CAPACITATED TWO-STAGE MULTI-ITEM PRODUCTION/ INVENTORY MODEL WITH JOINT SETUP COSTS

S. ANILY
Tel Aviv University, Tel Aviv, Israel

A. FEDERGRUEN
Columbia University, New York, New York

(Received March 1987; revisions received January 1989, January 1990; accepted January 1990)

We analyze a continuous-time, two-stage production/inventory system. In the first stage, a common intermediate product is produced in batches, and possibly stored. In the second phase, the intermediate product is fabricated into n distinct finished products. Several finished products may be included in a single production batch of *limited capacity* to exploit economies of scale. We propose a planning methodology to address the *combined* problem of joint setup costs and capacity limits (per setup). We restrict ourselves to a class of replenishment strategies with the following properties: a replenishment strategy specifies a collection of families (subsets of items) covering all end-items; if an item belongs to several families a specific fraction of its sales is assigned to each family. Each time the inventory of one item in a family is replenished, the inventories of all other items in the family are replenished as well. We derive a simple (roughly $0(n \log n)$) algorithm that results in a strategy whose long-run average cost comes within a few percentage points of a lower bound for the minimum achievable cost (within the above class of strategies).

Over the last couple of years, important progress has been made in the development of efficiently solvable inventory planning models for deterministic multi-echelon systems with *batch* production runs and batch distribution activities; see e.g., Roundy (1985, 1986), Maxwell and Muckstadt (1985), Zheng (1987) and a recent survey by Muckstadt and Roundy (1988). The cost structure in these, as well as virtually all other, inventory models consists of: i) inventory carrying costs, assumed to be proportional with the inventory levels of all relevant work-in-process and finished goods items; ii) variable production costs, assumed to be proportional with the production (shipment) volumes of the individual items; and iii) a fixed cost structure reflecting the costs of setups of production runs or distribution activities.

As is well known from the simplest of the above models (the well known single item Economic Order Quantity model), optimal replenishment frequencies, and hence, average inventory levels, depend critically on the assumed fixed cost parameters, and it is therefore of crucial importance that an adequate representation of the setup cost structure be employed. All of the above mentioned inventory models assume that a *single* setup suffices for the generation of an *unlimited* production run or analogously that a single fixed cost suffices for the shipment of an unlimited volume. This restriction applies both to existing models with a *separable* setup cost structure, in which it is assumed that all items are replenished on an *individual* basis, incurring a fixed (item specific) cost per setup as well as those with nonseparable *joint* cost structures (Roundy 1986, Zheng 1987, Federgruen and Zheng 1988). Such joint cost structures reflect economies of scale that may be exploited when different items are combined in the same production batch or by performing several operations or distribution activities together.

For many practical production and distribution activities, a single setup merely suffices to cover an activity volume up to a given *capacity* limit. In other words, the setup cost is a *step function* of the activity's volume instead of the conventional modeling assumption of a *flat* setup cost curve. This situation occurs, for example, in the following cases:

i. Many a production or distribution activity is performed with equipment of limited physical capacity. Examples include production vessels or ovens of limited volume in which blending, drying or cooking operations are performed (e.g., chemical or pharmaceutical production processes), automatic guided vehicles (AGVs) used to transfer items from one production stage to the next, and limited capacity trucks used to ship items from one level in the distribution network to the next. In all these examples, different items may often be combined into a single production run (or shipment). This is sometimes achieved by compartmentalization of the vessels, ovens, AGVs or trucks.

Subject classification: Inventory/production: multi-item, echelon stage: policies for two-stage capacitated multi-item models.

0030-364X/91/3902-0000 $01.25

ii. Many production processes are interrupted after a given number of production hours or after a given number of units produced, so as to perform a maintenance or cleaning job, to assess the quality of the units produced, to readjust the equipment or to replace tools. Each interruption period represents a setup so that the total setup cost of a production run varies with the total production volume as a step function.

iii. Production orders are often assigned to several parallel machines, production lines or employee pools to ensure that the entire order be completed within a specified lead-time limit. This lead-time limit translates into a capacity limit for the volume that can be produced on a single *production unit* with a single setup.

