International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

The Stochastic Quasi-Gradient Method Applied to a Facility Location Problem

Ermoliev, Y.M., Leonardi, G. and Vira, J.

IIASA Working Paper

WP-81-014

February 1981

THE STOCHASTIC QUASI-GRADIENT METHOD APPLIED TO A FACILITY LOCATION PROBLEM

Y.M. Ermoliev, G. Leonardi,
and J. Vira

February 1981
WP-81-14

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

THE AUTHORS

Dr. Giorgio Leonardi has been at IIASA since October 1979 on leave from the Polytechnical Institute of Turin, Italy. He has previously been a researcher at the Italian Association for Housing Research (AIRE) and a consultant on urban and regional planning for the Regional Authority of Piemonte.

Prof. Yuri Ermoliev has been at IIASA since December 1979 on leave from the Institute of Cybernetics of Kiev, USSR. He was previously head of the Department of Mathematical Methods of Operation Research at the Institute of Cybernetics.

Juhani E. Vira was a participant in the young scientists summer program at IIASA in 1980. He came from the Technical Research Centre of Finland where he is a research scientist in the Nuclear Engineering Laboratory.

FOREWORD

The public provision of urban facilities and services often takes the form of a few central supply points serving a large number of spatially dispersed demand points: for example, hospitals, schools, libraries, and emergency services such as fire and police. A fundamental characteristic of such systems is the spatial separation between suppliers and consumers. No market signals exist to identify efficient and inefficient geographical arrangements, thus the location problem is one that arises in both East and West, in planned and in market economies.

This problem is being studied at IIASA by the Public Facility Location Task which started in 1979. The expected results of this Task are a comprehensive state-of-the-art survey of current theories and applications, an established network of international contacts among scholars and institutions in different countries, a framework for comparison, unification, and generalization of existing approaches, as well as the formulation of new problems and approaches in the field of optimal location theory.

This paper is an outcome of an interaction between the Human Settlements and Services Area and the Systems and Decision Sciences Area. Its main aim is to test several numerical procedures for solving a class of stochastic programming problems using data on high school location in Turin, Italy. It is a sequel to an earlier theoretical working paper (WP-80-176) on the same subject.

Although the test problem is highly simplified, the results obtained encourage the development of further generalizations that can better exploit the potential use of this stochastic programming method.

A list of related IIASA publications appears at the end of this paper.

Andrzej P. Wierzbicki
Chairman
System and Decision
Sciences Area

Andrei Rogers
Chairman
Human Settlements
and Services Area

ABSTRACT

This paper explores the computational aspects of using the stochastic quasi-gradient method (SQG) to solve some facility location problems. The problems addressed belong to a general class of resource allocation problems with random demand. An algorithm is first developed for the simplest formulation, where a convex objective function is minimized, and results are shown for the location of high schools in Turin, Italy.

Fixed charges are then introduced in the objective function, giving rise to a non-convex problem possessing many local minima, and some numerical results for the same case study are reported.

CONTENTS

THE STOCHASTIC QUASI-GRADIENT METHOD APPLIED TO A FACILITY LOCATION PROBLEM

Y.M. Ermoliev, G. Leonardi,
and J. Vira

1. INTRODUCTION

The data on the location of high schools have already been used as a test problem for some deterministic location techniques (Erlenkotter and Leonardi, forthcoming; and Leonardi and Bertuglia, 1981). However, it has been recognized in Ermoliev and Leonardi (1980) that more realism is captured when random features are introduced. Among the possible types of randomness considered, uncertainty in the customers choice behavior is of special interest. When assignment of customers to facilities is not normatively predetermined, and freedom of choice is allowed, then the number of customers attracted to each facility may be subject to random fluctuations. The difficulty, therefore, is to find those locations and sizes for facilities that in some sense minimize the cost of such fluctuations.

This problem belongs to the following more general class of stochastic programming problems:

$$\min F(X) = E\, f(X,\Omega) \tag{1}$$

subject to the constraints

$$G_i(X) = E\{g_i(X,\Omega)\} \geq 0 \quad , \qquad i = 1,\ldots,m \tag{2}$$

where E denotes the mathematical expectation, X is a vector of decision variables, Ω is a vector of random parameters and $f(X,\Omega)$, $g_i(X,\Omega)$, $i = 1,\ldots m$ are known functions.

In this paper the computational feasibility of the so-called stochastic quasi-gradient method is discussed and applied to a special, simple form problem (1)-(2). The procedure is based on moving iteratively to the direction determined by an estimate of the generalized gradient of the objective function. Under rather general conditions the method has been proved to converge to the solution of the stochastic programming problem.

First the computation procedure is described with emphasis on its practical applicability. At this point some methods are presented by which the computation time can often be significantly reduced. Then some practical results are presented for a stochastic test problem, which deals with optimal sizes of school facilities. Real data from Turin, Italy, have been used in the tests, and the results are compared to those obtained by other methods. Finally some minimization results are reported from tests where the objective function is not even continuous.

2. PROBLEM FORMULATION

A simple model of optimal resource allocation can be stated as follows (Ermoliev and Leonardi, 1980):

Find a vector $X = (x_1,x_2,\ldots,x_n)$ that will minimize the function

$$F(X) = \sum_{j=1}^{n} E\{f_j(x_j,\omega_j\} \tag{3}$$

in the special case

$$F(X) = \sum_{j=1}^{n} E\{\max[\alpha_j(x_j - \omega_j),\ \beta_j(\omega_j - x_j)]\} \qquad (4)$$

$$j = 1,\ldots,n$$

subject to some constraints $0 \leq x_j \leq s_j$. Ω is a random vector and α_j and β_j are given nonnegative parameters. If the probability distribution function for ω_j is $H_j(\omega_j)$, the problem is then to find the minimum of

$$F(X) = \sum_{j=1}^{n} [\alpha_j \int_0^{x_j} (x_j - \omega_j)\, dH_j(\omega_j) + \beta_j \int_{x_j}^{\infty} (\omega_j - x_j)\, dH_j(\omega_j)] \qquad (5)$$

as $0 \leq x_j \leq s_j$, $j = 1,\ldots,n$. In the special case where $F(X)$ has continuous derivatives, the minimization of $F(X)$ by analytical means would lead to the consideration of the partial derivatives

$$\frac{\partial}{\partial x_1} F(X) = \alpha_j \int_0^{x_j} dH_j(\omega_j) - \beta_j \int_{x_j}^{\infty} dH_j(\omega_j) \qquad (6)$$

The solution would then require the determination of $X = (x_1,\ldots x_n)$, such that

$$H_j(x_j) = \frac{\beta_j}{\alpha_j}[1 - H_j(x_j)] \quad , \qquad j = 1,\ldots n \qquad (7)$$

In general this equation may not be solvable by anaytical means. Usually, however, the solution can be easily approximated. In particular, if $\alpha_j = \beta_j$, then the problem becomes finding the medians for the distribution functions. If, however, only observations of the random vector Ω can be made available while the distribution function itself is unknown, the solutions based on equations (6)-(7) are not feasible.

