MANAGEMENT SCIENCE
Vol. 36, No. 1, January 1990
Printed in U.S.A.

ONE WAREHOUSE MULTIPLE RETAILER SYSTEMS WITH VEHICLE ROUTING COSTS*

S. ANILY AND A. FEDERGRUEN
Faculty of Commerce and Business Administration, The University of British Columbia, Vancouver, BC, Canada
Faculty of Management, Tel-Aviv University, Tel Aviv, Israel
Graduate School of Business, Columbia University, New York, New York 10027
Graduate School of Business, Hebrew University, Jerusalem, Israel

We consider distribution systems with a depot and many geographically dispersed retailers each of which faces external demands occurring at constant, deterministic but retailer specific rates. All stock enters the system through the depot from where it is distributed to the retailers by a fleet of capacitated *vehicles* combining deliveries into *efficient routes*. Inventories are kept at the retailers but not at the depot.

We wish to determine feasible replenishment strategies (i.e., inventory rules and routing patterns) minimising (*infinite horizon*) long-run average transportation and inventory costs. We restrict ourselves to a class of strategies in which a collection of *regions* (sets of retailers) is specified which cover all outlets: if an outlet belongs to several regions, a specific fraction of its sales/operations is assigned to each of these regions. Each time one of the retailers in a given region receives a delivery, this delivery is made by a vehicle who visits all other outlets in the region as well (in an efficient route).

We describe a class of low complexity heuristics and show under mild probabilistic assumptions that the generated solutions are asymptotically optimal (within the above class of strategies). We also show that lower and upper bounds on the system-wide costs may be computed and that these bounds are asymptotically tight under the same assumptions. A *numerical* study exhibits the performance of these heuristics and bounds for problems of *moderate* size.

(INFINITE HORIZON INVENTORY RULES AND ROUTING PATTERNS; ONE WAREHOUSE MULTIPLE RETAILER SYSTEMS; ASYMPTOTIC OPTIMALITY)

Introduction and Summary

In many distribution systems important cost reductions and/or service improvements may be achieved by adopting efficient inventory replenishment strategies for all items and facilities concerned. Such strategies often need to exploit economies of scale that arise e.g. when shipping full (or close to full) truck loads or rail car loads of goods. The latter can often only be achieved by combining deliveries to distinct locations into efficient routes.

These efficiency improvements and service enhancements clearly require an *integrated* approach towards various logistical planning functions; in particular the areas of *inventory control* and *transportation planning* need to be *closely coordinated*; for example, shipping in smaller quantities and with higher frequency generally leads to reductions in inventory investments but requires additional transportation costs.

In this paper we consider distribution systems with a single depot and many geographically dispersed retailers each of which faces a specific demand process for a given item. All stock enters the system through the depot from where it is distributed to (some of) the retailers by a fleet of trucks, combining deliveries into efficient routes. Inventories are kept at the retailers but not at the depot, i.e. the distribution system is *coupled.* (This term was introduced in Rosenfield and Pendrock 1980.) In other words, the depot serves as a coordinator of the replenishment process and it acts as a "breakbulk" or transhipment

* Accepted by L. Joseph Thomas; received March 14, 1988. This paper has been with the authors 7 months for 2 revisions.

0025-1909/90/3601/0092$01.25

point from where all vehicles start and return to. (In some applications, the depot may in fact fail to correspond with a specific physical facility, in which case it is assumed that all stock is picked up directly from a single outside supplier.)

Our objective is to determine long-term *integrated* replenishment strategies (i.e., inventory rules and routing patterns) enabling all retailers to meet their demands while minimising long-run average system-wide transportation and inventory costs.

We assume that at each outlet, customer demands occur at a constant, deterministic but *outlet specific* rate. These demand rates are assumed to be integer multiples of some base rate $\mu > 0$. Inventory carrying costs are incurred at a constant rate per unit of time, per unit stored at one of the outlets. (This rate is assumed to be identical for all retailers.) The transportation costs include a fixed (leasing or renting) cost per route driven by one of the vehicles and variable costs proportional to the total (Euclidean) distance driven. Delivery patterns are restricted by a volume or weight capacity for the trucks, i.e. an upper bound on the total load a vehicle can carry.

Optimal policies can be very complex; note e.g. that even when an extremely *simple* inventory strategy is fixed specifying replenishments to all outlets at times $0, T, 2T, \ldots$ for some time interval $T > 0$, optimal delivery routes remain to be determined and this combinatorial problem is equivalent to the classical Vehicle Routing Problem which is notoriously hard. Moreover, the complexity of the structure of optimal policies makes them difficult, if not impossible, to implement even if they could be computed efficiently. Instead we restrict ourselves to a class of replenishment strategies Φ with the following properties: a replenishment strategy specifies a collection of *regions* (subsets of outlets) covering all outlets: if an outlet belongs to several regions a specific fraction of its sales/operations is assigned to each of these regions. Each time one of the outlets in a given region receives a delivery, this delivery is made by a vehicle who visits all other outlets in the region as well (in an efficient sequence or *route*). We use the terms regions and routes interchangeably.

Our restriction is similar to that applied in many other joint replenishment problems; see e.g., Chakravarty et al. (1982, 1985) and Barnes et al. (1987), discussed below. See also Federgruen and Zheng (1988), and Schwarz (1981) for a discussion of this and other restriction approaches in multi-echelon (item) inventory models. The choice of the class Φ is motivated by the following considerations. First, we note that a large amount of flexibility is preserved within the class Φ, by allowing retailers to be assigned to several regions, i.e., by allowing regions to overlap. In addition, for any replenishment strategy in Φ it is easy to evaluate its average cost per unit time or to demonstrate that it is dominated by a simpler strategy whose average cost *can* be evaluated. For policies outside of Φ, the mere determination of the average cost may be intractably hard. Thirdly, a relatively simple procedure allows one to compute a strategy which comes close to being optimal (in the class Φ). See below for precise characterizations of the simplicity of the procedure and the magnitude of the optimality gap.

Most importantly, the restriction is often imposed by the sales/distribution system under consideration. A basic distinction among such systems is between those based on the "Pre Sell" and those based on the "Route Sales" concept. In the former the sales and delivery functions are completely separated while they are integrated in the latter. The system employs a force of salesmen/distributors each of which is assigned to a given region of outlets: each "salesman" is required to visit the outlets in his region periodically in a given sequence (*route*) determining replenishment quantities (in the form of definite sales or unbinding consignments) and delivering them as well.

For many industries, Route Sales distribution systems are exceedingly popular because of a variety of perceived advantages. For example, according to a recent survey by O'Neil and Meegan (1987) only 17.4% of the softdrink producers and distributors employ "Pre-Sell" (down from 32% in 1983) (in its pure form) all others either using a pure "Route

Sales" system or (to an increasing degree) some hybrid variation thereof; see also Jabbonsky (1987). (These surveys fail to indicate to what extent hybrid systems employ fixed partitions of the customers.)

It is often necessary or desirable to consider (within the above specified class of replenishment strategies) additional constraints of the following types:

—upper bounds for the frequency with which deliveries may be made in each of the regions;

—upper bounds for the total sales volume to be assigned to each of the regions.

The need for such constraints may arise from limited vehicle fleet sizes, scarce loading, administrative and transportation facilities at the warehouse; sales volume constraints may arise in the case of the above Route Sales systems.

We derive *easily computable lower* and *upper* bounds for the minimal system-wide costs (within the above defined class of strategies) and prove under weak probabilistic assumptions that these bounds are asymptotically accurate, i.e., the ratio (upper bound/lower bound) decreases to one as the number of outlets increases to infinity. Such bounds provide cost estimates which may be used in various design studies. In addition we design *simple* $O(n \log n + Kn)$ heuristics achieving a similar, provable (asymptotic) degree of optimality, with n the number of retailers and K the demand rate of the largest retailer.

Experimental studies, reported in §3, indicate that the gap between the lower bounds and the costs of the constructed solutions is rather small, even when the number of outlets is only of moderate size. The availability of rigorous (asymptotic) optimality and accuracy results contrasts with the tradition in the vehicle routing literature which mostly relies on limited empirical testing for the validation of its proposed heuristics.

The simplicity of the proposed algorithms contrasts, moreover, with the complexity of at least the mathematical programming based heuristics used in most standard routing models. The latter usually require the availability of (large-scale) linear programming codes and branch and bound or cutting plane procedures to be used as subroutines as well as rather sophisticated matrix generators for the specification of the models. Availability of the latter and the size of the (mixed integer) programs involved, usually imply that the algorithms need to be run on high speed/large memory computers. Even though our procedures are currently only available on a mainframe computer as well, we are confident that they can easily be adapted for personal or micro-computers.

The generated replenishment strategies, exhibit in addition, a number of interesting structural properties.

(a) There exists a uniform upper bound $M^*\mu$ for the total demand rate in a single region.

(b) If the same delivery frequency and sales volume bounds apply to each of the regions, and the total sales volume in the system is an integer multiple of $M^*\mu$ then *each* sales territory, associated with a single route, corresponds with a total sales volume of $M^*\mu$ units per unit of time. If the total sales volume of the system is of the form $(k_1 M^* + k_2)\mu$ for $k_1 \geq 0$, $1 \leq k_2 \leq M^* - 1$ then a single route serves a number of retailers closest to the depot such that the corresponding sales territory has a sales volume of exactly $k_2\mu$ units per unit of time. All other sales territories represent a sales volume of exactly $M^*\mu$ units per unit of time. (Note again that regions may overlap.) In other words, in view of (a), the number of different sales territories is the minimum required and all sales territories, with the possible exception of one, correspond to the maximum feasible total sales volume per unit of time, namely $M^*\mu$.

(c) All sales territories are visited by a vehicle at equally spaced epochs and the delivery sizes to the retailers in the territory remain constant over time. Since routes (territories) may overlap, the quantities delivered at and the time intervals between consecutive replenishments of a given *retailer* may, however, fail to be constant when the retailer belongs to more than one route (territory).

(d) There exists a critical distance $\bar{R}$ such that only fully loaded vehicles depart to retailers at a distance from the depot exceeding $\bar{R}$.

The above results are obtained by verifying that the combined inventory control and vehicle routing problems reduce to special cases of a class of Euclidean routing problems (with general route cost functions) addressed in Anily and Federgruen (1987).

Existing inventory models with multiple stocking points assume rather *simple* structures for the delivery costs: most commonly, *separability* across locations is assumed, the cost of a single replenishment to a given outlet consisting of (at most) a fixed term and a variable component proportional with (or convex in) the quantity delivered. Some non-separable structures have been considered as well, however under restrictive assumptions which fail to be satisfied in the vehicle routing context.

For example, the most extensively studied inventory replenishment model with *joint delivery costs* is the so-called "joint replenishment problem", see Brown (1967), Goyal (1973, 1974a, b, 1985), Goyal and Belton (1979), Goyal and Soni (1984), Graves (1979), Nocturne (1973), Schweitzer and Silver (1983), Shu (1971), Silver (1976), Jackson et al. (1985), Chakravarty (1984a, b, c), Chakravarty and Goyal (1986). This model assumes that a major setup cost is incurred for each joint delivery, independent of which locations are involved. In addition, location specific setup costs are incurred for each location included in the joint delivery.