If all items are produced on an *individual* basis, i.e., if the setup cost structure is *separable* across the items, it is relatively easy to incorporate the above capacity limits into existing planning models; see e.g., Zheng (Chapter 6), where this objective is achieved for production/distribution networks of general topology representing general bills of materials. It appears, however, that existing methodologies are incapable of incorporating *capacity limits* when different items may be combined into a single production run, i.e., when the setup cost structure fails to be separable among the items.

In this paper, we propose a planning methodology to address the combined problem of joint setup costs and capacity limits (per setup) for a continuous-time, two-stage, multi-item production or distribution system. (To avoid repetitious statements of the analogies between production and distribution planning problems, we cast the remainder of our description entirely within the production sphere.) We first describe the assumptions of this model.

In the first stage, a common intermediate product is produced in batches and possibly stored. In the second phase, the intermediate product is fabricated into n distinct finished products; several finished products may be included in a single production batch to exploit economies of scale. In particular, assume that a fixed cost K_1 is incurred for any (second stage) production run. Likewise, a fixed cost K_0 is incurred whenever a new production run for the intermediate product is initiated.

We assume that customer demands for the end-items occur at constant, deterministic but item-specific rates. These demands must be filled from available inventories, i.e., backlogging is not allowed. While different items may be combined in a single production batch, the total production volume per batch cannot exceed a given capacity limit. As the above examples demonstrate, this capacity limit may be expressed, for example, as a restriction on the total volume, weight, (variable) production time or the number of packaging units (bottles, boxes) that can be assigned to a single production run.

To simplify the notation, and without loss of generality, we express all demand rates in the *same*, *common* unit as is used to express the capacity restriction (gallons, pounds, hours, bottles, boxes, etc.). Let b denote the capacity limit ($b \leqslant \infty$). (A similar capacity limit b_0 may be imposed on production runs for the intermediate product. Our analysis and conclusions are easily extended for this case as long as the capacity ratio b_0/b is a power-of-two value.) Simultaneous production of several batches is permitted; on the other hand, if each production run is to be performed on one of a limited number (say L) of machines or equipment pools, no more than L batches may be produced at any given point in time.

Let d_i denote the demand rate for item i ($i = 1, \ldots, n$) and assume that these rates are integer valued. Inventory carrying costs are incurred at a constant rate per unit of time, per unit stored. The cost rates for the intermediate product and the n end-items may all be distinct.

We are interested in determining a production/inventory strategy minimizing long-run average costs. We assume that the variable production costs (in both stages) are linear in the production volumes; hence, these cost components may be ignored because their long-run average value is identical for all relevant replenishment strategies, with long-run average production rates equal to the demand rates.

Optimal policies may be very complex (even without joint setup costs or capacity limits per setup, see, e.g., Roundy 1985) and their complexity makes them difficult to implement even if they could be computed efficiently. As a consequence, we restrict ourselves to the following class of *family-based* replenishment strategies Φ. A replenishment strategy in this class specifies a collection of *families* (subsets of items) covering all end items; if an item belongs to several families, a specific fraction of its sales is assigned to each family. Each time the inventory of one item in a family is replenished, the inventories of all other items in the family are replenished as well.

Our restriction is similar to that applied in many other joint replenishment problems, see e.g., Chakravarty, Orlin and Rothblum (1982, 1985) and Barnes, Hoffman and Rothblum (1989), discussed below. Note that a large amount of flexibility is preserved within the class Φ by allowing items to be assigned to several families, i.e., by allowing families to overlap.

The use of a fixed collection of families, all of whose items are replenished simultaneously, allows one to assign these families to dedicated machine cells or work centers, as well as to specific (groups of) administrative, materials handling and sales force personnel. For these reasons, one would often impose an upper bound M^* on the total (sales) volume which may be assigned to a single family. When the number of dedicated work centers, or groups of administrative, materials handling and sales force personnel is fixed, one may wish to specify the number of families, L, or to impose a lower or upper bound on L. In other settings, L may be treated as an unconstrained variable. It is also relatively simple to extend our analysis to include bounds on the *frequency* with which families may be produced; see Anily (1987).