The practical problem that leads to the minimization of an equation (4) type function is common in operations research. For example it can be understood as a facility allocation problem or as a storage inventory control problem where some capacities have to meet random demand and both surpluses and deficits cause penalty costs. In this study the test problem consisted of determining the optimal size of school facilities using data from Turin, Italy. Under certain assumptions the objective function can be stated in the form of equation (4).

3. STOCHASTIC MINIMIZATION

This experimental work concentrated on testing the practicability of the stochastic quasi-gradient method applied to the minimization problem outlined above. The algorithm can be presented as follows (Ermoliev, 1976 and 1978):

(1) Choose an initial approximation X^0.

(2) For s = 0,1... compute successively

$$X^{s+1} = \Pi[X^s - \rho^s H^s] \tag{8}$$

where H^s is the estimate for the generalized gradient $\hat{F}_x(X^s)$ of the function F(X) at X^s such that

$$E\{H^s \mid X^0,\ldots,X^s\} = \hat{F}_x(X^s) \tag{9}$$

and Π is the projection to the feasible set; ρ^s are some step multipliers.

In practice, after the initial values have been chosen, a sequence Ω^s of random deviates is generated. Each random value is then used to determine the current estimate for the generalized gradient of the test problem. In our test problem the estimates

H^s are defined simply by (Ermoliev and Nurminski, 1980)

$$h_j^s = \begin{cases} \alpha_j & \text{if } x_j^s > \omega_j^s \\ -\beta_j & \text{if } x_j^s \le \omega_j^s \end{cases} \tag{10}$$

The execution of the recursion loop should not pose any difficulties nor use much computer time. However, as with the gradient methods in deterministic nonlinear optimization problems, the manner of choosing the step multipliers is crucial to the speed of convergence. In principle, the convergence will be obtained if the step multipliers $\rho^s (s = 0,1,...)$ are chosen so that (Ermoliev, 1976)

$$(1) \quad \sum_{s=0}^{\infty} \rho^s = \infty \tag{11a}$$

$$(2) \quad \rho^s \to 0 \quad \text{as} \quad s \to \infty \tag{11b}$$

$$(3) \quad \sum_{s=0}^{\infty} (\rho^s)^2 < \infty \tag{11c}$$

For the practical construction of the step-size control equations (11a,b,c) are of small importance.

4. PRACTICAL COMPUTATIONS

4.1 Basic Computation Procedure

The methods of controlling the step size in stochastic minimization are usually based on keeping the step multiplier constant during a number of iterations and then reducing it according to certain rules. In the course of the iterations a succession of the function values $F_s = \Sigma_j f_j(x_j^s, \omega_j^s)$ is observed.

Usually these values vary over a wide range. However, the sequence

$$E_k = \frac{1}{k} \sum_{s=0}^{k} F_s = \frac{1}{k} \sum_{s=0}^{k} \sum_{j=1}^{n} f_j(x_j^s, \omega_j^s) \qquad (12)$$

shows smoother behavior as can be seen in Figure 1. Indeed, E_k could be expected to approach a stationary value. One rule of controlling the step size is based on this fact. The method can be summarized as follows:

(1) Choose the initial value ρ^0 for the step multiplier

(2) Using ρ^0 for the step multiplier calculate the value of E_k according to equation (12)

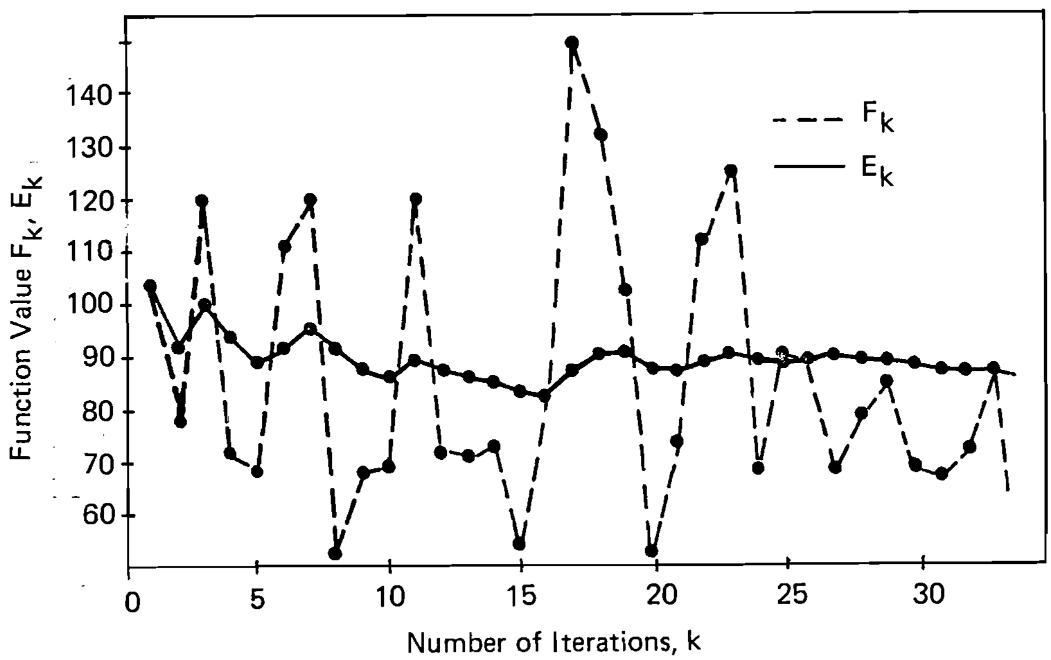

Figure 1. The behavior of the sequences $\{F_k\}$ and $\{E_k\}$ as a function of the iteration number.

(3) When a stationary sequence $\{E_k\}$ is observed, reduce the step multiplier by one half

(4) Go back to step (2) until no improvement in the test function E_k is observed.

There are some unanswered questions in the procedure outlined above. First, how should the initial step multiplier be chosen? If it is too large, both the sequence $\{E_k\}$ and the iterates x^s will oscillate heavily and no decrease in the objective function will be observed. If the initial step multiplier is too small, the rate of decrease will be very small and perhaps hardly noticeable. From the computational point of view the latter situation is more harmful and should be avoided, while the situation arising from too large a step multiplier is rapidly recognized and hence can be corrected. As a rule of thumb the initial step should be chosen to satisfy

$$\rho h_j \approx r\bar{x}_j \tag{13}$$

where $r \in (0,1)$ and $\bar{x}_j$ is the estimated value for the j^{th} component of the solution.