Chakravarty et al. (1982), (1985) consider a generalization of this cost structure where the cost of a joint delivery is given by a general nondecreasing concave function of the sum of the location-specific "setup costs". Barnes et al. (1987) consider additional generalizations of this structure, where each location may be characterized by several "setup cost components". Chakravarty et al. and Barnes et al. restrict themselves to a class of replenishment strategies which is similar to ours. Roundy (1986) considers a general "family cost structure", with a given list of sets of locations ("families" in his terminology): there exists a setup cost for each of the families which is incurred once whenever one of the locations in the family receives a delivery. Queyranne (1985) and Zheng (1987) consider quite general joint setup cost structures merely assuming *submodularity*, i.e. the increment in the setup cost due to the addition of a new retailer (product) to a given collection of locations (items) is assumed to be no larger than if the same retailer (product) were added to a subset of this collection. While submodularity is satisfied in many production and distribution settings, it fails to hold in our context, see the Appendix at the end.

The literature on vehicle routing problems on the other hand, usually assumes that replenishment frequencies and delivery sizes are exogenously determined (possibly by some external inventory control model). In this view, routing problems decompose into separate single period problems, and as a consequence virtually all of the literature deals with variants of the following generic problem, usually referred to as the VRP (Vehicle Routing Problem) i.e., the design of a set of routes, of minimal total length, covering a given collection of retailers such that each route starts from and eventually ends at the warehouse while satisfying the vehicles' capacity constraints as well as meeting retailers' demands. We refer to the introduction of Anily and Federgruen (1987) for a brief discussion of the VRP literature and for references to review articles.

Initial models integrating inventory allocation and vehicle routing problems in one warehouse, multiple retailer systems (Federgruen and Zipkin 1984, Federgruen, Prastacos and Zipkin 1986) consider a single planning period, but allow for *random* demands at the retailers. These authors show that the integrated models may be solved by constructive and/or interchange heuristics as well as mathematical programming based methods in an amount of time which is comparable to that of corresponding methods for the standard VRP. They also demonstrate that the ability to adapt delivery sizes (and hence inventories) in an integrated model may result in quite significant cost savings. Federgruen, Rinnooy

Kan and Zipkin (1985) consider a stylised Euclidean version of this model and derive easily computable and asymptotically accurate upper and lower bounds for the minimum system-wide costs as well as simple, asymptotically optimal partitioning schemes.

Bell et al. (1984) developed a computerized planning system based on a multi-period, combined inventory control/vehicle scheduling model. Their system has been implemented in several companies. (The first prize in the 1983 CPMS competition was awarded for the initial implementation at Air Products.) Numerical experiments with heuristic solution methods for integrated multi-period models have been reported in Golden et al. (1984), Assad et al. (1984), Dror et al. (1987), Dror and Ball (1986) and Dror and Levy (1986). The first *integrated infinite horizon* model appears to be due to Burns et al. (1985); as in our model variable transportation costs are assumed to be proportional with Euclidean distances and demands at the retailers occur at a constant deterministic rate. Confining themselves to the case where all retailers face *identical* demand rates, these authors obtain several solution heuristics and cost approximations, on the basis of a number of simplifying assumptions and heuristic derivations. See Anily (1986, p. 36) for a more detailed discussion of this model.

The remainder of this paper is organized as follows: In §1, we analyze our basic model (without sales volume constraints) and derive the above mentioned bounds, heuristics and asymptotic optimality and accuracy results. In §2, we briefly discuss a number of variants of the basic model. In §3, we present the results of a series of numerical experiments to assess the accuracy and performance of the bounds and heuristics (respectively) for problems of *moderate size*.

1. The Basic Model

Consider a system with one warehouse and n retailers. Let μ_j denote the demand rate of retailer j, $j = 1, \ldots, n$. As pointed out in the Introduction we assume that these demand rates are multiples of some common quantity μ, i.e., $\mu_j = k_j\mu$, $j = 1, \ldots, n$ with k_j an integer between 1 and K for some $K \geq 1$. (For example, all demand rates may be expressed in tons per month, thousands of gallons per month, or number of "standard" packages per month.) We define a *demand point* as a point in the plane facing a demand rate of μ. Each retailer j ($j = 1, \ldots, n$) can thus be viewed as consisting of k_j *demand points*, all facing identical demand rates of μ units per unit of time and all located at the same geographic point as retailer j. We restrict ourselves to Φ, the class of replenishment strategies which *partition* the set of demand points into a collection of *regions*; each time a delivery is made to one of the outlets in a given region, this delivery is made by a vehicle which visits all other demand points in the region as well.

Let $X = \{x_1, \ldots, x_N\}$ be the set of demand points in the Euclidean plane, with r_i the distance of demand point x_i from the depot x_0. We choose x_0 as the origin of the plane. We assume that the demand points are numbered in ascending order of their distances to the depot, i.e., $r_1 \leq r_2 \leq \cdots \leq r_N$ and that the retailers' initial stock at time 0 equals zero.

Let

h^+ = the inventory holding cost per unit of time per unit stored at the retailers.

c = the fixed cost per route driven.

(We assume without loss of generality, that the variable transportation cost per mile equals one.)

b = the capacity of a vehicle. (In uncapacitated models we let $b = \infty$.)

f^* = the upper bound on the frequency with which a given route may be driven.

(Without loss of generality, we assume $\mu/b \leq f^*$).

We use $\chi = \{X_1, X_2, \ldots, X_L\}$ to denote a partition of X. Thus L indicates the number of regions which may either be fixed or variable, and X_l ($l = 1, \ldots, L$) denotes the collection of demand points on route l. We write $m_l = |X_l|$ ($l = 1, \ldots, L$). Note that for a *given* partition of X into regions, the remaining problem reduces to a separate (constrained) EOQ-problem for each of the regions involved. It then follows that each strategy in Φ is dominated by one under which the demand points of each region receive deliveries at equidistant epochs which are therefore of *constant* size. We hence restrict ourselves to strategies of the latter type. (Note that since a retailer may be assigned to more than one region, its (*total*) delivery sizes may vary over time.)

Thus, for $l = 1, \ldots, L$ let

Q_l = the total amount delivered to X_l each time route l is driven;

$TSP(X_l^0)$ = the length of an optimal traveling salesman tour through X_l and the depot.

Also, let

$V^*(X)$ = the minimal long-run average cost among all strategies in the class Φ.

We first consider the basic model, with both $b < \infty$ and $f^* < \infty$ but *no* upper bounds on the regions' sales volumes.

LEMMA 1. a.

$$V^*(X) = \min\left\{\sum_{l=1}^{L} \min\left[\frac{1}{2} h^+ Q_l + \frac{\mu m_l}{Q_l}(TSP(X_l^0) + c):\right.\right.$$

$$\left.\left.\chi = \{X_1, \ldots, X_L\} \text{ is a partition of } X \text{ and } \frac{\mu m_l}{f^*} \le Q_l \le b\right]\right\}. \tag{1}$$

b. *For a given partition* $\chi = \{X_1, \ldots, X_L\}$ *let*

$$Q_l^* = \min\left\{b, \max\left[\frac{\mu m_l}{f^*}; (2\mu m_l(TSP(X_l^0) + c)/h^+)^{1/2}\right]\right\}. \tag{2}$$

Q_l^* *achieves the minimum within the square brackets in* (1).

c. *Any feasible partition* $\chi = \{X_1, \ldots, X_L\}$ *has* $m_l \le M^* \stackrel{\text{def}}{=} [f^*b/\mu]$.

d. *If* $m_l = M^*$, $M^*/(M^* + 1)b < Q_l^* \le b$. *In fact, if* f^*b/μ *is integer*, $Q_l^* = b$.

PROOF. a. Since all retailers have no initial stock, one easily verifies that it is optimal to make deliveries to each region when its stock drops to zero. Note that for a given set of demand points X_l the average holding cost per unit of time is given by $h^+Q_l/2$ and the average transportation costs per unit of time are given by $[TSP(X_l^0) + c]$ multiplied by the frequency region X_l is visited, i.e., $\mu m_l/Q_l$. The upper bound $Q_l \le b$ reflects the vehicle capacity constraint while the lower bound on Q_l reflects the frequency constraint.

b. This formulation is well known from the Economic Order Quantity (EOQ) model with bounded delivery sizes. It is also immediate from the convexity of the expression within the square brackets in (1), viewed as a function of Q_l.

c. Immediate from the bounds on Q_l in (1).

d. Note that $f^*b/\mu < M^* + 1$, so $f^*b/\mu M^* < 1 + 1/M^*$ and $\mu M^*/f^*b > M^*/(M^* + 1)$. □

Substituting (2) into (1) we obtain:

$$V^*(X) = \min\left\{\sum_{l=1}^{L} f(TSP(X_l^0), m_l): \chi = \{X_1, \ldots, X_L\} \text{ is a partition of } X\right\} \quad \text{where}$$

$$f(\theta, m) = \begin{cases} h^+\mu m/(2f^*) + f^*(\theta + c) & \text{if} \quad \theta + c \le \mu m h^+/(2f^{*2}), \\ (2h^+\mu m(\theta + c))^{1/2} & \text{if} \quad \mu m h^+/(2f^{*2}) \le \theta + c \le b^2h^+/(2\mu m), \\ h^+b/2 + \dfrac{\mu m}{b}(\theta + c) & \text{otherwise.} \end{cases} \tag{3}$$

It follows that the problem of determining an optimal strategy within the class Φ reduces to the problem of partitioning the set X into L feasible routes with minimal total cost where the cost of a route with length θ and m points is given by $f(\theta, m)$. (A route is feasible if it covers no more than M^* demand points.) This class of routing problems with general route cost function has been addressed in Anily and Federgruen (1987) and in the remainder we draw on the results of this companion paper. It is easily verified that f is nondecreasing in θ. We thus conclude from lemma 1 in Anily and Federgruen (1987) that

$$(\underline{P}^1)\text{: } \underline{V}^1(X) = \min \{ \sum_{l=1}^{L} f(2 \sum_{j \in X_l} r_j / m_l, m_l) \text{: } \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ and } m_l \leq M^*, l = 1, \ldots, L\} \quad \text{and} \tag{4}$$

$$(\underline{P}^2)\text{: } \underline{V}^2(X) = \min \{ \sum_{l=1}^{L} f(2 \max_{j \in X_l} r_j, m_l) \text{: } \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ and } m_l \leq M^*, l = 1, \ldots, L\} \tag{5}$$

are both lower bounds for $V^*(X)$ and moreover $V^*(X) \geq \underline{V}^2(X) \geq \underline{V}^1(X)$.

Let $F^2(j) = \underline{V}^2(\{x_{j+1}, \ldots, x_N\})$. (Note that $F^2(0) = \underline{V}^2(X)$.)

Before proceeding, we first need to review a few properties of partitions and partitioning problems.