We derive a simple algorithm which results in a replenishment strategy whose long-run average cost (UB) is guaranteed to come within a few percentage points of a lower bound (LB) for the minimum achievable cost under any strategy in Φ. For example, when L is variable, and the total sales volume $N = \sum_{i=1}^{n} d_i$ is an integer multiple of M^*, the optimality gap $(= (UB - LB)/LB)$ is less than 6.1%. If M^* fails to divide N the gap is bounded by $0.06 + 2M^*/N$, which for fixed M^* quickly decreases to 0.06 as the number of items increases. Also when $b = \infty$, i.e., in the absence of production capacity constraints, the gap can be reduced to 2% regardless of any other considerations. The complexity of the algorithm is bounded by $0(n \log n + d_{\max} n)$ where $d_{\max} = \max_{i=1,\ldots,n} d_i$ is the demand rate of the largest item.

The proposed strategy specifies a collection of families, each of which is replenished at constant intervals. Production runs of the intermediate product are conducted at constant intervals as well. All replenishment intervals are powers-of-two times the smallest such interval; this strategy is thus easy to implement. Such policies are referred to as *power-of-two policies*. Note that the replenishment epochs of any given product are equidistant as well, even though a product may belong to more than one family. This follows immediately from the power-of-two property of the replenishment intervals. The inventories of any given family of items are, at each production run, replenished by constant amounts, but consecutive production volumes of the intermediate product may vary, according to a simple periodic pattern.

Note that under a power-of-two policy, production runs for a family with the lowest replenishment frequency, coincide with production runs for all other families. (Assume, for example, that three families are employed. A power-of-two policy may prescribe, for example, that the first family be replenished on a daily basis, the second family every other day and the third one, once per eight days. Assuming that at time zero we start with an empty system, three *simultaneous* production runs are required at times $t = 8, 16, 24, \ldots$ etc., one for each family.) Thus, if each production run is to be performed on one of a limited number of machines or equipment pools, it is required that L, the number of families, be bounded by the number of machines.

The assumption that all demand rates are constant and deterministically known, although common to all of the above mentioned planning models, represents a serious restriction. This assumption may be valid when applying the model to the production of components by assembly plants (e.g., in the automobile, chemical and pharmaceutical industries; these plants often operate under deterministic regular schedules for their end items). In many other settings, however, sales volumes are subject to a considerable degree of uncertainty or nonstationarity. We believe that our model continues to be useful in such settings when applied in a *hierarchical* planning mode, for the purpose of determining *optimal* replenishment frequencies for the intermediate product and each end-item. With these parameters fixed, inventory rules could be determined on the basis of some of the available two-stage stochastic inventory models, see e.g., Eppen and Schrage (1981) and Federgruen and Zipkin (1984a, b, c and 1988). This would, for example, allow for an adequate determination of safety stocks.

We complete this Introduction with a review of related inventory planning models and an outline of the remainder of the paper. In our review, we restrict ourselves to models addressing single-stage or two-stage production or distribution systems. We refer to Maxwell and Muckstadt (1985), Roundy (1986), Zheng (1987) and Muckstadt and Roundy (1988) for a discussion of more general network topologies.

One of the most extensively studied inventory replenishment problems with *joint setup costs* is the so-called Joint Replenishment Problem; see Brown (1967), Goyal (1973, 1974a, b), Goyal and Belton (1979), Graves (1979), Nocturne (1972), Schweitzer and Silver (1983), Shu (1971), Silver (1976), and Jackson, Maxwell and Muckstadt (1985). The Joint Replenishment Problem may be viewed as a special case of the previously described model, in which no inventories of the intermediate product may be kept (i.e., the distinction between the two production stages vanishes) and in which no restrictions need to be addressed regarding the number of items or the total volume produced in a single batch (i.e., with $M^* = \infty$ and $b = \infty$). On the other hand, a somewhat more general joint setup cost structure may be handled: as in the above described model, there is a fixed setup cost for each production batch, independent of which items are produced; in addition, however, item specific setup costs may be incurred for each item included in the batch. Jackson, Maxwell and Muckstadt (1985) derive an $0(n \log n)$ algorithm that generates a power-of-two policy whose long-run average cost comes within 6% of the Joint Replenishment Problem's minimum average cost per unit time. (Even for this relatively simple model no algorithm is known to result in an exactly optimal solution.)