The use of step (3) also needs further explanations. The ideal way of controlling the procedure would be an on-line code, where the program continuously plots the values of the sequence $\{E_k\}$ on the screen and where the iterations could be manually interrupted to cut down the step multiplier. This is not always possible and the iterations must be performed in small batches, whereafter the values of E_k are plotted and possible adjustments of the step multiplier can take place. A definite way to find the stationary phase of the sequence is to rescale the coordinate axes before plotting the values of a new batch. In this case the stationary phase is in fact recognized as smooth oscillations around a fixed value.

Figure 2 shows an example of the behavior of E_k as a function of the iteration number k. The values for coefficients are $\alpha_j = \beta_j = 1.00$, $j = 1,\ldots,23$, $\rho^0 = 1.00$, and the components of the initial estimate and the solution are known to differ at most by five units. Note that the rate of decrease of the sequence $\{E_k\}$ is fast during the first iteration batches but becomes slower as the step size decreases. Hence a crude estimate of the result is obtained after a rather small number of iterations, but for greater accuracies the number of iterations needed grows rapidly.

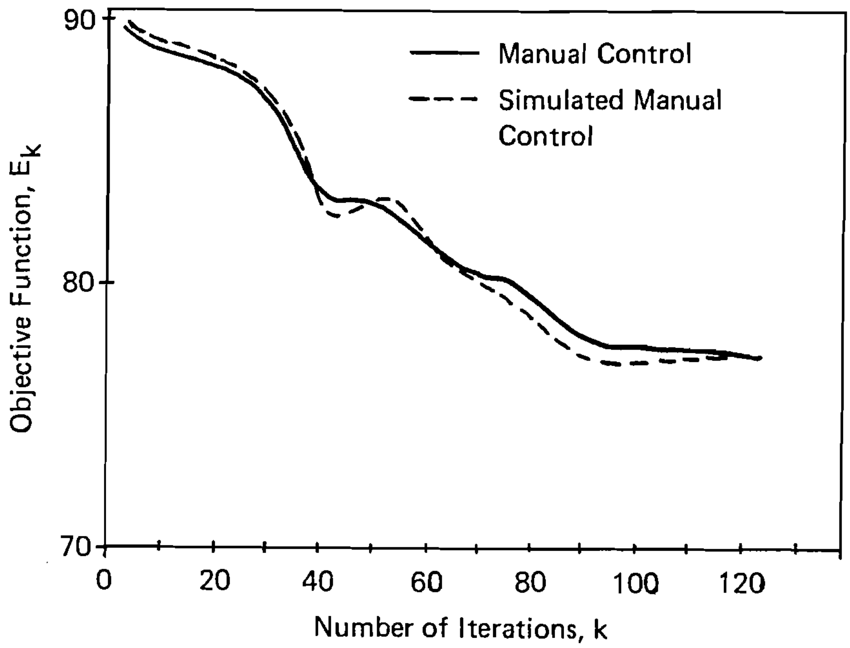

Figure 2. The convergence behavior of $\{E_k\}$ in the manual control and simulated manual control cases.

4.2 Speeding up the Convergence

If rigorously followed, the basic procedure for the step-size control may lead to a slow solution algorithm. First the manual step-size control with many I/O operations requires considerable effort from the person who calculates and usually effects a slow computer code. This happens especially in a time-sharing computer environment where the number of users is large and the average response time is long. Second, the number of iterations needed can be often significantly reduced.

To overcome the need of numerous manual I/O operations a simple automatic version of the manual step-size control was designed. Given three parameters the procedure simulates the behavior of the controlling person and reduces the step multiplier as soon as it observes a stationary or an oscillatory sequence $\{E_k\}$. Let the three input parameters be NB, DIF1, and DIF2. The first parameter NB fixes the batch size, i.e., the iterations will be performed in batches of NB iterations. Let the step multiplier used during the iteration batch be equal to ρ . A test indicator is defined as:

$$d_m = \frac{E_{(m-1)\cdot NB} - E_{m\cdot NB}}{\sum_{s\in M} \rho^s ||H^s||} , \qquad m = 1,\ldots, \tag{14}$$

The procedure then checks the two conditions

$$d_m \le \text{DIf1} \tag{15a}$$

and

$$\frac{\sum_{s\in M} \Delta^+ E_s}{\max_{s\in M} E_s - \min_{s\in M} E_s} \ge \text{D1F2} \tag{15b}$$

where

$$\Delta^+ E_s = \max\ (0, E_s - E_{s-1}) \quad (16a)$$

$$M = \{s \mid (m-1) \cdot NB \leq s \leq m \cdot NB\} \quad (16b)$$

In case either of these conditions holds the step multiplier is reduced by one half. The first condition (15a) tests if the decrease of the sequence proportioned to the step size used is less than the given limit. The second condition (15b) then checks if the sequence is oscillatory. This is done by considering the ratio of the sum of positive jumps of the sequence $\{E_k\}$ to the maximum change in the sequence that takes place during the iteration batch.

With DIF1 = 0.01 and DIF2 = 0.30 the procedure simulates the manual control very closely (Figure 2). Depending on the starting values used for x^0 and ρ^0 sometimes a few more iterations were performed than the manual control would have required, but the total computing time still usually remained smaller than in the case of manual control.

With the aforementioned values for DIF1 and DIF2 the automatic step-size control normally guarantees that the solution is eventually reached, independent of the initial values for x^0 and ρ^0. Often the algorithm can be made faster by using a greater value for DIF1. If for example, DIF1 = 1.00, the use of the control would reduce the step multiplier as soon as the total decrease of the objective function during a batch is less than the total change of the components in that batch. If the solution can be only roughly estimated initially, the number of iterations can be kept moderate. This can be done by choosing an initial value for ρ that will reach the solution region during a few iterations and by cutting down the step size as soon as the rate of decrease of the objective function slows down. Using the test indicator d_m of equation (14) the program checks if

$$d_m \leq \text{DIF1} \tag{17a}$$

or

$$d_m \leq d_{m-1} \tag{17b}$$

Instead of E_m an average of a few neighboring values of E_m can be used to calculate the indicator d_m. If any of conditions (17) holds, the step multiplier is cut down by a factor r, which is given as an input.