A set of points $Y \subset X$ is *consecutive* if the indices of the points are consecutive integers. A partition $\chi = \{X_1, \ldots, X_L\}$ is called consecutive if $X_1 = \{x_1, \ldots, x_{m_1}\}$, $X_2 = \{x_{m_1+1}, \ldots, x_{m_1+m_2}\}, \ldots, X_L = \{x_{m_1+\cdots+m_{L-1}+1}, \ldots, x_N\}$ for some $m_1, m_2, \ldots, m_L$ with $\sum_{l=1}^{L} m_l = N$. (For example, the partition $\{\{x_1, x_2\}, \{x_3, x_4, x_5\}, \{x_6, x_7\}\}$ is consecutive while $\{\{x_1, x_3\}, \{x_2, x_4, x_5\}, \{x_6, x_7\}\}$ is not.) A partition is called *monotone* if it is consecutive and if $m_1 \leq m_2 \ldots \leq m_L$.

A partitioning problem of the set X into L regions with capacities $\{M_1^*, \ldots, M_L^*\}$ is determined by a cost function U which assigns a cost $U(\chi)$ to each partition χ; the partitioning problem consists of determining (P: $\min \{U(\chi) \text{: } \chi = \{X_1, \ldots, X_L\}$ and $m_l \leq M_l^*\}$.

A partitioning problem is said to be extremal if the following two properties are satisfied:

(i) an optimal monotone partition exists;

(ii) let $\chi = \{X_1, \ldots, X_L\}$ be a monotone partition of X. The cost of a partition χ' obtained by transferring the highest indexed element from some set X_l ($1 \leq l < L$) to X_{l+1} is less than or equal to the cost of χ.

Computation of the Lower Bounds $\underline{V}^1$ and $\underline{V}^2$

The results in Anily and Federgruen (1986) and Anily (1986) show that both $\underline{V}^1$ and $\underline{V}^2$ may be computed efficiently. Since f is nondecreasing, it follows from Theorem 7 ibid, that an optimal consecutive partition exists for the partitioning problem $\underline{P}^2$. This allows us to solve (5) by the following dynamic program

$$F^2(j) = \min_{1 \leq j'-j \leq M^*} \{f(2r_{j'}/(j'-j), j'-j) + F^2(j')\}.$$

The resulting algorithm has complexity $O(NM^*)$ in case L is variable and $O(NLM^*)$ in case L is fixed. (Observe that for a fixed value of M^*, L must grow linearly with N to ensure that a feasible solution exists.) The complexity of the latter case is thus $O(N^2M^*)$ (see Table 1 in Anily and Federgruen 1987). Further simplification in the computation of $\underline{V}^2$ does not appear to be possible, since $\underline{P}^2$ fails to be extremal (as pointed out in Example 8 in Anily and Federgruen 1986). Below we show, however, that $\underline{P}^1$ is extremal which permits us to employ the simple Extremal Partitioning Algorithm (EPA) presented in Anily and Federgruen (1986) requiring no more than $3\lceil N/M^*\rceil$ elementary operations and no evaluations of the function f, see Table 1 in Anily and Federgruen (1987) as well as the discussion ibid.

THEOREM 1. (a) *The function f defined by* (3) *is concave in both arguments* (*separately*).

(b) *The function $h(\Theta, m) = f(\Theta/m, m)$ has antitone differences*, i.e. $[h(\Theta_1, m)$ $h((\Theta_2, m)]$ *is nonincreasing in m for all $\Theta_1 > \Theta_2$.*

(c) $\underline{P}^1$ *is extremal.*

PROOF. (a) We first show that $\partial f/\partial\theta$ ($\partial f/\partial m$) exists and is continuous in $\theta(m)$. Since $\partial^2 f/\partial\theta^2$ ($\partial^2 f/\partial m^2$) exists and is nonpositive almost everywhere, we conclude that $\partial f/\partial\theta$ ($\partial f/\partial m$) is nonincreasing in $\theta(m)$ i.e. f is concave in $\theta(m)$. To verify existence and continuity in θ of $\partial f/\partial\theta$, it clearly suffices to establish that $\partial^+ f/\partial\theta$ and $\partial^- f/\partial\theta$, the right and left side partial derivatives of f with respect to θ, are equal in points (θ^0, m^0) satisfying (i) $\theta^0 + c = \mu m^0 h^+/(2f^{*2})$ and (ii) $\theta^0 + c = b^2h^+/(2\mu m^0)$. But for (i) $\partial^+ f/\partial\theta = (h\mu m^0/(2(\theta^0 + c)))^{1/2} = f^* = \partial^- f/\partial\theta$ and for (ii)

$$\frac{\partial^+ f}{\partial\theta} = \frac{\mu m^0}{b}\frac{(h^+\mu m^0)^{1/2}}{(2(\theta^0 + c))^{1/2}} = \frac{\partial^- f}{\partial\theta}.$$

Similarly, to verify existence and continuity in m of $\partial f/\partial m$, it suffices to show that $\partial^+ f/\partial m = \partial^- f/\partial m$ are equal in points (θ^0, m^0) satisfying (i) and (ii). But for (i),

$$\frac{\partial^+ f}{\partial m} = \frac{(h^+\mu(\theta^0 + c))^{1/2}}{(2m^0)^{1/2}} = \frac{h^+\mu}{2f^*} = \frac{\partial^- f}{\partial m}$$

and for (ii)

$$\frac{\partial^+ f}{\partial m} = \frac{\mu(\theta^0 + c)}{b} = \frac{\partial^- f}{\partial m}.$$

(b) Note that

$$h(\Theta, m) = \begin{cases} h^+\mu m/(2f^*) + f^*\left(\dfrac{\Theta}{m} + c\right) & \text{if} \quad \dfrac{\Theta}{m} + c \le \dfrac{\mu m h^+}{2(f^*)^2}, \\[2ex] (2h^+\mu(\Theta + cm))^{1/2} & \text{if} \quad \dfrac{\mu m h^+}{2(f^*)^2} < \dfrac{\Theta}{m} + c \le \dfrac{b^2 h^+}{2\mu m}, \\[2ex] h^+ b/2 + \dfrac{\mu}{b}(\Theta + cm) & \text{otherwise.} \end{cases}$$

Since $\partial f/\partial\theta$ exists everywhere (see proof of part (a)) it follows from the chain rule that $\partial h/\partial\Theta$ exists everywhere. We show that $\partial h/\partial\Theta$ is continuous in m. Since $\partial^2 h/\partial m\partial\Theta$ exists and is nonpositive almost everywhere, it follows that $\partial h/\partial\Theta$ is nonincreasing in m, hence h has antitone differences. Note that

$$\frac{\partial h}{\partial\Theta} = \begin{cases} \dfrac{f^*}{m} & \text{if} \quad \dfrac{\Theta}{m} + c \le \dfrac{\mu m h^+}{2(f^*)^2}, \\[2ex] [h^+\mu/(2(\Theta + cm))]^{1/2} & \text{if} \quad \dfrac{\mu m h^+}{2(f^*)^2} < \dfrac{\Theta}{m} + c \le \dfrac{b^2 h^+}{2\mu m}, \\[2ex] \dfrac{\mu}{b} & \text{otherwise.} \end{cases} \qquad (6)$$

To establish continuity of $\partial h/\partial\Theta$ with respect to m, it clearly suffices to verify that

$$\lim_{m\downarrow m^0}\frac{\partial h}{\partial\Theta}=\lim_{m\uparrow m^0}\frac{\partial h}{\partial\Theta}$$

for points (Θ^0, m^0) satisfying (i) $\Theta^0/m^0+c=\mu m^0h^+/(2f^{*2})$ or (ii) $\Theta^0+cm^0=b^2h^+/2\mu$. But the latter follows from (6).

(c) Part (c) follows immediately from parts (a) and (b), f nondecreasing in θ and Theorem 5 in Anily and Federgruen (1986). □

Since $\underline{P}^1$ is extremal, and all regions have the same capacity, it follows that a closed form expression for $\underline{V}^1(X)$ may be given. Consider first the case where L is variable. Let $N=k_1M^*+k_2$, $0\le k_2<M^*$. Then

$$\underline{V}^1(X)=f(2(\sum_{i=1}^{k_2}r_i)/k_2,k_2)+\sum_{l=0}^{k_1-1}f\left(\frac{2}{M^*}\sum_{i=lM^*+k_2+1}^{(l+1)M^*+k_2}r_i,M^*\right),\tag{7}$$

i.e., the k_2 points closest to the depot are assigned to a single region and all other points are assigned to consecutive regions of cardinality M^* each.

In case an additional constraint on the number of sales regions is imposed (i.e., $l\le\bar{L}$ with $\bar{L}$ a given parameter), we either have $\bar{L}<\lceil N/M^*\rceil$ in which case no feasible solution exists, or the constraint is nonbinding and $\underline{V}^1(x)$ is still given by (7).

If a *specific* number of regions is required, i.e., $L=\bar{L}$ for some $\bar{L}$ satisfying $\lceil N/M^*\rceil\le\bar{L}\le N$, the (EPA) partitions X into *three* (possibly empty) consecutive sets $W=\{x_1,\ldots,x_k\}$, $Y=\{x_{k+1},\ldots,x_{k'}\}$ and $Z=\{x_{k'+1},\ldots,x_N\}$ with $0\le k\le k'\le N$ and $k'-k\le M^*-1$ such that each point in W constitutes a sales region by itself, the set Y, if not empty, is a single sales region of cardinality less than M^* (but bigger than one), and the points in Z are assigned to consecutive sales regions of cardinality M^* each. We conclude that even in this case, there are at most three distinct regional cardinalities. In the remainder of this section we confine ourselves to the case where L is variable. (Modifications of the other cases are straightforward, in view of the above observations.)

Upper Bound and a Heuristic Solution

We now derive an upper bound for V^* and design a heuristic which are asymptotically accurate and asymptotically optimal, respectively, for extremely general stochastic sequences $\{x_1,x_2,\ldots\}$. We start by determining an optimal partition χ^* for $\underline{P}^1$. Let $X^{(m)}=\{x_i\in X: \chi^*$ assigns x_i to a set of cardinality $m\}$, $m=1,\ldots,M^*$ denote the set of demand points which are assigned to a region with m points (in total).

Anily and Federgruen (1987) proposes a general partitioning scheme (the so-called Modified Circular Regional Partitioning (MCRP)-procedure) which is applicable to routing problems with general route cost function $f(\cdot,\cdot)$. This procedure operates on each of the sets $\{X^{(m)}: m=1,\ldots,M^*\}$ separately and clusters the points in $X^{(m)}$ into regions of cardinality m each.

Note from the above characterization of the partition χ^* optimising $\underline{P}^1$ that in our case only the points in $X^{(M^*)}$ need to be partitioned into regions: recall that both for L variable and for L fixed at most one of the sets $X^{(m)}$ with $1<m<M^*$ may be nonempty and this set consists of exactly m points which necessarily need to be assigned to the same region.

The (MCRP) (as applied to the set $X^{(M^*)}$) is specified as follows:

Modified Circular Regional Partitioning Scheme.