Chakravarty, Orlin and Rothblum (1985) building on earlier work (Chakravarty, Orlin and Rothblum 1982) restrict themselves to strategies that employ a fixed partition of items into families: Each time the inventory of a given item is replenished, it is replenished jointly with the other members of the family and the setup cost of that family is incurred. (This class of strategies bears similarity to the class Φ considered here.) Federgruen and Zheng address general joint setup cost structures in which the setup cost is given by a general (set) function of the collection of items to be jointly replenished. (This setup cost function is required merely to reflect economies of scale, in the sense of submodularity, see *ibid.*) As in all previous references, it is assumed that an unlimited production volume may be generated at the expense of a single setup.

For a given collection of families, and in the absence of production capacity constraints, i.e., when $b = \infty$, the remaining problem reduces to identifying an optimal inventory replenishment strategy in the one-warehouse, multiple retailer lot sizing model analyzed by Roundy (1985), where the intermediate product plays the role of the warehouse and each family plays the role of a single retailer. (As before, Roundy's model allows for the fixed procurement cost to be retailer dependent.) Roundy (1985) has identified a simple $0(n \log n)$ procedure for this "one-warehouse-multiple-retailer" model which results in a power-of-two-policy whose cost is guaranteed to come within 2% of the minimal cost, as achievable under *any* strategy! (Here, n represents the number of retailers.) Queyranne (1987) has shown that this algorithm can be implemented in $0(n)$ time. Muckstadt and Roundy (1987) consider a related one-warehouse, multiple retailer, multiple item model with a fixed order cost at each retailer which is independent of the specific items ordered. In this paper, the class of policies is restricted to *nested* power-of-two policies. The nestedness condition means that every time a shipment of an item is received at the warehouse, a shipment of the item is made to each retailer as well. (Roundy 1985 explains that this restriction may lead to a serious loss in optimality.) With N representing the number of retailer/item combinations, the proposed algorithm requires $0(N \log N)$ operations and results in a nested power-of-two policy which comes within 6% of an optimal such policy. If nonnested policies are to be considered as well, the $0(N^4)$ general algorithm in Roundy (1986) may be invoked.

As just indicated, ours appears to be the first continuous-time, multistage inventory replenishment model with an explicit capacity constraint on the total production volume per batch. (Aggregate capacity constraints on the total number of production runs per unit of time have been handled within the context of the above discussed Joint Replenishment Problem and the one-warehouse, multiple retailer model; see Jackson, Maxwell and Muckstadt (1988) and Muckstadt (1985).) Finite horizon, discrete-time planning models with similar production capacity constraints have been treated by Florian and Klein (1971), Florian, Lenstra and Rinnooy Kan (1980), Baker et al. (1978), Bitran and Yanasse (1982) and Barany, Van Roy and Wolsey (1984); see Federgruen and Zipkin (1986a, b) and Wijngaard (1972) for a treatment of stochastic, single-item, infinite horizon models with similar capacity constraints.

In Section 1, we derive some preliminaries and summarize the analysis. In Section 2, we show how a lower bound for the minimum system-wide costs (within the class Φ) may be computed efficiently. In Section 3, we describe a simple procedure that results in a feasible replenishment strategy of the above described structure; we show that its long-run average cost comes within a few percentage points of the lower bound, obtained in Section 2. These results are obtained with the help of efficient algorithms for a general class of structured partitioning problems addressed in Anily and Federgruen (1990).

1. PRELIMINARIES AND SUMMARY OF ANALYSIS

Assume that inventories of the intermediate product incur carrying costs at a rate h_0 per unit and per unit of time while inventories of end-item i are charged at a rate h'_i, $i = 1, \ldots, n$. Let $h_i = h'_i - h_0$ $(i = 1, \ldots, n)$ denote the *echelon holding cost* rate and assume that $h_i \geqslant 0$ $(i = 1, \ldots, n)$. Since holding cost rates usually increase with the (cumulative) value added, this assumption is almost always satisfied.

Recall that the demand rates of all items are assumed to be integer-valued, or more generally, that they are integer multiples of some common quantity d, i.e., $d_i = k_i d$, $i = 1, \ldots, n$ with k_i an integer between 1 and K for some $K \geqslant 1$; for notational simplicity, and without loss of generality, we assume that $d = 1$. We define a *demand-item* as an item with a demand rate of $d = 1$. Each end-item i $(i = 1, \ldots, n)$ can thus be viewed as consisting of d_i *independent demand-items*, each with unit demand rate and all with the same holding cost rate h_i.

