0.50
D	1.25	0.50	1.75	0.00	1.00
E	1.00	0.50	0.50	1.00	0.00

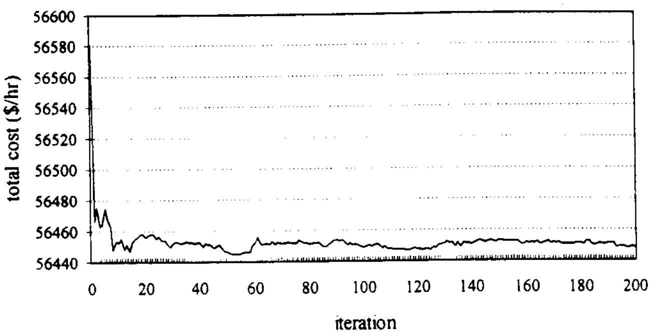

FIGURE 5 Objective function value for optimizing implementation times of 50% capacity expansions at locks 2, 3, and 4.

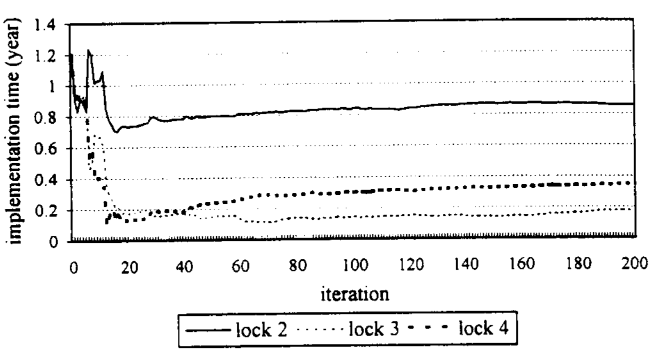

FIGURE 6 Optimal implementation times for pre-set capacities at locks 2, 3, and 4.

TABLE 4 Results of Optimal Implementation Times with Different Starting Points

starting point	lock 2	lock 3	lock 4
(0.5,0.5,0.5)	0.838	0.297	0.353
(1.2,1.2,1.2)	0.838	0.298	0.353
(1.5,1.5,1.5)	0.836	0.298	0.353

TABLE 5 Results with the Same Random Number But Different Gain Sequences $\{c_k\}$

c	lock 2	lock 3	lock 4
0.5	0.858	0.427	0.482
1.0	0.838	0.298	0.353
1.5	0.703	0.052	0.095

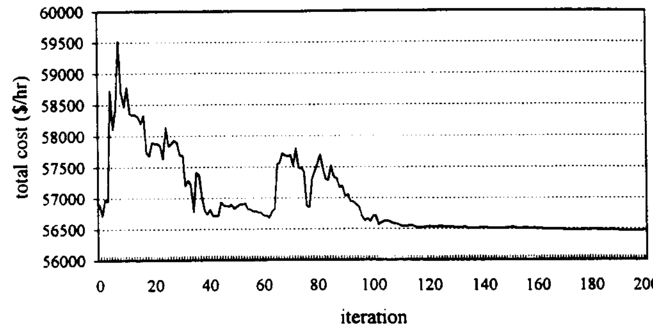

FIGURE 7 Objective function value for jointly optimizing capacities and implementation times.

converge to a different local optimum if the gain sequence values are chosen differently. However, these random samples show substantial improvement from the initial values.

Case 3: Joint Optimization of Capacities and Implementation Times

With the same traffic flows as in Case 2, the capacities and implementation times are jointly optimized. Figure 7 shows the objective function value of jointly optimizing implementation times and capacities. The total cost is comparatively stable after 160 iterations. The optimal capacity ratios and implementation times for each lock are presented in Figures 8 and 9, respectively. Figure 8 shows that capacity ratios for Locks 3 (1.63) and 4 (1.626) are close, as expected. The implementation times for Locks 3 (0.866) and 4 (0.897) are closer than those (0.179 for Lock 3 and 0.348 for Lock 4) shown in Figure 6, in which the capacity ratio is the same (1.5) for both locks. The difference between Cases 2 and 3 shows the interdependence of capacity ratios and implementation times. For a given planning period, a larger capacity expansion implies a later implementation.