Step 1. $m:=M^*$, $n_m:=|X^{(m)}|$, $R_m:=\max\{r_i|x_i\in X_{(m)}\}$ and $q_m:=\lfloor n_m/(m\lfloor n_m^{1/2}\rfloor)\rfloor$.

Step 2. Partition the circle with radius R_m into $\lfloor n_m^{1/2}\rfloor$ consecutive sectors containing mq_m points each and potentially one additional sector containing $n_m-\lfloor n_m^{1/2}\rfloor mq_m$ points. Let K denote the number of generated sectors. (Note $K=\lfloor n_m^{1/2}\rfloor$ or $K=\lfloor n_m^{1/2}\rfloor+1$.) Let $S_k^{(m)}$ denote the kth generated sector, $k=1,\ldots,K$.

Step 3. For each $k = 1, \ldots, K$ partition the sector $S_k^{(m)}$ by circular cuts such that each of the subregions contains m retailers and denote by $S_{k,l}^{(m)}$ the lth subregion in the kth sector, $l = 1, \ldots, |S_k^{(m)}|/m$.

Step 4. For each of the generated subregions, determine the optimal traveling salesman tour through the depot and the m points in the subregion.

Since f is nondecreasing and concave in θ, it follows from Theorem 3 in Anily and Federgruen (1987) and the subsequent discussion that the generated solution is asymptotically optimal under very general conditions regarding the sequence $\{x_1, x_2, \ldots\}$, see the discussion ibid. This remains true if instead of computing optimal traveling salesman tours in Step 4 of (MCRP) we apply a (TSP) heuristic whose worst case relative error is bounded (for example, Christofides' algorithm, see Christofides 1976). Likewise, the cost associated with the solution and $\underline{V}^1(X)$ provide asymptotically accurate upper and lower bounds (under the same conditions). We refer to Anily and Federgruen (1987) for the computation of alternative (asymptotically accurate) upper bounds which avoid the explicit determination of traveling salesman tours in the sales regions.

Also, if $\underline{V}^2(X)$ were determined separately (by the above described dynamic programming algorithm), it would provide a somewhat superior (and hence certainly asymptotically accurate) lower bound. In fact, as pointed out in Anily and Federgruen (1987), the partitioning scheme may be applied to the sets $X^{(m)}$, $m = 1, \ldots, M^*$ associated with any partition minimizing $\underline{P}^2$. Note, however, that such a partition may not need to be monotone so that the convex hulls of $X^{(m)}$ $(m = 1, \ldots, M^*)$ may overlap. In addition, it appears that all of the sets $X^{(m)}$, $(m = 1, \ldots, M^*)$ may be nonempty in which case the (MCRP) has to be applied M^* times. In view of these complications we ignore this alternative in our discussion below.

We complete this section by summarizing the proposed algorithm. Let TRVS(S) be a subroutine which determines for any $S \subset X$ the length of either an optimal or a heuristic traveling salesman tour through the points in S and the depot. (If a heuristic is employed, it is assumed that its relative error is uniformly bounded.) Also, let EOQ(θ, m) be the "optimal" delivery size Q_l^* for a region containing m demand points when its demand points are procured in a route of length θ, as determined by (2). Let also $R_{m,m+l} \stackrel{\text{def}}{=} \sum_{i=m}^{m+l} r_i$.

The Combined Routing and Replenishment Strategies Algorithm (CRRSA).

Step 1. Initialize $n := N$; LB $:= 0$; $M^* := \lfloor f^* b/\mu \rfloor$; $k_1 := \lfloor N/M^* \rfloor$;

Step 2. (compute the lower bound LB $= \underline{V}^1(X)$):

```
begin R := 0
      if n ≥ M* then
              begin
                R := R_{n-M*+1,n};
                LB := LB + f(2R/M*, M*);
                n := n − M*; repeat step 2;
              end;
      else
              begin
                if k_2 = 0 go to Step 3;
                otherwise
                  begin
                  R := R_{1,k_2};
                  LB := LB + f(2R/k_2, k_2);
                  end;
              end;
      end;
```

Step 3. Apply the (MCRP) to the set $\{x_{k_2+1}, x_{k_2+2}, \ldots, x_N\}$ with $m = M^*$. Let $S_1, S_2, \ldots, S_{k_1}$ be an enumeration of the generated regions.

Step 4. (computation of a solution and upper bound UB)

```
begin
    if k2 > 0 then
        begin
            l := 0;
            θ_l := TRVS({x_1, ..., x_k2}); (length of the tour)
            UB := f(θ_l, k2);
            Q_l := EOQ(θ_l, k2); (delivery size)
            f_l := k2 μ/Q_l; (replenishment frequency)
        end;
    l := 1;
    while l ≤ k1 do
        begin
            θ_l := TRVS(S_l);
            UB := UB + f(θ_l, M*);
            Q_l := EOQ(θ_l, M*);
            f_l := M* μ/Q_l;
            l := l + 1;
        end;
end;
```

As explained in Anily and Federgruen (1987), Step 3 of (CRRSA) requires $O(N \log N)$ and Step 4 only $O(N)$ operations. The overall complexity of the entire algorithm is thus $O(N \log N)$. In terms of the number of retailers n, the complexity may be bounded by $O(n \log n + Kn)$. (Note that the $O(N \log N)$ complexity bound for the MCRP arises from the need to rank the demand points with respect to both of their polar coordinates. This may be achieved by ranking the retailers according to these two attributes first, thus requiring $O(n \log n + N) = O(n \log n + Kn)$ operations only.) Since a problem instance is specified by $O(n)$ input parameters (the retailers' coordinates, the demand rate multiples $\{k_j, j = 1, \ldots, n\}$ and a few cost and constraint parameters), the algorithm is strictly speaking, not fully polynomial in the usual "complexity theoretical" sense. We argue, however, that in practical applications, $K = \max\{k_j, j = 1, \ldots, N\}$ is relatively small and for fixed values of K the algorithm is $O(n \log n)$ only! It is worth noting that, in view of Lemma 1 part (d), the delivery sizes and hence the replenishment intervals of all but the first region are rather close to each other, and they are in fact identical in case $f^* b / \mu$ is integer. Indeed, in the latter case the delivery sizes to all sales regions of size M^* are equal to b.

2. Some Variants of the Model

In this section we discuss a few variants of the basic model.

Uncapacitated Models

Uncapacitated models arise whenever the truck capacity is large relative to the delivery sizes or the demand rates are small relative to the holding costs. In that case we can ignore the vehicles' capacity constraints (i.e., $Q_l \le b$, $l = 1, \ldots, L$) and the function $f(\theta, m)$ simplifies to:

$$f(\theta, m) = \begin{cases} h^+ \mu m/(2f^*) + f^*(\theta + c) & \text{if} \quad \theta + c \le \mu m h^+/(2f^{*2}), \\ (2h^+ \mu m(\theta + c))^{1/2} & \text{otherwise.} \end{cases} \tag{8}$$

Clearly, the properties specified in Theorem 2 are retained. However, the number of demand points in the sets belonging to an optimal partition for $\underline{P}^1$ or $\underline{P}^2$, may no longer be uniformly bounded in N. In fact, the partition generated by the (EPA) (again optimal for $\underline{P}^1$ since $\underline{P}^1$ remains extremal) assigns all demand points to a single set. The same conclusions are reached in the limiting case where $f^* = \infty$ (and $b < \infty$). These results are analogous to a single traveling salesman tour being optimal in classical single-period *uncapacitated* routing models.

Nonidentical Vehicles; Sales Volume Constraints

Observe that the (CRRSA) algorithm may easily be generalized (and the above optimality and accuracy results extended) to the case where the vehicle fleet consists of vehicles with *nonidentical capacities* and/or different frequency bounds apply to different regions. This would result in *nonidentical* upper bounds $\{M_l^*: l = 1, \ldots, L\}$ for the number of demand points which may be assigned to the L regions, i.e. constraints of the form $m_l \leq M_l^*$ $(l = 1, \ldots, L)$.

Such *nonidentical* regional capacities $\{M_l^*: l = 1, \ldots, L\}$ may also arise when for reasons explained in the Introduction, an upper bound is imposed on each of the regions' sales volumes. Given our restriction to the class Φ and assuming all upper bounds are specified as multiples of μ, these translate into upper bounds on the number of demand points that may be assigned to the sales regions, i.e., constraints of the form $m_l \leq N_l^*$ $(l = 1, \ldots, L)$ for given integers N_l^* $(l = 1, \ldots, L)$. The effect of such additional constraints is again limited to replacing the upper bounds for m_l in (4) and (5) by $m_l^* = \min\{M^*, N_l^*\}$ $(l = 1, \ldots, L)$. In models with uncapacitated vehicles i.e., $M^* = \infty$ the constraints, of course, simplify to $m_l \leq N_l^*$, thus restricting the regions' sizes whereas they would otherwise not be. Note that an optimal partition for $\underline{P}^1$ may still be computed with the (EPA) algorithm, even when the regional capacities $\{m_l^*: l = 1, \ldots, L\}$ are *nonidentical*. However, more than two of the sets $\{X^{(m)}\}$ associated with this optimal partition for $\underline{P}^1$ may now be nonempty, thus requiring multiple applications of (MCRP) in Step 3 of (CRRSA); see Anily and Federgruen (1987) for details.

A Backlogging Option

If demands may be backlogged, a retailer with a demand rate of $k\mu$ $(k = 2, 3, \cdots)$ is no longer equivalent to k independent points each facing a demand rate of μ. For if these k points would be assigned to different regions, a solution may be generated in which their replenishment frequencies are different and some of these k demand points may thus have backlogs while (at the same time) inventories are carried in others. Representing the retailer as k independent demand points could thus lead to serious overestimates of the average inventory-related costs under any replenishment strategy. We thus restrict ourselves initially to the case where the retailers face identical demand rates (of μ units each) so that each retailer can be represented as a single demand point.

Assume that in addition to the cost components of the basic model, a cost h^- is incurred per unit of time for each unit backlogged at one of the retailers. Restricting ourselves again to the class of strategies Φ, consider a sales region X_l consisting of m_l demand points and receiving a total delivery Q_l each time it is visited by a vehicle. Given this replenishment structure one easily verifies along the lines of Carr and Howe (1962), that it is optimal for each retailer i to receive deliveries of constant size (say q_i) when its backlog has reached a given (constant) size s_i $(s_i \geq 0)$. The minimal long-run average cost for the region is thus given by, see, e.g., Hadley and Whitin (1963):

$$\frac{\mu m_l}{Q_l}(TSP(X_l^0) + c) + \min_{\substack{s_1, \ldots, s_{m_l} \\ q_1, \ldots, q_{m_l}}} \left\{ \sum_{i=1}^{m_l} \left[\frac{h^+(q_i - s_i)^2}{q_i} + \frac{h^-}{2} s_i^2 / q_i \right] : \sum_{i=1}^{m_l} q_i = Q_l \right\}. \quad (9)$$

Note that each of the terms within square brackets in (9) is jointly convex in (q_i, s_i). The minimum in (9) is thus achieved by equalizing all q_i and s_i-values $(i = 1, \ldots, m_l)$. Hence, $q_i = Q_l/m_l$ and $s_i = s$ $(i = 1, \ldots, m_l)$. Let $S_l = m_l s$. Substitution in (9) results in the following expression:

$$[h^+(Q_l - S_l)^2 + h^- S_l^2]/(2Q_l) + \frac{\mu m_l}{Q_l}(TSP(X_l^0) + c). \tag{10}$$

Since S_l is unconstrained, the partial derivative of this expression with respect to S_l must equal zero for any optimal combination (Q_l^*, S_l^*). It is thus easily verified that $S_l^* = h^+/(h^+ + h^-)Q_l^*$. Substitution into (10) shows that the minimal long-run average cost for region X_l is given by:

$$\frac{\hat{h}}{2}Q_l + \frac{m_l\mu}{Q_l}(TSP(X_l^0) + c), \qquad \text{where} \qquad \hat{h} = \frac{h^+h^-}{h^+ + h^-}.$$

We conclude that (1) continues to represent an exact expression for $V^*(X)$ as long as h^+ is replaced by $\hat{h}$. With this substitution, the entire analysis proceeds as before.