We restrict ourselves to Φ, the class of strategies which employ a fixed partition of the set of demand-items: Each time the inventory of one item in a family is replenished, the inventories of all other items in the family are replenished as well. Let $X = \{x_1, \ldots, x_N\}$ be the collection of all demand-items (with $N = \sum_{i=1}^{n} d_i$) numbered in ascending order of their incremental holding cost rates, i.e., $h_1 \leqslant h_2 \leqslant \ldots \leqslant h_N$.

We use $\chi = \{X_1, \ldots, X_L\}$ to denote a partition of X. For any given partition χ of X, let $U(\chi)$ denote the minimal long-run average cost under the best replenishment strategy employing the collection of families χ.

For a given feasible partition $\chi = \{X_1, \ldots, X_L\}$, a power-of-two policy is characterized by a vector of replenishment intervals $T = (T_0, T_1, \ldots, T_L)$ where

$T_0 =$ the (constant) replenishment interval of the intermediate product;

$T_l =$ the (constant) replenishment interval of family l, l = 1, . . . , L.

Under a power-of-two policy, all T_l's are powers-of-two times the smallest such interval. Let $m_l = |X_l|$ (l = 1, . . . , L). One easily verifies (Roundy 1985) that the long-run average cost under this policy is given by

$$C_\chi(T) = \frac{K_0}{T} + \sum_{l=1}^{L} \left\{ \frac{K_1}{T_l} + \frac{1}{2} H_l T_l + \frac{1}{2} m_l h_0 \max(T_0; T_l) \right\} \quad (1)$$

where $H_l \overset{\text{def}}{=} \sum_{i \in X_l} h_i$, $l = 1, \ldots, L$. Note that at each production run for family l, $m_l T_l$ units need to be produced. A power-of-two policy T is thus feasible if and only if $T_l \leqslant b/m_l$ $(l = 1, \ldots, L)$. An optimal power-of-two policy for partition χ is therefore obtained from the minimization problem

$$U^H(\chi) = \inf\{C_\chi(T): 0 \leqslant T_0;\ 0 \leqslant T_l \leqslant b/m_l;\ T_l/T_0 = 2^{n_l},\ n_l \text{ integer } (l = 1, \ldots, L)\}.$$

Let $\underline{U}(\chi)$ represent the value of the *continuous* relaxation of this minimization problem in which the power-of-two ratio requirements are relaxed

$$\underline{U}(\chi) \overset{\text{def}}{=} \inf\{C_\chi(T): 0 \leqslant T_0;\ 0 \leqslant T_l \leqslant b/m_l\ (l = 1, \ldots, L)\}. \quad (2)$$

The following lemma shows that $\underline{U}(\chi)$ provides a lower bound for $U(\chi)$ even though $U(\chi)$ represents the minimal cost under *any* replenishment strategy which employs the collection of families χ, even though no power-of-two policy needs to be optimal, and even though $C_\chi(T)$ may fail to represent the average cost for vectors T which are not power-of-two. The proof of this lemma is given in the Appendix and extends that of Theorem 1 in Roundy (1985) for uncapacitated models.

Lemma 1. *$\underline{U}(\chi) \leqslant U(\chi)$ for every partition χ.*

Thus, a lower bound for V^*, the minimal long-run average cost among *all* strategies in Φ, is obtained as

$$V^* \geqslant \underline{V} \overset{\text{def}}{=} \min\{\underline{U}(\chi): \chi \text{ is a feasible partition}\}.$$

To express $\underline{V}$ differently, let the function $c_T(\cdot,\cdot)$ for any $T > 0$, be defined by

$$c_T(H, m) = \inf_{0 \leqslant \vartheta \leqslant b/m} \cdot \left\{ K_1 \vartheta^{-1} + \frac{1}{2} H \vartheta + \frac{1}{2} h_0 m \max(T; \vartheta) \right\} \quad (3)$$

and note from (1)–(3) and by an interchange of infimum operators that

$$\underline{V} = \inf_{T_0 > 0} \left\{ K_0 T_0^{-1} + \min \left[\sum_{l=1}^{L} c_{T_0}(H_l, m_l): \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ and } m_l \leqslant M^*,\ l = 1, \ldots, L \right] \right\}. \quad (4)$$

Below we show with the help of the results in Anily and Federgruen (1990) that the following partition χ^* achieves the minimum in (4) for all values of $T_0 > 0$!