The effect of the accelerated procedure is seen in Figure 3 where the curves correspond to the accelerated step-size control. The reduction coefficient r is 0.5 in both cases but in the first case the batch size is 10, in the latter case, 5. DIF1 has now

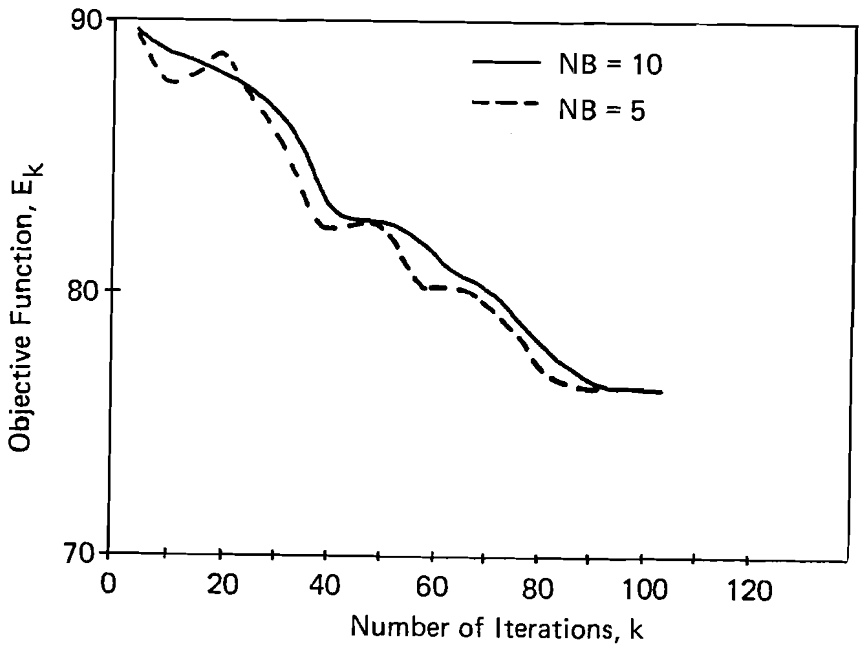

Figure 3. The convergence behavior of $\{E_k\}$ in the accelerated step-size control case.

been set to 1.0. It is seen that some decrease in the number of iterations have been obtained in both cases compared to the situation of Figure 2 but the difference is quite small. However, in this example a good estimate of the solution is known in advance and the number of iterations is rather small with any kind of step-size control. Note that if the initial estimate for X is far from the actual solution and a small initial value is used for ρ, then the accelerated procedure may reduce the step too rapidly, and an excessive number of iterations is needed to obtain the solution. This danger can be normally eliminated by selecting an initial ρ^0 estimate that is too big rather than too small. (The instructions for the user of the computer code SQG are given in Appendix A.)

5. A CASE STUDY

A simple example of a resource allocation problem that minimizes costs to meet uncertain demand will be discussed in this section. The problem is high school location in Turin, Italy. The physical setting and the data for this problem are described in Leonardi and Bertuglia (1981). For the purpose of the analysis, Turin is divided into 23 districts, each district being both a demand source and a possible high school facility location. Customers are assumed to behave according to a gravity-type model. For simplicity, travel time is assumed as the only explanatory variable for the choice behavior (some theoretical underpinnings for such models are described in Leonardi, 1980a and 1980b).

However, unlike in the standard use, the gravity model will be given a stochastic interpretation here, as suggested in Ermoliev and Leonardi (1980) (an earlier interpretation of the gravity model as a stochastic process is found in Bertuglia and Leonardi, 1979). That is, the relative distribution of students among facilities is looked at as a discrete multinomial Bernoulli distribution, rather than as a set of deterministic fractions. This is put in mathematical terms in the following.

Let s_i, $i = 1,\ldots,n$, be the total number of students at point i. The problem is to determine the size x_j of the facilities at points j, $j = 1,\ldots,n$, when it is known that the students at point i choose the facility at point j with probability

$$p_{ij} = \frac{e^{-\lambda c_{ij}}}{\sum_{j=1}^{n} e^{-\lambda c_{ij}}} \tag{18}$$

where λ is a constant and c_{ij} are empirical coefficients that depend on the distance between i and j (in the example: travel times in minutes). The use of (18) for the probabilities has theoretical and empirical justifications. Model (18) is a simplified form of the logit model discussed in McFadden (1973, and 1974) for example. If the flow of students between i and j is denoted ϕ_{ij}, the stochastic demand at point j is then

$$\omega_j = \sum_{i=1}^{n} \phi_{ij} \tag{19}$$

while the number of students at point i can be written as

$$s_i = \sum_{j=i}^{n} \phi_{ij} \tag{20}$$

The numbers s_i are now deterministic and given as an input. If the unit cost of capacity surplus is α and that of deficit is β and no other costs are considered, then our cost minimization problem is of the equation (4) type, $\alpha_j = \alpha$, $\beta_j = \beta$, $j = 1,\ldots n$.

The ability to generate random realizations, ω^s, of the demand vector ω is essential for the quasi-gradient method that is being discussed. The direct determination of the distribution of ω_j is practically impossible in this case. Instead, random vectors can be generated by simulating individual choices of the students according to the probabilities p_{ij} in (18). This still may lead to a time-consuming procedure if the total number of students, s_i, at points i is large. In this case the s_i number should be first scaled down by a factor η common to all the components $i = 1,\ldots,n$ (i.e., $\eta s_1, \eta s_2, \ldots, \eta s_\eta$). The final solution is then obtained by rescaling the solution of the smaller problem by $1/\eta$.

Table 1 shows the solutions obtained for $\alpha = \beta = 1.0$. In this case the solution $x_j = \Sigma_i s_i \cdot p_{ij}$ of a deterministic problem that is based on an entropy approach. The first column in Table 1 contains the labels of each district, numbered from 1-23. The second column of Table 1 gives the vector $S = (s_1,\ldots,s_{23})$ of total demands in each district; S was also used as the initial estimate for the iteration. Here the original data from Turin have been multiplied by 1/100. The next three columns show the results originating from the use of different starting values for the iteration. The last column shows the solution based on the deterministic model. In general, a good agreement exists between all the solutions; they are usually within two digits of each other. There are, however, some significant discrepancies. These can be partly explained by the stochastic nature of the convergence and by the flatness of the objective function near the solution. They associate somewhat with the slow convergence of the algorithm as the number of iterations increases. Then, while the scaling of the number of students saves computational effort that is required for the generation of the random realizations, the need for accuracy may soon counteract this benefit.

The discrepancies between the solutions in Table 1 can be associated with the shape of the probability densities underlying the probabilities of (18). The values that are used for the coefficients c_{ij} are listed in Appendix B, the value of the constant λ is 0.15. Probability densities can be numerically approximated from this data. Densities for several of the components are drawn in Figure 4. The densities are mostly symmetric and strongly peaked. In these cases the stochastic minimization solution, which corresponds to the median of this distribution,

Table 1. Optimal location of Turin high schools. Solutions obtained for penalty costs $\alpha = \beta = 1.0$.