In addition, in our basic model with *identical* regional capacities, i.e. $M_l^* = M^* = (f^*\mu)/b$ the results apply even to systems with *nonidentical* retailers provided f^*b/μ is integer. While in general the representation of a retailer with demand rate $k_i\mu$ as k_i independent demand points may lead to a serious over estimate of the costs, no such distortions can arise in this specific model. Recall from our discussion in §1 that if (CRRSA) is applied to the partition generated by the (EPA) (which minimizes $\underline{P}^1$), all but at most one of the regions generated by the (CRRSA) have identical replenishment intervals (equal to $1/f^*$) and have backlogs during the same sequence of intervals. For the remaining region the replenishment interval may, of course, be modified to be $1/f^*$ and the reorder point S_1 may be reset, so that the sequence of backlog periods of this region is identical to the others' as well. This modification of course does not affect asymptotic optimality or accuracy. Thus when applying this (slightly modified) solution of the (CRRSA), the inventory of each retailer follows a sawtooth pattern even if it is assigned to more than one region and the computed average system-wide costs is exactly correct.

3. A Numerical Study

In this section we summarize a series of numerical experiments which were conducted to assess, for problems of *moderate* size the computational requirements of the (CRRSA) algorithm as well as the optimality gap of the generated solutions (or alternatively the degree of accuracy of the computed lower and upper bounds). Computation of a truly optimal solution is entirely intractable even for the smallest of the analyzed problems (with only 100 demand points). Thus, the exact optimality gap of the generated (heuristic) solutions cannot be measured. Instead, we report the ratios of the computed upper and lower bound (UL/LB) which serve as *upper bounds* for these optimality gaps.

Our conclusion from this study is that our procedures have very modest computational requirements which grow roughly *linearly* with the number of locations. For example, for a problem with 1000 demand points in which no route visits more than four distinct points, the entire solution procedure (i.e., computation of the lower bound, upper bound, routes and inventory strategies) requires no more than about 2 CPU seconds when encoded in Fortran (Tops 20—Version 2) and run on a DEC system-2065 computer. The generated solutions come within a relatively small percentage of a *lower bound* for the *minimal* system-wide costs, even for problems of moderate size. For example, for problems with 500 demand points, the average bound for the optimality gap is only 7% and for problems with 100 demand points only 14% (see Table 1).

TABLE 1
Summary Results

N	m^*	No. of Problems	UB/LB $\mu^{(*)}$	UB/LB $\sigma^{(**)}$	CPU Time (seconds) μ	CPU Time (seconds) σ
100	2	8	1.100	0.044	0.233	0.016
100	4	16	1.188	0.098	0.322	0.024
500	2	8	1.043	0.019	0.772	0.030
500	4	16	1.073	0.037	1.168	0.060
500	7	4	1.138	0.069	65.941	0.640
1000	2	8	1.031	0.013	1.592	0.077
1000	4	16	1.053	0.027	2.347	0.137
1000	7	4	1.103	0.051	137.131	1.294
5000	2	8	1.014	0.006	10.824	0.569
5000	4	16	1.025	0.012	13.786	0.408
5000	7	4	1.057	0.020	691.264	6.555
10000	2	8	1.010	0.004	26.692	0.969
10000	4	16	1.018	0.009	30.432	0.693
10000	7	4	1.034	0.017	1426.626	7.782

(*) μ indicates the mean value within the relevant category of problems.
(**) σ indicates the standard deviation of the results within the relevant category of problems.

In all of our models, the retailers' locations are randomly generated in a square of 200 $\times$ 200 with the depot placed in its center (i.e., both coordinates are generated independently from the uniform distribution on the interval $[-100, 100]$). The problem set summarized in Table 1, consists of eight distinct categories numbered I–VIII. Their characteristics are reported in Table 2. In each of the eight categories, the number of retailers is varied from $N = 100$ till $N = 10{,}000$. In all eight problem categories the retailers have identical characteristics, with a demand rate which, without loss of generality, is set equal to one. Recall that $m_l^* = \min\{M^*, N_l^*\}$ represents the upper bound for the number of demand points on the lth route ($l = 1, \ldots, L$). By choosing $N_1^* = \cdots = N_L^* = N^*$ we can henceforth drop the subscript of m^*. Only for categories V and VI where $m^* = 7$ have we omitted the runs for $n = 100$ since the generated solution would consist of only $15 = (\lceil 100/7 \rceil)$ sales regions. While the bounds for the optimality gaps are likely to be relatively large for such small problems, it should be remembered that no alternative algorithms appear to exist. All instances with a given number of retailers (N), share the same exact set of retailers $X_{(N)}$. Moreover, the sets $\{X_{(N)} : N = 100, 500, 1000, 5000, 10000\}$ are nested, i.e. $X_{(100)} \subseteq X_{(500)} \subseteq X_{(1000)} \subseteq X_{(5000)} \subseteq X_{(10000)}$. Also

TABLE 2
List of Parameter Values in Each of the Nine Categories

	I	II	III	IV	V	VI	VII	VIII	IX
b	∞	0.4, 0.8, 2.0	∞	2	∞	2	∞	0.4	∞
f^*	5	5	1.0, 10.0	1, 10.0	5	5	5	5	5
h^+	100	100	100	100	100	100	50, 150	50, 150	100
N^*	4	4	4	4	7	7	4	4	4, 22
N	100, 500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000	500, 1000, 5000, 10000	500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000	100, 500, 1000, 5000, 10000
c	10, 250	10, 250	10, 250	10, 250	10, 250	10, 250	10, 250	10, 250	10, 250
No. of problems	10	30	20	20	8	8	20	20	14

in all eight categories, the models are systematically evaluated for two different values of the fixed cost per route: $c = 10$ and $c = 250$; the former (latter) represents situations where most of the transportation costs are variable (fixed).

Table 1 summarizes the performance of the (CRRSA) algorithm in all 136 problems in these eight categories. The traveling salesman tours in the generated regions were obtained by complete enumeration. This step in the (CRRSA) algorithm accounts for the bulk of the CPU time as may be inferred by comparing the average CPU time between problem categories with identical number of retailers (N) but different values for m^*. Note that increasing m^* from 2 to 4 leads to an increase in the average CPU time by a factor of 1.3 (approximately) while increasing m^* from 4 to 7 leads to an increase by a factor of 50 (approximately). Note from the description of the algorithm that the required number of elementary operations and evaluations of the function $f(\cdot, \cdot)$ depends on N and m^* only. Moreover, except for the evaluation of the lengths of the traveling salesman tours, the computational requirements of all other parts of the algorithm actually decrease with m^*, for any given value of N! For values $m^* \geq 7$ (say) it is thus advisable to determine the optimal traveling salesman tours by a more sophisticated exact solution method. Alternatively, as pointed out in Anily and Federgruen (1987), a heuristic *TSP*-solution method with bounded worst case performance may be employed.

As with the CPU times, we observe that the ratios UB/LB are quite predictable as a function of N and m^* only. Moreover, the ratios UB/LB are quite low even for small problems with 100 demand points only. The results in Table 1 are consistent with the asymptotic optimality results, derived in §1. It is worth pointing out that the ratios UB/LB for the case of nonidentical retailers become significantly smaller than the respective ratios for identical retailers. Also, for any fixed value of N, the ratios UB/LB increase with m^*, which is consistent with the analyses in Anily and Federgruen (1987), exhibiting the error gaps as a function of the number of sales regions $L = \lceil N/m^* \rceil$. Note, in addition, that the lower bound approximation for the length of a traveling salesman tour by two times the average value of the radial distances, becomes increasingly less accurate when the number of demand points per region increases.

Category I represents our basic model. While the basic model is uncapacitated, model II investigates the impact of progressively more severe vehicle capacity constraints. In categories III and IV we vary the maximum permissible replenishment frequency in uncapacitated and capacitated models respectively. Categories V and VI are designed to assess the impact of relaxed sales volume constraints, in both types of models. Finally in categories VII and VIII we vary the holding cost rate to a value 50% higher and one 50% lower than the basic value of $h^+ = 100$, both in a model where the maximum number of retailers is determined by the sales volume constraint (category VII) and one in which this number is determined by the combination of the vehicle capacity and frequency constraints (category VIII).

In each of the Tables 3–7, we list for each of the scenarios the following performance measures:

$m^* = \min\{N^*, \lfloor f^* b/\mu \rfloor\}$ = the number of retailers in all but at most one sales region,

LB = value of the lower bound $\underline{V}^1(X)$,

UB = long-run average system-wide costs of the generated solution,

UB/LB = upper bound on the achieved optimality gap,

TRC = average transportation related costs,

NR = number of sales regions in the generated solution,

CPU = CPU time in seconds.

Consider first the results of Category I. Recall that for two instances with identical values for N (hence with identical sets of retailers) and identical values for m, the (CRRSA) procedure generates the same collection of sales regions. Thus any such pair of instances

TABLE 3
Category I

N	c	m^*	LB	UB	UB/LB	TRC	NR	CPU Time
100	10.0	4	8780.24	10585.07	1.205	5292.53	25	0.365
100	250.0	4	14106.05	15289.98	1.084	7644.99	25	0.347
500	10.0	4	44032.37	47731.05	1.084	23865.52	125	1.141
500	250.0	4	70643.97	72951.26	1.033	36475.63	125	1.124
1000	10.0	4	88813.46	94257.72	1.061	47128.86	250	2.285
1000	250.0	4	141703.82	145109.48	1.024	72554.74	250	2.269
5000	10.0	4	442723.61	455525.67	1.028	227762.83	1250	13.671
5000	250.0	4	707695.64	715546.98	1.011	357773.49	1250	13.550
10000	10.0	4	884969.73	903708.23	1.021	451854.14	2500	30.063
10000	250.0	4	1415056.69	1426496.89	1.008	713248.44	2500	30.071

share the same values for NR. This suggests that both LB and UB increase by similar amounts (hence decreasing the ratio UB/LB) when increasing the value of c (the fixed cost per route), even though the cost function $f(\cdot, \cdot)$ is nonlinear in this parameter. This hypothesis is confirmed by the results in Table 3.