If L is variable: $X_1^* = \{x_1, \ldots, x_r\}$,

$X_l = \{x_{r+(l-2)M^*+1}, \ldots, x_{r+(l-1)M^*}\}$,

$l = 2, \ldots, \lceil N/M^* \rceil$

where $r = M^*$ if N is a multiple of M^* and

$$r = N \text{ (modulo } M^*) = N - \lfloor N/M^* \rfloor M^* \text{ otherwise}. \quad (5a)$$

If L is fixed: (Without loss of generality, $\lceil N/M^* \rceil \leqslant L \leqslant N$. Indeed, if $L < \lfloor N/M^* \rfloor$, then $LM^* \leqslant (\lceil N/M^* \rceil - 1)M^* < N$, i.e., no feasible partition exists.) There exists an index

$$l^* \overset{\text{def}}{=} \left\lfloor \frac{(LM^* - N)}{(M^* - 1)} \right\rfloor \text{ such that } X_l^* = \{x_l\} \quad (5b)$$

for $l = 1, \ldots, l^*$ and $\{x_{l^*+1}, \ldots, x_N\}$ is partitioned as in the case where L is variable.

Note that when L is variable, χ^* employs the lowest possible number of families $L = \lceil N/M^* \rceil$. For L fixed, the l^* demand items with the lowest (incremental) holding cost rates each form a family by itself (i.e., are replenished by themselves) while the remaining demand items are partitioned as in the case where L is variable. Let $\hat{l}$ be the number of families which are not filled to capacity (i.e., $m_l < M^*$). Note that $\hat{l} \leqslant 1$ if L is variable and $l^* \leqslant \hat{l} \leqslant l^* + 1$ if L is fixed.

The partition χ^* is obtained by an exceedingly simple, linear time algorithm. We also identify a simple linear time algorithm that results in a vector of replenishment intervals T^* such that the system-wide lower bound $\underline{V} = C_{\chi^*}(T^*)$. Since the vector T^* may fail to be a power-of-two, the corresponding replenishment strategy may fail to be feasible. However, we exhibit a simple rounding procedure which transforms the vector T^* into a power-of-two vector T^H; the average cost of this policy, implemented with the collection of families χ^*, is shown to come within a few percentage points of the lower bound $\underline{V}$. (See the Introduction and the discussion below for a more precise statement of the optimality gap.)

2. EVALUATION OF THE LOWER BOUND $\underline{V}$

We first show that the functions $c_T(\cdot, \cdot)$ may be evaluated in closed form. Let

$$\tau'(H, m) = \left[2K_1/(H + mh_0)\right]^{1/2}$$

and (6)

$$\tau(H, m) = \left[2K_1/H\right]^{1/2}$$

be the order intervals obtained by the EOQ formula with a fixed cost of K_1, a demand rate of one and holding cost rates of $(H + mh_0)$ and H, respectively.

Lemma 2

a. *If $T \leqslant b/m$, then*

$$c_T(H, m)$$

$$\begin{aligned} & mK_1/b + {}^1\!/_2(H + mh_0)b/m, \; b/m < \tau' && \text{(7a)}\\ = {} & \left[2K_1(H + mh_0)\right]^{1/2}, \quad T < \tau' \leqslant b/m && \text{(7b)}\\ & K_1/T + {}^1\!/_2(H + mh_0)T, \quad \tau' \leqslant T \leqslant \tau && \text{(7c)}\\ & \left[2K_1 H\right]^{1/2} + {}^1\!/_2 mh_0 T, \quad \tau < T. && \text{(7d)} \end{aligned}$$

b. *If $b/m \leqslant T$, then*

$$c_T(H, m)$$

$$\begin{aligned} & mK_1/b + {}^1\!/_2 Hb/m + {}^1\!/_2 mh_0 T, = b/m < \tau && \text{(8a)}\\ & \left[2K_1 H\right]^{1/2} + {}^1\!/_2 mh_0 T, \quad \tau \leqslant b/m. && \text{(8b)} \end{aligned}$$