SUMMARY

The technique of simultaneous perturbation stochastic approximation (SPSA) has been applied to the optimization of improvement projects for an inland waterway network. The technique offers

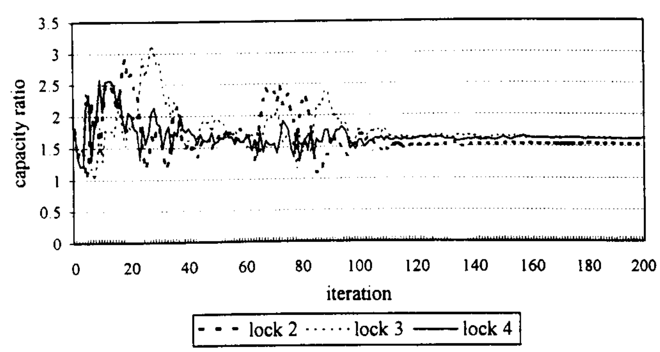

FIGURE 8 Optimal capacity ratios at locks 2, 3, and 4 when jointly optimizing capacities and implementation times.

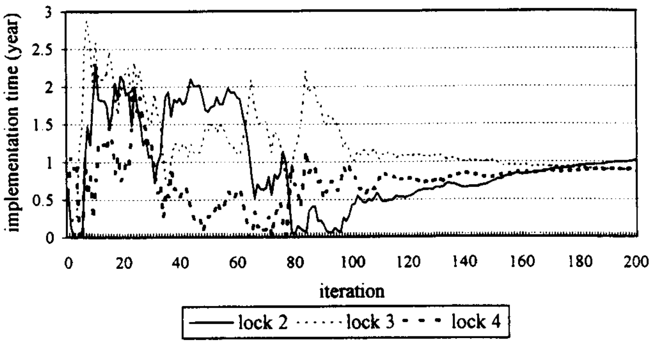

FIGURE 9 Optimal implementation times at locks 2, 3, and 4 when jointly optimizing capacities and implementation times.

significant computational savings over traditional finite-difference methods. The authors considered three different cases and conducted experiments to investigate the viability of the technique. These simulation results, though not comprehensive, are quite promising and should help encourage further experimental work on the application of SPSA to the optimization of dynamic systems. One direction for further study would be to use this optimization method to optimize the maintenance scheduling problem. Another direction of study would be to consider other control strategies and other objectives such as minimizing fuel consumption.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support received from the Institute for Water Resources of the Corps of Engineers, especially from Robert Pietrowsky, and Shilpa Patel. The authors also wish to thank Michael Fu of the College of Business and Management at the University of Maryland for his advice on the SPSA algorithm and the anonymous TRB reviewers for their constructive comments.

REFERENCES

1. Chien, S. I. J., and P. Schonfeld. Effects of Lock Interdependence on Tow Delays. *Compendium on Waterway Transportation Reliability: Lock Congestion and Lock Queues, IWR Report 93-R-9,* April, 1993, pp. 21–50.
2. Ting, C. J., and P. Schonfeld. Effects of Tow Sequencing on Capacity and Delay at a Waterway Lock. *Journal of Waterway, Port, Coastal, and Ocean Engineering,* Vol. 122, No. 1, 1996, pp. 16–26.
3. Spall, J. C. Multivariate Stochastic Approximation Using a Simultaneous Perturbation Gradient Approximation. *IEEE Transactions on Automatic Control,* Vol. 37, No. 3, 1992, pp. 332–341.
4. Kiefer, J., and J. Wolfowitz. Stochastic Estimation of a Regression Function. *Annals of Mathematical Statistics,* Vol. 23, 1952, pp. 462–466.
5. Fleming, M., D. Wood, and R. Goodwin. *Lock Performance Monitoring System User's Manual for Data Analysis.* Institute for Water Resources, U.S. Army Corps of Engineers, Fort Belvoir, Va., 1985.
6. *SAS/STAT User's Guide Version 6* 4th ed. Vols. 1 and 2, SAS Institute, Cary, N.C., 1990.

Publication of this paper sponsored by Committee on Inland Water Transportation.