District	Number of students	$\rho^0 = 1.00$ NB = 20	$\rho^0 = 1.00$ NB = 10	$\rho^0 = 1.00$ NB = 5	Deterministic solution
1	14.0	15.6	17.0	17.8	17.5
2	13.0	12.8	12.8	13.6	13.0
3	15.0	18.6	17.9	17.1	18.7
4	11.0	18.0	18.3	18.9	18.9
5	14.0	17.0	16.4	15.3	16.4
6	14.0	13.0	13.8	14.0	13.7
7	11.0	11.2	10.1	10.0	11.0
8	12.0	10.0	10.0	10.0	10.5
9	12.0	12.9	12.9	13.5	13.2
10	23.0	19.2	19.6	20.1	19.3
11	26.0	25.4	26.7	26.9	26.2
12	23.0	19.9	20.0	19.1	20.3
13	22.0	16.2	16.0	15.5	16.1
14	18.0	15.0	15.6	15.0	15.3
15	14.0	13.9	14.0	14.0	14.3
16	15.0	13.4	13.6	13.8	13.2
17	14.0	13.0	13.0	13.1	12.9
18	14.0	15.0	16.1	15.7	15.8
19	10.0	9.8	10.0	10.2	9.8
20	10.0	10.0	10.9	10.1	10.5
21	5.0	5.0	5.0	5.0	5.1
22	8.0	10.8	11.7	10.8	10.6
23	21.0	16.5	16.1	16.0	16.9

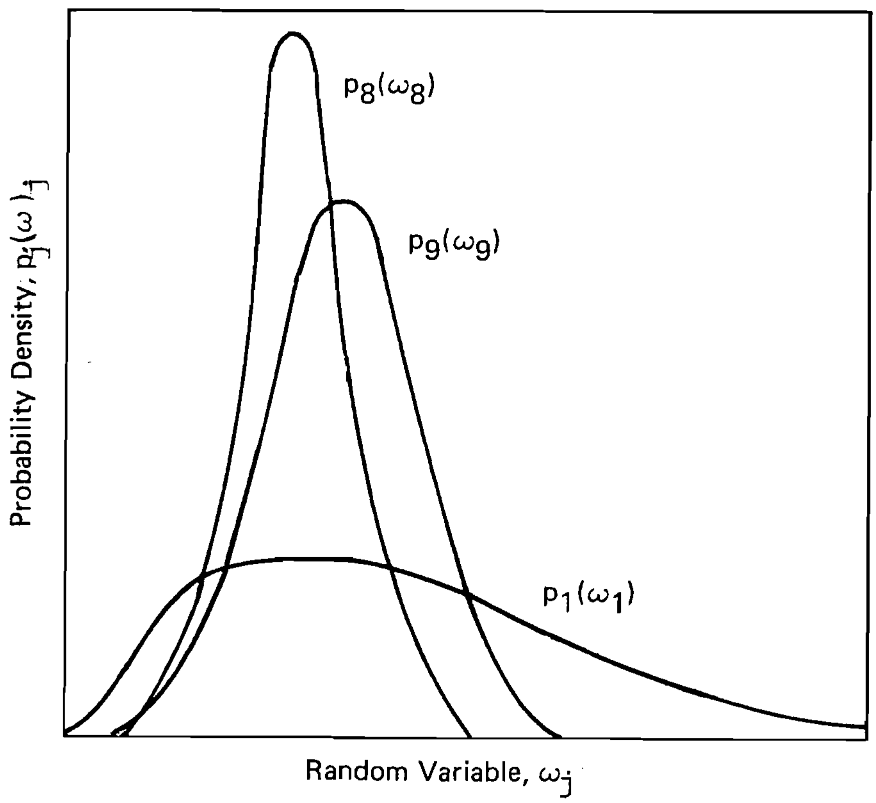

Figure 4. The probability densities for random demand ω_j at location j = 1.8, or 9.

and the deterministic solution, which corresponds to the expected value, should be close to each other. This is in fact demonstrated, for instance, by the facility sizes in districts 8 and 9, where the discrepancies are small. However, for district 1 the density is flat and skew, and the median and expected values are not equal. On the other hand, in the solutions for x_1 the discrepancies are large. The flatness of the density also explains the large discrepancies between the different solutions obtained from the stochastic minimization procedure.

Table 2. Optimal location of Turin high schools. Solutions obtained for different values of penalty costs α and β.

District	$\alpha = 1.00$ $\beta = 1.50$	$\alpha = 1.00$ $\beta = 2.00$	$\alpha = 1.50$ $\beta = 1.00$	$\alpha = 2.00$ $\beta = 1.00$
1	16.9	20.7	12.9	10.9
2	13.9	15.0	11.5	10.8
3	19.7	21.1	17.0	16.0
4	19.6	20.7	17.1	16.0
5	17.5	18.0	15.2	14.7
6	14.6	15.0	13.0	12.0
7	11.2	12.0	10.0	9.4
8	11.1	11.9	9.8	9.1
9	13.6	14.2	12.0	11.0
10	20.4	21.4	18.4	17.6
11	27.4	28.4	25.0	23.2
12	21.9	22.6	19.1	17.2
13	16.7	18.0	14.3	13.9
14	16.0	16.6	14.1	13.4
15	15.0	15.0	14.0	12.5
16	13.8	14.8	12.3	12.0
17	13.2	14.0	12.2	12.0
18	16.7	16.8	14.8	13.9
19	9.9	10.9	9.6	8.8
20	10.9	12.0	9.0	9.0
21	5.0	5.0	5.0	5.0
22	10.9	12.9	8.9	8.5
23	17.4	18.1	15.4	14.3

In Table 2 solutions are presented for cases where α and β differ from each other. As one could expect, the increase in the relative cost of deficit compared to the cost of surplus leads to larger values in the solution vector. If however, the probability density of the corresponding component of ω_j is very peaked, as in the case of ω_{21}, the change in the relative costs does not have any significant influence on the solution.

6. A NON-CONVEX OBJECTIVE FUNCTION

The problem discussed so far lacks some of the main features that are usually considered typical for optimal location problems. For instance, economies of scale, usually considered as that which makes location problems non-trivial, are absent in our earlier formulation. In deterministic models, economies of scale are usually introduced by means of fixed charges, to be paid when a facility is established, no matter what the number of attracted customers. This formulation is typical of the well known plant-location problems of Operations Research. Its extension to a gravity-type demand model has been developed in Erlenkotter and Leonardi (forthcoming). Related ways to introduce scale effects are by means of suitable constraints, as on the total number of facilities (the so-called "p median" problem (see ReVelle and Swain, 1970), or on the minimum feasible size for facilities (as in Leonardi and Bertuglia, 1981).

Here the first formulation will be explored. Let a fixed cost γ be defined, to be paid when a facility is established. For simplicity, the same value of γ for all districts will be assumed (as in Erlenkotter and Leonardi, forthcoming). Then the minimization of the expected cost calls for finding the minimum of the function.

$$G(X) = \sum_{j=1}^{n} \gamma\delta(x_j) + E\{\sum_{j=1}^{n} \max[\alpha(x_j-\omega_j),\ \beta(\omega_j-x_j)]\} \tag{21}$$

where δ (x) is the unit step function at zero. It is easy to see that with non-negative x_j, G(X) is not convex and usually has several local minima. The problems of this form are normally treated with mixed integer programming methods. Here we attempt to apply the general idea of stochastic quasi-gradients to finding the global minimum. Approximating the step function by a logarithmic function, the estimate

$$h_j^s = \frac{\gamma}{x_j^s+\varepsilon} + \begin{cases} \alpha \text{ if } x_j^s < \omega_j^s \\ -\beta \text{ if } x_j^s \geq \omega_j^s \end{cases} \tag{22}$$

with ε a small positive constant, is used for the generalized gradient at $X = X^s$. Otherwise the procedure in equation (8), remains as before.