Proof

a. Let

$$D_T(\vartheta, H, m) = \begin{cases} K_1\vartheta^{-1} + {}^1\!/_2 H\vartheta + {}^1\!/_2 mh_0\vartheta, & \vartheta \geqslant T\\ K_1\vartheta^{-1} + {}^1\!/_2 H\vartheta + {}^1\!/_2 mh_0 T, & \vartheta < T. \end{cases}$$

If $T \leqslant b/m$ observe that in cases (7b), (7c) and (7d) $\inf_{\vartheta > 0} D_T(\vartheta, H, m)$ is achieved at a point that is smaller than or equal to b/m (τ', T and τ, respectively). Thus, $\inf_{0 < \vartheta \leqslant b/m} D_T(\vartheta, H, m)$ is achieved in the same values as well. In case (6a), where $T \leqslant b/m < \tau'$, $\inf\{D_T(\vartheta, H, m)\colon T \leqslant \vartheta \leqslant b/m\}$ is achieved at $\vartheta = b/m$, while for $\vartheta \leqslant T < \tau' < \tau$ the function $D_T(\vartheta, H, m) = K_1\vartheta^{-1} + {}^1\!/_2 H\vartheta + {}^1\!/_2 mh_0 T$ is nonincreasing. Thus, $\vartheta = b/m$ represents the best possible value, in this case.

b. If $T \geqslant b/m$, only values of ϑ with $\vartheta \leqslant T$ are feasible. Since τ is the unconstrained minimum (over ϑ) of the function $\{K_1\vartheta^{-1} + {}^1\!/_2 H\vartheta + {}^1\!/_2 mh_0 T\}$, it follows that the function $D_T(\vartheta, H, m)$ achieves its minimum over $(0, b/m]$ in $\min(b/m; \tau)$.

Note that the inner minimization in (4)

$$\min\left\{\sum_{l=1}^{L} c_{T_0}(H_l, m_l)\colon \chi = \{X_1, \ldots, X_L\} \text{ partitions } X \text{ and } m_l \leqslant M^*, \; l = 1, \ldots, L\right\} \quad (9)$$

represents a partitioning problem in which the N demand items are to be assigned to L families to minimize the sum of all family costs where the cost of a single family merely depends on the total value of the (incremental) holding cost rates as well as the number of demand items in the family, in accordance with the $c_T(\cdot, \cdot)$ function. These partitioning problems are NP-complete for general group cost functions $\gamma(H, m)$; see Chakravarty, Orlin and Rothblum (1982) and Anily and Federgruen (1990). However, the properties of the group cost function $c_T(\cdot, \cdot)$, identified in Lemma 3, guarantee that an optimal partition χ^* is easily identified and of an extremely simple structure.

A function $g(x, y)$ of two real-valued variables is said to have antitone differences, if

$$g(x_2, y_1) - g(x_1, y_1) \geqslant g(x_2, y_2) - g(x_1, y_2) \quad \text{for all } x_1 < x_2, \; y_1 < y_2. \quad (10)$$

(See Anily and Federgruen for an extensive discussion of this property.)

Lemma 3

a. *For any $T > 0$, the function $f_T(\tilde{h}, m) \stackrel{\text{def}}{=} c_T(m\tilde{h}, m)$ is concave in both arguments.*

b. *The function $c_T(\cdot,\cdot)$ has antitone differences for any $T > 0$.*

Proof

a. Fix $T > 0$. Treating m as a continuous variable, it is easily verified from (7) *and* (8) *that $\partial^2 f_T / \partial \tilde{h}^2$ and $\partial^2 f_T / \partial m^2$ exist and are nonpositive almost everywhere. It thus suffices to show that $(\partial f_T / \partial \tilde{h})(\partial f_T / \partial m)$ exists and is continuous in $\tilde{h}(m)$ everywhere. As in Lemma* 2, *we distinguish between the following two cases.*