Note that the ratios UB/LB decrease by an approximate factor of 2.5, 1.35, 2.15 and

TABLE 4
Category II

N	b	c	m^*	LB	UB	UB/LB	TRC	NR	CPU
100	0.4	10.0	2	41015.62	47364.43	1.155	46364.43	50	0.264
100	0.8	10.0	4	21007.81	29812.72	1.419	28812.72	25	0.316
100	2.0	10.0	4	10502.48	14025.09	1.335	11525.09	25	0.309
100	0.4	250.0	2	101015.62	107364.43	1.063	106364.43	50	0.229
100	0.8	250.0	4	51007.81	59812.72	1.173	58812.72	25	0.314
100	2.0	250.0	4	22503.13	26025.09	1.157	23525.09	25	0.328
500	0.4	10.0	2	206824.66	220672.69	1.067	215672.69	250	0.771
500	0.8	10.0	4	105912.33	122546.70	1.157	117546.70	125	1.167
500	2.0	10.0	4	52850.99	59518.68	1.126	47018.68	125	1.160
500	0.4	250.0	2	506824.68	520672.72	1.027	515672.72	250	0.757
500	0.8	250.0	4	255912.33	272546.70	1.065	267546.70	125	1.155
500	2.0	250.0	4	112864.93	119518.68	1.059	107018.68	125	1.154
1000	0.4	10.0	2	419259.05	439066.25	1.047	429066.25	500	1.590
1000	0.8	10.0	4	214629.52	239169.95	1.114	229169.95	250	2.308
1000	2.0	10.0	4	106825.50	116666.82	1.092	91682.03	250	2.288
1000	0.4	250.0	2	1019258.97	1039066.21	1.019	1029066.21	500	1.597
1000	0.8	250.0	4	514629.54	539169.96	1.048	529169.98	250	2.265
1000	2.0	250.0	4	226851.80	236667.98	1.043	211667.98	250	2.333
5000	0.4	10.0	2	2084403.80	2129292.91	1.022	2079291.89	2500	10.680
5000	0.8	10.0	4	1067201.98	1123249.59	1.053	1073249.61	1250	13.680
5000	2.0	10.0	4	531742.18	554277.64	1.042	429965.83	1250	13.658
5000	0.4	250.0	2	5084403.81	5129292.00	1.009	5079292.12	2500	10.609
5000	0.8	250.0	4	2567201.78	2623249.50	1.022	2573249.41	1250	13.799
5000	2.0	250.0	4	1131880.78	1154300.02	1.020	1029300.01	1250	13.571
10000	0.4	10.0	2	4163597.97	4228082.31	1.015	4128082.34	5000	25.475
10000	0.8	10.0	4	2131798.97	2213326.53	1.038	2113326.66	2500	29.907
10000	2.0	10.0	4	1062428.59	1095247.41	1.031	845808.60	2500	30.083
10000	0.4	250.0	2	10163597.12	10228082.12	1.006	10128082.12	5000	25.490
10000	0.8	250.0	4	5131798.25	5213326.87	1.016	5113326.50	2500	29.949
10000	2.0	250.0	4	2262719.56	2295330.75	1.014	2045330.56	2500	29.493

1.35 when increasing N from 100 to 500, from 500 to 1000, 1000 to 5000 and 5000 to 10000, respectively. Observe, in addition, that the CPU time increases roughly linearly with N. The above observations regarding (CRRSA)'s computational requirements and the sensitivity of its performance with respect to variations in the parameters and N hold for all other categories of scenarios as well, see Tables 4–7. The same or similar explanations may be brought to bear.

Turning next to Category II, note from Table 4 that (UB − LB)/UB more than doubles when b increases from 0.4 to 0.8 but decreases as the vehicle capacity is increased from $b = 0.8$ to $b = 2$ and from $b = 2$ to $b = \infty$. (For $b = \infty$, cf. the results in Table 3.) This pattern may be explained as follows: when b is increased from 0.4 to 0.8, m^* (the number of retailers in each region) increases from 2 to 4. Clearly the lower bound

TABLE 5
Categories III/IV

N	c	f^*	m^*	LB	UB	UB/LB	TRC	b	NR	CPU
100	10.0	1.0	4	8990.97	10641.06	1.184	5103.29	∞	25	0.290
100	10.0	10.0	4	8780.24	10585.07	1.206	5292.53	∞	25	0.297
100	250.0	1.0	4	14106.05	15289.98	1.084	7644.99	∞	25	0.301
100	250.0	10.0	4	14106.05	15289.98	1.084	7644.99	∞	25	0.293
500	10.0	1.0	4	45130.47	48346.36	1.071	22146.53	∞	125	1.121
500	10.0	10.0	4	44032.37	47731.05	1.084	23865.52	∞	125	1.134
500	250.0	1.0	4	70643.97	72951.26	1.033	36475.63	∞	125	1.151
500	250.0	10.0	4	70643.97	72951.26	1.033	36475.63	∞	125	1.185
1000	10.0	1.0	4	90817.19	95584.33	1.052	43571.20	∞	250	2.306
1000	10.0	10.0	4	88813.46	94257.72	1.061	47128.86	∞	250	2.230
1000	250.0	1.0	4	141703.82	145108.48	1.024	72554.74	∞	250	2.279
1000	250.0	10.0	4	141703.82	145109.48	1.024	72554.74	∞	250	2.311
5000	10.0	1.0	4	452952.68	463913.43	1.024	207254.54	∞	1250	14.00
5000	10.0	10.0	4	442723.61	455525.67	1.029	227762.83	∞	1250	13.617
5000	250.0	1.0	4	707695.64	715546.98	1.011	357773.49	∞	1250	13.800
5000	250.0	10.0	4	707695.64	715546.98	1.011	357773.49	∞	1250	13.781
10000	10.0	1.0	4	905423.09	921385.44	1.018	409457.41	∞	2500	29.742
10000	10.0	10.0	4	584969.73	903708.28	1.021	451854.14	∞	2500	30.321
10000	250.0	1.0	4	1415056.69	1426496.89	1.008	713248.44	∞	2500	30.546
10000	250.0	10.0	4	1415056.69	1426496.89	1.008	713248.44	∞	2500	30.372
100	10.0	1.0	2	13003.13	14272.89	1.098	9272.89	2	50	0.253
100	10.0	10.0	4	10502.48	14025.09	1.336	11525.09	2	25	0.337
100	250.0	1.0	2	25003.13	26272.89	1.051	21272.89	2	50	0.239
100	250.0	10.0	4	22503.13	26025.09	1.157	23525.09	2	25	0.346
500	10.0	1.0	2	65364.93	68134.53	1.042	43134.54	2	250	0.805
500	10.0	10.0	4	52850.99	59518.68	1.126	47018.68	2	125	1.253
500	250.0	1.0	2	125364.94	128134.54	1.022	103134.54	2	250	0.832
500	250.0	10.0	4	112864.93	119518.68	1.059	107018.68	2	125	1.368
1000	10.0	1.0	2	131851.82	135813.24	1.030	85813.24	2	500	1.652
1000	10.0	10.0	4	106825.50	116666.82	1.092	91682.03	2	250	2.543
1000	250.0	1.0	2	251851.83	255813.24	1.016	205813.25	2	500	1.753
1000	250.0	10.0	4	226851.80	236667.98	1.043	211667.98	2	250	2.799
5000	10.0	1.0	2	656880.86	665858.49	1.014	415858.32	2	2500	12.230
5000	10.0	10.0	4	531742.18	554277.64	1.042	429465.83	2	1250	14.954
5000	250.0	1.0	2	1256880.59	1265858.42	1.007	1015858.45	2	2500	11.105
5000	250.0	10.0	4	1131880.78	1154300.02	1.020	1029300.01	2	1250	13.412
10000	10.0	1.0	2	1312719.25	1325616.64	1.010	825616.61	2	5000	27.142
10000	10.0	10.0	4	1062428.59	1095247.41	1.031	845808.59	2	2500	31.710
10000	250.0	1.0	2	2512719.50	2525616.75	1.005	2025616.56	2	5000	28.615
10000	250.0	10.0	4	2262719.56	2295330.75	1.014	2045330.56	2	2500	32.336

approximation for the length of a region's optimal route, by 2 times the average value of the radial distances ($TSP(X_l^0) \geq 2 \sum_{i \in X_l} r_i / |X_l|$) is exact when $m^* = 1$ and becomes progressively worse as m^* increases. When b is larger than 0.8 the sales volume constraint becomes the determinant of m^*, which therefore remains equal to 4. As the vehicle capacity is increased, both ratios TRC/UB and UB/LB decrease.

With category III, we investigate the impact of the parameter f^* on the performance of the algorithm. Recall that in category I, with $f^* = 5$, the frequency constraints are redundant in all scenarios and for all of the generated sales regions. The same is true, a fortiori, when $f^* = 10$. The solutions under $f^* = 5$ and $f^* = 10$ are therefore identical, cf. Table 3 and Table 5. For $f^* = 1$, the frequency constraints are binding for some (though not all) sales regions. Decreasing f^* from 10 to 1 results in an increase in the cost of the generated solution (UB) and an even bigger increase in the value of the lower bound (LB) thus reducing the ratios UB/LB.

In category IV, the vehicle capacity b is held constant with $b = 2$. As pointed out above, with $f^* = 5$ (category II), the capacity constraint has no impact on m^*, the number of retailers per region. In this case the vehicle capacity constraints do, however, determine the replenishment intervals (and hence quantities), for all sales regions when $c = 250$, and for a majority of sales regions when $c = 10$. Recall also that when $f^* = 5$ the frequency constraints are redundant. This holds a fortiori when $f^* = 10$ so that the same solutions are generated under $f^* = 5$ (Table 4) and $f^* = 10$ (Table 5).

When f^* is reduced to 1, the combination of the frequency and vehicle capacity constraints become the bottleneck in the determination of the maximum permissible number of retailers per region (which is the actual number of retailers in all generated regions, since N is a multiple of 2, throughout, and m^* is reduced from 4 to 2). It follows from (3) that for $c = 250$, the vehicle capacity constraints are binding for the determination of all the regions' replenishment intervals. For $c = 10$, only one replenishment interval is feasible for each region, see (1), i.e. the lower bounds imposed by the frequency constraints coincide with the upper bounds imposed by the capacity constraints. Since a reduction of f^* from 10 to 1 results in a reduction of m^* from 4 to 2, it is to be expected that the ratios UB/LB decrease (see our comments with respect to category II for the appropriate arguments).