In general, the procedure rapidly finds a minimum which is at least local. After that, however, some difficulties arise with the control of the iteration process. In principle, the approximation

$$G_k^1(X) = \gamma \sum_{j=1}^{n} \delta(x_j) + \frac{1}{k} \sum_{s=0}^{k} \sum_{j=1}^{n} \max\,[\alpha(x_j^s-\omega_j^s), \beta(\omega_j^s-x_j^s)] \tag{23}$$

can be used again to follow the course of iterations. Now, however, after a number of iterations the function $G_k^1(X^k)$ may achieve a minimum. On the other hand, some components of the estimate for the generalized gradient as calculated from equation (22) may still show a trend toward the origin, where another (at least) local minimum would be found. Note that with a small ε the origin becomes a fixed point for the iteration: if $X^{s0} = 0$ for one s_0, then $X^s = 0$ for all $s > s_0$. To overcome these difficulties, the initial value X^0 should be large enough and the initial step multiplier ρ should be chosen such that the step size is a small fraction of x_j. In this way a fallacious convergence towards zero during the first iterations can be excluded. To assess the behavior of the function G(X) at the various minima, a test function

$$G_k^2(X^k) = \gamma \sum_{j=0}^{n} \frac{x_j^k}{x_j^{k-m}} + \frac{1}{k} \sum_{s=0}^{k} \sum_{j=1}^{n} \max\,[\alpha(x_j^s-\omega_j^s), \beta(\omega_j^s-x_j^s)] \tag{24}$$

could be used. In this case m is a small integer, the choice of which slightly depends on the relative magnitude of α, β, and γ.

Figure 5 shows the behavior of the functions $G_k^1(X^k)$ and $G_k^2(X^k)$ with increasing k for $\alpha = \beta = 0.5$, $\gamma = 5.0$, $m = 6$. It is seen that $G_k^2(X^k)$ is monotonically decreasing toward the global minimum while $G_k^1(X^k)$ has two local maxima. Table 3 shows the vector X^k at $k = 180$, which corresponds to one local minimum of $G_k^1(X^k)$, and at the end of the iteration ($k = 280$). It cannot be proved that the solution obtained is the exact solution of the optimization problem. Indeed, the deterministic counterparts shown in Erlenkotter and Leonardi (forthcoming), are quite different. On the other hand, the computational effort that is needed for an estimation by the stochastic quasi-gradient method is also relatively small when compared to some integer programming methods, for instance.

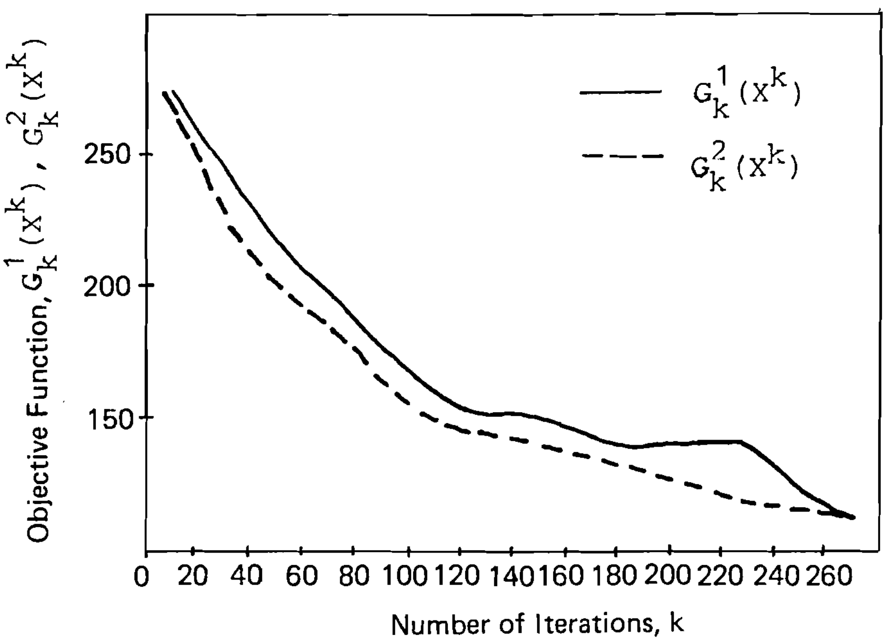

Figure 5. The behavior of $G_k^1(X^k)$ and $G_k^2(X^k)$ as a function of k.

Table 3. Optimal location of Turin high schools. Solutions obtained after 180 and 280 iterations with penalty costs $\alpha = \beta = 0.5$ and fixed charge $\gamma = 5.0$.

District	k = 180	k = 280
1	-	-
2	-	-
3	4.4	-
4	8.7	-
5	8.3	-
6	-	-
7	-	-
8	-	-
9	-	-
10	16.7	16.1
11	23.2	22.5
12	14.0	13.0
13	7.3	-
14	9.1	-
15	-	-
16		-
17	2.7	-
18	8.7	-
19	-	-
20	-	-
21	-	-
22	-	-
23	11.5	10.1

The solutions obtained depend mostly on the relative magnitudes of α, β, γ. With increasing fixed costs, γ, more facilities are likely to remain closed. When the β are increased the deficits are more penalized and thus more facilities remain open. Table 4 shows results from a sensitivity analysis on the values of α and β. The aim of the analysis is to find which values of α and β will cause the smallest facility (district 21) to disappear from

Table 4. Optimal location of Turin high schools. Results of a sensitivity analysis for changing values of penalty costs α and β. The fixed charge is fixed and equal to γ = 5.0.

District	α, β = 1.0	α, β = 1.5	α, β = 1.75	α, β = 2.0
1	9.2	11.2	11.4	12.9
2	9.8	11.0	11.4	11.4
3	16.0	16.5	17.1	17.7
4	16.4	17.1	17.5	17.3
5	14.0	15.0	15.1	15.2
6	11.1	12.0	12.0	12.1
7	8.1	9.3	10.0	9.7
8	8.3	9.0	9.0	9.0
9	10.0	11.9	11.9	12.0
10	17.0	18.3	18.5	18.7
11	24.7	24.9	24.9	24.9
12	18.0	18.1	18.5	19.0
13	13.8	14.3	14.4	14.5
14	13.8	14.0	14.0	14.0
15	12.4	13.0	13.0	13.0
16	11.6	12.0	12.3	12.4
17	11.7	12.0	12.0	12.0
18	13.9	14.1	14.6	14.9
19	8.4	8.8	9.0	9.0
20	7.0	8.7	9.0	9.2
21	-	-	-	5.0
22	6.4	8.0	8.1	9.0
23	14.7	15.1	15.7	15.8

the solution. This will happen almost certainly when β is less than 1.5. However, for a large range of values of x_{21}, between zero and five, the objective function remains almost constant. Hence, with these parameter values, opening or closing that facility does not have great influence on the value of the objective function. Table 5 shows the results of a sensitivity analysis on the fixed charge γ. The aim of this analysis is to find the least value of γ leading to a solution with a single facility open.

Table 5. Optimal location of Turin high schools. Results of a sensitivity analysis for changing values of fixed change γ. The penalty costs are fixed and equal to $\alpha = \beta = 1.5$.

District	$\gamma = 10.0$	$\gamma = 15.0$	$\gamma = 20.0$
1	-	-	-
2	-	-	-
3	13.9	-	-
4	14.7	12.3	-
5	12.5	-	-
6	9.9	-	-
7	-	-	-
8	-	-	-
9	8.7	-	-
10	16.2	14.4	-
11	24.5	23.7	19.1
12	18.6	16.0	-
13	14.3	11.2	-
14	13.9	-	-
15	11.8	-	-
16	11.8	-	-
17	11.5	-	-
18	13.1	-	-
19	-	-	-
20	-	-	-
21	-	-	-
22	-	-	-
23	14.1	5.7	-

A few comments are appropriate here on the comparison between the deterministic solutions, as determined in Erlenkotter and Leonardi (forthcoming) or Leonardi and Bertuglia (1981), and the solutions obtained with the stochastic quasi-gradient method. Some general tendencies are shared in common among all solutions, such as the low ranking of district 21 and the high ranking of district 11. The general clusters of open locations show also some similarity. A cluster of central districts (between 1-6), one of first-ring districts (between 9-18) and a few peripheral districts (usually district 23 only) appear in deterministic solutions as well. However, when one looks at the detailed composition of these clusters, no two of them are the same. Sometimes very striking differences are found, such as the closing or opening of district 1 (the downtown district), which would be difficult to justify to a public authority. The main cause for such lack of robustness of stochastic methods is the existence of many local minima and many near optimal solutions, with values of the objective function lying within a very narrow range. Of course a deterministic algorithm of an ennumerative nature can still detect small differences, even though it may take a long time. In a stochastic formulation, random fluctuations might well be of the same order of magnitude of the range of the objective function values. This seems to be the case in our examples.

7. CONCLUDING REMARKS

The purpose of this study has been to consider the stochastic quasi-gradient method for solving a resource allocation problem. The main advantages of the method are undoubtedly its computational simplicity and the small amount of information required - explicit probability distributions are not needed, random observations from a Monte Carlo simulation process will do.

The computational procedure for the basic recursion equation can be written by using only a few program statements and the storage requirements of the method are minimal. The generation of the random observations, however, may be time-consuming and hence the need for an optimized algorithm exists. The standard step-size control is based on the interactive use of the computer and normally guarantees that the solution is found after a moderate number of iterations. In this paper some methods are presented that do not necessarily require continuous control from the person who calculates and that often reduce the computation time.

Tests are also made for a case where the objective function is non-convex. In the deterministic formulation, problems of this type lead to integer programming methods that are often slow, unless for some special assumptions (like linearity) concerning the objective function and constraints. Here the solution is based on the same iteration algorithm as in the convex case. The existence of several local minima may cause some difficulties with the control of the iteration process, but the experience shows that with regard to its simplicity and speed the method can be efficiently applied to obtain good estimates for the solutions of these difficult problems.

The practical results for the problem of determining the size of school facilities in Turin were generally seen to be in agreement with the solutions derived by other means although differences in details are found. It is true that, given the special probability structure of equation (18), some simple deterministic algorithms are available (Erlenkotter and Leonardi, forthcoming). However, these algorithms do not apply to more general cases, where the stochastic procedure might be advantageous.

APPENDIX A

THE USE OF COMPUTER CODE SQG

For practical computations a FORTRAN program SQG was designed and implemented on a PDP 11/70 computer. The code has been meant for interactive use, but for some parts of the input a few files must be prepared in advance. This appendix describes the program to the extent necessary for its use.

INPUT

The input that is required for a successful computation with SQG consists of:

a. the problem specification
b. the control of the iteration process

Problem Specification

Prior to the execution of SQG three input files must be specified. These are referred to with the following device numbers:

2 the numbers of the customers (the first rows with format 10i5)
  the upper capacity bounds (the last rows with format 10i5)
3 the initial approximation (10i5)
4 the coefficients c_{ij} of the exponentials, see equation (18) (23f3.0)

The program asks for the rest of the input via the terminal. What remains to define the problem is the values of the coefficients α, β, γ, and c. These are required as

- INPUT alfa beta gamma c (4f6.2)

Control of the Iteration Process

- NORMAL = O AVERAGE = N (i1)

Every iteration batch can start from the previous estimate for the solution. If desired, the program can also utilize the average of the last NB/N estimates, where NB is the number of iterations in one batch. Therefore, type 'O' for normal batch startup, N, where N is an integer, if average is desired.

- DRAW ONLY = 2 ? (i1)

The program plots the objective function values (E_k or G_k) on the terminal screen (rotated 90^0). The objective function values as well as the random function values will also be printed unless '2' is typed at this point.

- STEP SIZE CONTROL ? (i1)

The step multiplier can be controlled in the three ways presented earlier in this report. Type

'1' for manual control (default)
'2' for simulated manual control
'3' for the rate-of-decrease based control

If '1' is Typed

- INPUT rho nb m (f6.4,2i3)

where

rho = ρ
nb number of the iterations in one batch (default 10)
m see equation (24)

(If no fixed cost is included, m can be disregarded.)

After NB iterations the following question is asked:

- WHAT NEXT rho xx.xx change xx.xx obj xx.xx

where 'change' is the sum $\Sigma_j |x_j' - x_j|$; x_j' and x_j are the values of the j^{th} component of the result estimate in the beginning of the batch and at the end of the batch, respectively. The current value for the objective function is 'obj'. Type an integer as follows:

negative	step iterations
zero	continue without adjustment of
positive	continue after adjustment of

In the last case the next value for ρ will be questioned.

If '2' is Typed

- INPUT rho nb dif1 dif2 m (f6.4,i3,2f5.2,i3)

where

rho	=	ρ
nb		the number of iterations in one batch (default 10)
dif1		see equations (15a,b)
dif2		
m		see equation (24)

If '3' is typed

- INPUT rho nb dif1 red m (f6.4,i3,2f5.2,i3)

where

rho	=	ρ
nb		number of iterations in one batch (default 10)
dif1		see equation (17a)
red		reduction coefficient r
m		see equation (24)

In the last two cases the only question asked after this is

- TERMINAL CONDITION ?

The iteration will be terminated as soon as the condition $\rho < 10^{-IER}$ holds, IER is given here as an answer (default 5).