TABLE 6

Categories V/VI

N	c	m^*	LB	UB	UB/LB	TRC	b	NR	CPU
500	10.0	7	33325.32	38065.99	1.142	19032.99	∞	72	66.552
500	250.0	7	53538.83	56565.54	1.057	28282.77	∞	72	66.616
1000	10.0	7	67171.15	74337.87	1.107	37168.93	∞	143	136.600
1000	250.0	7	107166.27	111726.88	1.043	55863.44	∞	143	136.972
5000	10.0	7	334701.98	351635.47	1.051	175817.73	∞	715	685.507
5000	250.0	7	535115.86	545647.62	1.020	272823.81	∞	715	684.771
10000	10.0	7	669002.41	693237.12	1.036	346618.56	∞	1429	1424.152
10000	250.0	7	1069792.94	1084743.55	1.014	542371.77	∞	1429	1438.482
500	10.0	7	47561.07	59236.41	1.245	52071.32	2	72	65.143
500	250.0	7	107573.56	119248.59	1.109	112048.59	2	72	65.483
1000	10.0	7	96169.81	113702.69	1.182	99402.69	2	143	135.724
1000	250.0	7	216151.82	233702.69	1.081	219402.69	2	143	139.23
5000	10.0	7	478329.38	518342.34	1.084	446905.94	2	715	694.109
5000	250.0	7	1078380.78	1118382.78	1.037	1046882.80	2	715	700.671
10000	10.0	7	955578.83	1012178.02	1.059	869320.62	2	1429	1427.011
10000	250.0	7	2155621.19	2212196.19	1.026	2069296.19	2	1429	1416.86

TABLE 7
Categories VII/VIII

N	c	h^+	m^*	LB	UB	UB/LB	TRC	b	NR	CPU
100	10.0	50.0	4	6208.57	7484.77	1.206	3742.39	∞	25	0.367
100	10.0	150.0	4	10753.56	12964.00	1.206	6482.00	∞	25	0.302
100	250.0	50.0	4	9974.48	10811.65	1.084	5405.82	∞	25	0.302
100	250.0	150.0	4	17276.31	18726.32	1.084	9363.16	∞	25	0.333
500	10.0	50.0	4	31135.59	33750.95	1.084	16875.47	∞	125	1.165
500	10.0	150.0	4	53928.42	58458.36	1.084	29229.18	∞	125	1.143
500	250.0	50.0	4	49952.83	51584.33	1.033	25792.16	∞	125	1.137
500	250.0	150.0	4	86520.84	89346.68	1.033	44673.34	∞	125	1.133
1000	10.0	50.0	4	62800.60	66650.27	1.061	33325.14	∞	250	2.303
1000	10.0	150.0	4	108773.83	115441.66	1.061	57720.83	∞	250	2.427
1000	250.0	50.0	4	100199.73	102607.90	1.024	51303.95	∞	250	2.341
1000	250.0	150.0	4	173551.04	177722.10	1.024	88861.05	∞	250	2.272
5000	10.0	50.0	4	313052.87	322105.27	1.029	161052.64	∞	1250	14.566
5000	10.0	150.0	4	542223.51	557902.70	1.029	278951.34	∞	1250	13.362
5000	250.0	50.0	4	500416.37	505968.15	1.011	252984.07	∞	1250	13.765
5000	250.0	150.0	4	866746.56	876362.48	1.011	438181.23	∞	1250	13.390
10000	10.0	50.0	4	625768.05	639018.36	1.021	319509.18	∞	2500	30.450
10000	10.0	150.0	4	1083862.25	1106812.09	1.021	553406.05	∞	2500	30.685
10000	250.0	50.0	4	1000596.36	1008685.55	1.008	504342.77	∞	2500	30.495
10000	250.0	150.0	4	1733083.44	1747094.66	1.008	873547.33	∞	2500	30.685
100	10.0	50.0	2	40515.62	46864.43	1.157	46364.43	0.4	50	0.217
100	10.0	150.0	2	41515.63	47864.43	1.153	46364.43	0.4	50	0.217
100	250.0	50.0	2	100515.63	106864.43	1.063	106364.43	0.4	50	0.226
100	250.0	150.0	2	101515.62	107864.43	1.063	106364.43	0.4	50	0.220
500	10.0	50.0	2	204324.66	218172.69	1.068	215672.69	0.4	250	0.761
500	10.0	150.0	2	209324.66	223172.69	1.066	215672.69	0.4	250	0.743
500	250.0	50.0	2	504324.68	518172.72	1.027	515672.72	0.4	250	0.737
500	250.0	150.0	2	509324.68	523172.72	1.027	515672.72	0.4	250	0.767
1000	10.0	50.0	2	414259.06	434066.24	1.048	429066.25	0.4	500	1.555
1000	10.0	150.0	2	424259.06	444066.25	1.047	429066.25	0.4	500	1.587
1000	250.0	50.0	2	1014258.97	1034066.21	1.020	1029066.21	0.4	500	1.496
1000	250.0	150.0	2	1024258.97	1044066.21	1.019	1029066.21	0.4	500	1.510
5000	10.0	50.0	2	2059403.77	2104291.87	1.022	2079291.89	0.4	2500	10.519
5000	10.0	150.0	2	2109403.84	2154291.91	1.021	2079291.89	0.4	2500	10.448
5000	250.0	50.0	2	5059403.75	5104292.06	1.009	5079292.12	0.4	2500	10.388
5000	250.0	150.0	2	5109403.69	5154292.06	1.009	5079292.12	0.4	2500	10.617
10000	10.0	50.0	2	4113598.00	4178082.34	1.016	4128082.34	0.4	5000	26.257
10000	10.0	150.0	2	4213597.87	4278082.37	1.015	4128082.34	0.4	5000	26.523
10000	250.0	50.0	2	10113597.12	10178081.87	1.006	10128082.12	0.4	5000	27.380
10000	250.0	150.0	2	10213597.37	10278082.12	1.006	10128082.12	0.4	5000	26.654

In comparing the results of category V with those in category I we note an increase in the ratios UB/LB, to be expected from the increase of m^* from 4 to 7 (see again our comments with respect to category II for the appropriate arguments). Note that each of the considered values of N fails to be a multiple of 7 so that in each case all but one of the generated regions have 7 retailers. CPU times are more than 60 times larger than the corresponding values in category I. As explained above, this is entirely due to the increased effort required for the determination, by a full enumeration scheme, of the optimal routes for the sales regions. Significant improvements may be achieved as explained above.

Category VI ought to be compared with the scenarios in category II for which $b = 2$. Once again, the increase from $m^* = 4$ to $m^* = 7$ results in an increase of the ratios UB/LB, CPU times are comparable to the ones obtained for category V, and again more than 60 times larger than in category II. The frequency constraints remain redundant except when $c = 10$ in the extremely unlikely event when the length of a region's optimal route is less than 4. The vehicle capacity constraints are now virtually always binding even when $c = 10$. As before, we notice somewhat increased ratios UB/LB when comparing the capacitated scenarios in category VI with the corresponding ones in category V.

Categories VII and VIII investigate the sensitivity with respect to h^+, the holding cost rate. Note first that this parameter has no impact on the generated collection of sales regions. Thus the values of NR and TD are identical to the corresponding ones in category I and II (the latter with $b = 0.4$).

Recall that in category I, with $h^+ = 100$, the frequency constraints are always redundant. This remains true, a fortiori when $h^+ = 50$. When $h^+ = 150$ the frequency constraint is binding only when $c = 10$, and only for extremely unlikely regions with an optimal route length of 2 or less. Since the models in category VII are uncapacitated, this suggests that the unconstrained replenishment intervals apply for all sales regions. This is confirmed by the results in Table 7. Our observations also suggest that both, in the lower bound (LB) and the upper bound (UB), the cost of *each* region increases (decreases) by a factor of $\sqrt{1.5}$ ($\sqrt{0.5}$) when h^+ is increased (decreased) from 100 to 150 (50). This explains why the ratios UB/LB are constant in h^+ over the interval [50, 150]. In category VIII, $m = 2$. The frequency constraints are always redundant and the capacity constraints are always binding. It thus follows that the cost of a given region, both in LB and UB, is *linear in but not proportional* with h^+. Nevertheless this observation suggests that the ratios UB/LB would be rather insensitive with respect to variations in this parameter. This is confirmed by the results in Tables 4 and 7.

In category IX we investigate the performance of the (CRRSA) procedure for systems with nonidentical retailers. The scenarios in this category are identical to the ones in category I except that the demand rates of the retailers are generated from a uniform distribution over the integers $\{1, \ldots, 10\}$. Since each retailer consists in average of 5.5 demand points, the ratios UB/LB are clearly expected to be smaller as compared to the corresponding cases with the same number of retailers each consisting of a single demand point. The ratios for scenarios with 100 (1000) retailers are in fact significantly lower than the ratios for the corresponding scenarios with 500 (5000) retailers in category I, even though the number of demand points are comparable (550 and 5500 respectively). This is explained by the fact that in category IX most regions, even though consisting of

TABLE 8

Category IX

Number of Retailers	Number of Demand Points	c	m^*	LB	UB	UB/LB	TRC	NR
100	569	10.0	4	51108.83	52587.98	1.029	26293.99	143
100	569	250.0	4	81088.78	82022.13	1.011	41011.06	143
500	2797	10.0	4	247111.35	251168.77	1.016	125584.38	670
500	2797	250.0	4	395829.59	398269.68	1.006	199134.84	670
1000	5497	10.0	4	487537.15	493030.05	1.011	246515.03	1375
1000	5497	250.0	4	778567.20	781889.65	1.004	390944.82	1375
5000	27358	10.0	4	2424421.44	2437950.19	1.006	1218975.09	6840
5000	27358	250.0	4	3873930.41	3881882.16	1.002	1940941.08	6840
10000	55019	10.0	4	4887563.06	4906828.87	1.004	2453414.44	13755
10000	55019	250.0	4	7796925.06	7808168.25	1.001	3904084.12	13755
100	593	10.0	22	22355.62	26883.44	1.203	13441.72	27
100	593	250.0	22	35797.55	38740.15	1.082	19370.08	27
500	2783	10.0	22	105495.88	115546.18	1.095	57773.09	127
500	2783	250.0	22	168362.16	174753.27	1.038	87376.63	127

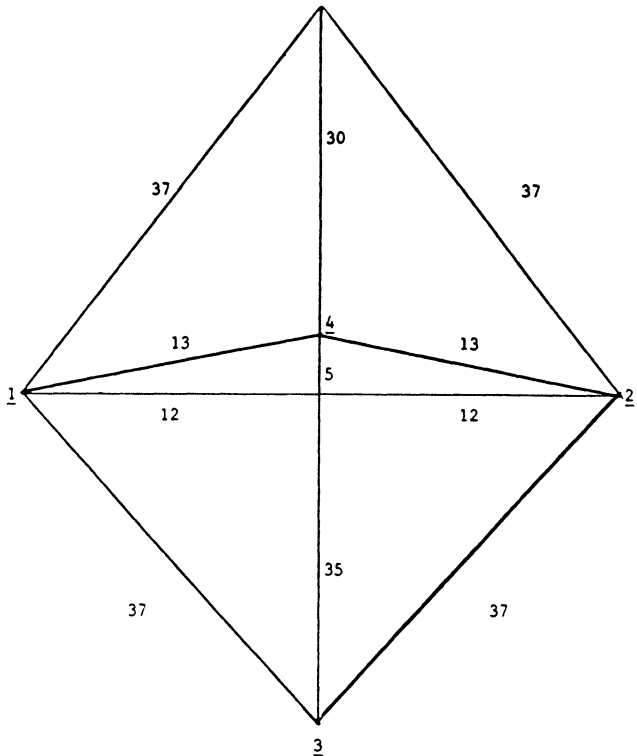

FIGURE 1

four demand points, correspond with only one or two retailers at distinct locations. The average length of the route is thus much smaller and the length of the optimal route for a given region is typically quite close to 2 times the average value of the radial distances in the region, the value used in the lower bound approximation. Another interesting comparison with identical retailer systems arises, when increasing N^* (and hence m^*) by a factor of 5.5 (the expected number of demand points per retailer). See the last four cases in Table 8 and compare with the scenarios in category I with the same number of retailers and c-values to conclude that the optimality gaps are virtually identical. Note that in these four cases the number of retailers per region potentially varies between 3 and 22.