OUTPUT

The output consists of two parts

1. the information that is necessary for the control of the iteration
2. the results

1. The objective function values (E_k or G_k) are plotted batch-wise on the screen as a function of the number of iterations performed. The axes, however, have been rotated 90^0 clockwise. After every batch of NB iterations the objective function value is also printed together with information on the change of

the result estimates during the iteration batch.

2. The current estimate for the solution as well as for the objective function is printed for every batch of NB iteration in the file specified with device number 9. If all objective function values (and the random function values) are desired, '2' should be answered to the appropriate question.

Note: The program is currently dimensioned for a demand vector of 23 locations. For other problem dimensions, change the first executable statement (nd = 23) and the dimensions of the tables in DIMENSION statements.

APPENDIX B

THE COEFFICIENTS c_{ij} FOR PROBABILITIES p_{ij}

i \ j	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	5	17	13	21	19	19	17	23	29	30	22	32	26	26	27	30	28	13	32	30	50	28	34
2	14	5	14	25	21	26	24	29	18	21	24	32	31	32	36	41	41	28	48	36	50	13	28
3	8	12	5	13	9	17	17	25	18	23	12	20	18	21	28	29	31	22	39	30	51	20	27
4	21	25	14	5	7	22	31	39	21	22	13	14	9	23	30	34	40	31	44	44	61	26	27
5	17	22	10	8	5	16	25	33	30	32	22	22	15	22	34	35	42	30	46	38	57	32	37
6	17	27	18	20	15	5	26	32	35	40	30	31	23	9	28	30	33	27	39	39	56	37	43
7	16	23	19	32	26	25	5	21	31	32	33	39	38	32	36	39	34	18	34	15	49	33	35
8	22	29	26	38	32	30	20	5	37	39	39	46	41	37	38	40	33	17	34	24	37	34	42
9	27	19	20	21	30	34	33	40	5	15	19	29	31	41	45	46	51	38	57	46	62	12	19
10	28	19	24	25	34	39	34	41	15	5	9	20	32	43	48	51	52	39	58	46	58	16	18
11	20	25	14	14	23	30	33	40	21	9	5	7	20	34	41	39	47	34	53	46	64	23	19
12	32	35	23	16	22	35	43	50	32	20	8	5	20	40	48	42	53	44	57	55	73	34	27
13	27	29	17	10	14	25	36	40	31	33	20	20	5	29	38	32	48	38	52	49	67	37	38
14	24	33	21	25	22	12	35	41	41	43	33	29	28	5	29	28	42	33	45	48	64	43	46
15	28	38	32	34	37	31	38	40	46	47	44	45	41	31	5	18	29	24	44	39	70	48	51
16	32	42	31	39	37	33	37	40	48	48	36	30	31	30	15	5	28	24	37	39	73	52	52
17	31	43	34	42	43	32	34	34	51	52	47	52	48	39	29	31	5	18	36	33	71	51	55
18	14	29	24	32	31	28	20	18	38	39	34	44	37	35	23	26	18	5	17	17	57	39	42
19	35	49	40	46	47	40	37	35	58	58	54	57	52	47	44	40	35	18	5	34	74	60	63
20	32	37	36	48	42	40	18	26	46	47	50	55	54	47	39	41	34	18	35	5	58	48	51
21	53	53	54	66	63	60	52	40	63	60	68	73	71	67	71	75	73	59	75	56	5	59	68
22	23	13	22	26	31	35	33	34	11	17	23	33	36	42	45	50	49	37	57	45	54	5	30
23	30	27	26	27	36	41	35	42	20	18	18	28	37	44	50	53	52	39	59	48	64	32	5

REFERENCES

Bertuglia, C.S., and G. Leonardi (1979) Dynamic models for spatial interaction. *Sistemi Urbani* 1(2):3-25.

Erlenkotter, D., and G. Leonardi, *Algorithms for Spatial Interaction-Based Facility Location Models*. WP-81-00. Laxenburg, Austria: International Institute for Applied Systems Analysis. (forthcoming)

Ermoliev, Y.M. (1976) *Methods of Stochastic Programming*. Moscow: Nauka. (in Russian)

Ermoliev, Y.M. (1978) *Methods of Nondifferentiable and Stochastic Optimization and their Applications*. WP-78-62. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ermoliev, Y.M., and G. Leonardi (1980) *Some Proposals for Stocahstic Facility Location Models*. WP-80-176. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Ermoliev, Y.M., and G. Leonardi, *Some Proposals for Stochastic Facility Location Models*. WP-81-00. Laxenburg, Austria: International Institute for Applied Systems Analysis. (forthcoming)

Ermoliev, Y.M., and E. Nurminski (1980) Stochastic quasi-gradient algorithms for minimax problems in stochastic programming. In: *Stochastic Programming*, edited by M.A.H. Dempster. London: Academic Press.

Leonardi, G. (1980a) *A Unifying Framework for Public Facility Location Problems*. WP-80-79. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Leonardi, G. (1980b) *A Multiactivity Location Model with Accessibility and Congestion-Sensitive Demand*. WP-80-124. Laxenburg, Austria: International Institute for Applied Systems Analysis.

Leonardi, G., and C.S. Bertuglia (1981) *Optimal High School Location First Results for Turin, Italy*. WP-81-5. Laxenburg, Austria: International Institute for Applied Systems Analysis.

McFadden, D. (1973) Conditional logit analysis of qualitative choice behavior. In: *Frontiers in Econometrics*, edited by P. Zarembka. New York: Academic Press.

McFadden, D. (1974) The measurement of urban travel demand. *Journal of Public Economics* 3:303-328.

RELATED PUBLICATIONS

1. David Hughes, Evgeni Nurminski, and Geoff Royston, *Nondifferentiable Optimization Promotes Health Care.* WP-79-90.

2. David Hughes and Andrzej Wierzbicki, *DRAM: A Model of Health Care Resource Allocation.* RR-80-23.

3. Giorgio Leonardi, *On the Formal Equivalence of Some Simple Facility Location Models.* WP-80-21.

4. Tony J. Van Roy and Donald Erlenkotter, *A Dual-Based Procedure for Dynamic Facility Location.* WP-80-31.

5. Donald Erlenkotter, *On the Choice of Models for Public Facility Location.* WP-80-47.

6. Giorgio Leonardi, *A Unifying Framework for Public Facility Location Problems.* WP-80-79.

7. Giorgio Leonardi, *A Multiactivity Location Model with Accessibility- and Congestion-Sensitive Demand.* WP-80-124.

8. Hedy Attouch and Roger Wets, *Approximation and Convergence in Nonlinear Optimization.* WP-80-142.

9. Yuri Ermoliev and Giorgio Leonardi, *Some Proposals for Stochastic Facility Location Models.* WP-80-176.

10. Giorgio Leonardi and Cristoforo Sergio Bertuglia, *Optimal High School Location: First Results for Turin, Italy.* WP-81-05.