See Anily and Federgruen (1988) for detailed observations and explanations of the sensitivity of the TRC/UB ratio in the different problem categories. The tables ibid give additional characteristics of the generated solutions.[1]

[1] The research of the first author has been partially supported by NSERC grant no. A4802.

The research of the second author has been partially supported by NSF grant no. ECS-8604409 as well as a grant of the Lady Davis Foundation.

Appendix

The following example shows that $TSP(S)$, the cost of servicing a given region $S \subset X$, fails to be submodular when viewed as a set function defined on the collection of subsets of X:

Example

Consider a problem with one warehouse (node $\underline{O}$) and four retailers (numbered from $\underline{1}$ to $\underline{4}$). Their locations are represented in Figure 1. The numbers associated with drawn links represent the Euclidean distances between their end points. Let $K(S)$ denote the length of the optimal traveling salesman tour through S, $S_2\{\underline{0}, \underline{1}, \underline{2}, \ldots\}$.

Note that $K(\{\underline{0}, \underline{1}, \underline{2}, \underline{4}\}) = 100$ (achieved by the tour $\underline{0}$-$\underline{1}$-$\underline{4}$-$\underline{2}$-$\underline{0}$), $K(\{\underline{0}, \underline{1}, \underline{2}, \underline{3}\}) = 148$ (achieved by the tour $\underline{0}$-$\underline{1}$-$\underline{3}$-$\underline{2}$-$\underline{0}$), $K(\{\underline{0}, \underline{1}, \underline{2}, \underline{3}, \underline{4}\}) = 154$ (achieved by the tour $\underline{0}$-$\underline{1}$-$\underline{3}$-$\underline{2}$-$\underline{4}$-$\underline{0}$). Thus, $K(\{\underline{0}, \underline{1}, \underline{2}, \underline{3}, \underline{4}\}) - K(\{\underline{0}, \underline{1}, \underline{2}, \underline{3}\}) > K(\{\underline{0}, \underline{1}, \underline{2}, \underline{4}\}) - K(\{\underline{0}, \underline{1}, \underline{2}\})$, violating submodularity.

References

ANILY, S., "Integrating Inventory Control and Transportation Planning," Ph.D. dissertation, Columbia University, New York, 1986.

——— AND A. FEDERGRUEN, "Structured Partitioning Problems." Working Paper, The Graduate School of Business, Columbia University, New York, 1986. *Oper. Res.* (to appear).

——— AND ———, "Class of Euclidean Routing Problems with General Route Cost Functions," Working Paper, The Faculty of Commerce and Business Administration, The University of British Columbia, Vancouver, (1987). *Math. Oper. Res.* (to appear).

——— AND ———, "One Warehouse Multiple Retailer Systems with Vehicle Routing Costs," Working Paper, The Faculty of Commerce and Business Administration, The University of British Columbia, Vancouver, 1988 (unabridged version).

ASSAD, A., B. GOLDEN, R. DAHL AND M. DROR, "Evaluating the Effectiveness of an Integrated System for Fuel Delivery," in *Proc. TIMS, Southeastern Nineteenth Annual Meeting*, Eatman (Ed.), 1983, 153–160.

BARNES, E. R., A. J. HOFFMAN, AND U. G. ROTHBLUM, "Optimal Partitions Having Disjoint Convex and Conic Hulls," Technion, Faculty of Ind. Engineering and Management Working Paper, Haifa, Israel, 1989.

BELL, W., L. M. DALBERTO, M. L. FISHER, A. J. GREENFIELD, R. JAIKUMAR, P. KEDIA, R. G. MACK AND P. J. PLUTZMAN, "Improving the Distribution of Industrial Gases with an On-line Computerized Routing and Scheduling Optimizer," *Interfaces* 13 (1983), 4–23.

BROWN, R. G. (1967), *Decision Rules for Inventory Management*, Holt, Rinehart, and Winston, New York, 1967, 46–55.

BURNS, L. D., R. W. HALL, D. E. BLUMENFELD AND C. F. DAGANZO, "Distribution Strategies That Minimize Transportation and Inventory Costs," *Oper. Res.*, 33 (1985), 469–490.

CARR, C. R. AND C. W. HOWE, "Optimal Service Policies and Finite Time Horizons," *Management Sci.*, 9 (1962), 126–140.

CHAKRAVARTY, A. K., "Multi-Stage Production/Inventory Deterministic Lot-Size Computations," *Internat. J. Production Res.*, 22 (May/June 1984a), 405–420.

———, "Group-Synchronized Replenishment of Inventory Items," *Engineering Costs and Production Economics*, 8 (July 1984b), 1–2, 129–134.

———, "Deterministic Lot-Sizing for Coordinated Families of Production/Inventory Items," *European J. Oper. Res.*, 17, (2), 207–214.

——— AND S. K. GOYAL, "Multi-Item Inventory Grouping with Dependent Set-Up Cost and Group Overhead Cost," *Engineering Costs and Production Economics*, 10 (March 1986), 13–24.

———, J. B. ORLIN AND U. G. ROTHBLUM, "A Partitioning Problem with Additive Objective with an Application to Optimal Inventory Groupings for Joint Replenishment," *Oper. Res.*, 30 (1982), 1018–1020.

———, ———, AND ———, "Consecutive Optimizers for a Partitioning Problem with Applications to Optimal Inventory Groupings for Joint Replenishment," *Oper. Res.*, 33 (1985), 820–834.

CHRISTOFIDES, N., "Worst-case Analysis of a New Heuristic for the Traveling Salesman Problem," Report 388, Graduate School of Industrial Administration, Carnegie Mellon University, 1976.

DROR, M., M. BALL AND B. GOLDEN, "A Computational Comparison of Algorithms for the Inventory Routing Problem," *Ann. Oper. Res.*, 4 (1986), 3–23.

——— AND ———, "Inventory/Routing: Reduction from an Annual to a Short-Period Problem," *Naval Res. Logist. Quart.*, 34 (1987), 891–905.

——— AND L. LEVY, "A Vehicle Routing Improvement Algorithm-Comparison of a "Greedy" and Matching Implementation for Inventory Routing," *Computers and Oper. Res.*, 13 (1986), 33–45.

FEDERGRUEN, A., G. PRASTACOS AND P. ZIPKIN, "An Allocation and Distribution Model for Perishable Products," *Oper. Res.*, 34 (1986), 75–82.

———, A. RINNOOY KAN AND P. ZIPKIN, "Combined Routing and Inventory Allocation Problems: Bounds and Heuristics," Working paper, Graduate School of Business, Columbia University, New York, 1985.

——— AND Y. ZHENG, "The Joint Replenishment Problem with General Joint Cost Structures: General Solution Methods and Performance Bounds," Working Paper, 1988.

——— AND P. ZIPKIN, "A Combined Vehicle Routing and Inventory Allocation Problem," *Oper. Res.*, 32 (1984), 1019–1037.

GOLDEN, B., "Analysis of Large Scale Vehicle Routing Problem with an Inventory Component," *Large Scale Systems*, (1984).

——— AND T. MAGNANTI, *Network Optimization*, (1987) (in preparation).

GOYAL, S., "Determination of Economic Packaging Frequency for Items Jointly Replenished," *Management Sci.*, 20 (1973), 232–235.

———, "Optimum Ordering Policy for a Multi-Item Single Supplier System," *Oper. Res. Quart.*, 25 (1974a), 293–298.

———, "Determination of Optimum Packaging Frequency of Items Jointly Replenished," *Management Sci.*, 21 (1974b), 436–443.

———, "On Improving Silver's Algorithm for the Joint Replenishment Problem," *IIE Trans.*, 17, 1 (March 1985), 99–101.

——— AND A. BELTON, "On a Single Method of Determining Order Quantities in Joint Replenishments Under Deterministic Demand," *Management Sci.*, 26 (1979), 604.

——— AND A. SONI, "Economic Packaging Frequency of Jointly Replenished Items with Multiple Manufacturing and Packaging Cycles," Presented at the Third Internat. Sympos. of Inventories, Budapest, Hungary, August 1984.

GRAVES, S. G., "On the Deterministic Demand Multi-Product Single Machine Lot-Scheduling Problem," *Management Sci.*, 25 (1979), 276–280.

HADLEY, G. AND T. WHITIN, *Analysis of Inventory Systems*, Prentice Hall, Englewood Cliffs, NJ, 1963.

JABBONSKY, L., "The Shelf is Obsolete," *Beverage World*, (October issue 1987), 72–76.

JACKSON, P. L., W. L. MAXWELL, AND J. A. MUCKSTADT, "The Joint Replenishment Problem with a Powers of Two Restriction," *IIE Trans.*, 17 (1985), 25–32.

NOCTURNE, D. J., "Economic Ordering Frequency for Several Items Jointly Replenished," *Management Sci.*, 19 (1973), 1093–1096.

O'NEIL, B. AND K. MEEGAN, "Soft Drink Reroutes in Distribution," *Beverage World*, (October issue 1987), 204–215.

QUEYRANNE, M., "A Polynomial-Time Submodular Extension to Roundy's 98% Effective Heuristic for Production/Inventory Systems," Working paper no. 1136, University of British Columbia, Vancouver, 1985.

ROSENFIELD, D. AND M. PENDROCK, "The Effects of Warehouse Configuration Design on Inventory Levels and Holding Costs," *Sloan Management Rev.*, 21, 4 (1980), 21–33.

ROUNDY, R., "A 98% Effective Lot-Sizing Rule for a Multi-Product, Multi-Stage Production Inventory System," *Math. Oper. Res.*, 11 (1986), 699–727.

SCHWARZ, L., *Multi-level Production/Inventory Control Systems: Theory and Practice*, North Holland, Amsterdam, 1981.

SCHWEITZER, P. J. AND E. A. SILVER, "Mathematical Pitfalls in the One Machine Multiproduct Economic Lot Scheduling Problem," *Oper. Res.*, 31 (1983), 401–405.

SHU, F. T., "Economic Ordering Frequency for Two Items Jointly Replenished," *Management Sci.*, 17 (1971), 406–410.

SILVER, E. A., "A Simple Method of Determining Order Quantities in Joint Replenishment Under Deterministic Demand," *Management Sci.*, 22 (1976), 1351–1361.

ZHENG, Y. S., "Replenishment Strategies for Production/Distribution Networks with General Setup Costs," Ph.D. dissertation, Graduate School of Business, Columbia University, 1987.