Case I. $T \leqslant b/m$: It suffices to verify the existence and continuity of $(\partial f_T / \partial \tilde{h})(\partial f_T / \partial m)$ in points $(\tilde{h}, m)$ where: (i) $b/m = [2K_1/m(\tilde{h} + h_0)]^{0.5}$, (ii) $T = [2K_1/m(\tilde{h} + h_0)]^{0.5}$, and, (iii) $T = [2K_1/m\tilde{h}]^{0.5}$. For (i), $\partial^+ f_T / \partial \tilde{h} = [{}^1\!/2 K_1 m/(\tilde{h} + h_0)]^{0.5} = b/2 = \partial^- f_T / \partial \tilde{h}$; for (ii) $\partial^- f_T / \partial \tilde{h} = [{}^1\!/2 K_1 m/(\tilde{h} + h_0)]^{0.5} = {}^1\!/2 mT = \partial^+ f_T / \partial \tilde{h}$, and for (iii) $\partial^+ f_T / \partial \tilde{h} = [{}^1\!/2 m K_1 / \tilde{h}]^{0.5} = {}^1\!/2 mT = \partial^- f_T / \partial \tilde{h}$. Similarly, for (i) $\partial^- f_T / \partial m = [K_1(\tilde{h} + h_0)/(2m)]^{0.5} = K_1/b = \partial^+ f_T / \partial m$; for (ii) $\partial^+ f_T / \partial m = [K_1(\tilde{h} + h_0)/(2m)]^{0.5} = T(\tilde{h} + h_0) = \partial^+ f_T / \partial m$, and for (iii) $\partial^- f_T / \partial m = [K_1 \tilde{h}/(2m)]^{0.5} + h_0 T = T(\tilde{h} + h_0) = \partial^+ f_T / \partial m$.

Case II. $T \geqslant b/m$: It suffices to verify the existence and continuity of $(\partial f_T / \partial \tilde{h})(\partial f_T / \partial m)$ in the points $(\tilde{h}, m)$, where $b/m = [2K_1/m\tilde{h}]^{0.5}$. But $\partial^+ f_T / \partial \tilde{h} = [{}^1\!/2 K_1 m / \tilde{h}]^{0.5} = b/2 = \partial^- f_T / \partial \tilde{h}$ and $\partial^- f_T / \partial m = [K_1 \tilde{h}/(2m)]^{0.5} + h_0 T = K_1/b + h_0 T = \partial^+ f_T / \partial m$.

b. Fix $T > 0$. As in Part a, we treat m as a real-valued variable. We first observe that $\partial c_T / \partial H$ exists everywhere and is continuous in m. The former follows from $\partial f_T / \partial \tilde{h}$ existing everywhere, invoking the chain rule. The latter is easily verified investigating all of the cases in Part a. Now fix $H_1 < H_2$. Since $\partial c_T / \partial H$ exists everywhere, we have $c_T(H_2, m) - c_T(H_1, m) = \int_{H_1}^{H_2} \partial c_T(H, m)/\partial H \, dH$. The fact that the above expression is nonincreasing in m, i.e., the fact that $c_T(\cdot,\cdot)$ has antitone differences (see (10)) now follows from the function $\partial c_T / \partial H$ being continuous in m (see above) and its derivative $(\partial^2 c_T / \partial H \partial m)$ being nonpositive almost everywhere.

To address the partitioning problem (9), we first need to distinguish between the following types of partitions. Recall that the demand items in X are numbered in ascending order of their holding cost rates. In any given partition $\chi = (X_1, \ldots, X_L)$ we number the families in ascending order of their cardinalities, i.e., $m_1 \leqslant m_2 \leqslant \ldots \leqslant m_L$. In a given ordered partition, let $l(i)$ denote the index of the family to which demand item i is assigned. We refer to the index function $l(\cdot)$ as the *family index function*. An ordered partition is *monotone* if the group index function is nondecreasing. (For example, $\chi = (X_1, X_2) = (\{1, 2\}; \{3, 4, 5\})$ is a monotone partition of $X = \{x_1, \ldots, x_5\}$ but $\chi' = (\{4, 5\}; \{1, 2, 3\})$ is not because $l(5) < l(1)$. In a monotone partition, a high holding cost item is not assigned to a set of lower cardinality than a low holding cost item.) A partitioning problem is called *extremal* if a monotone optimal partition exists and the cost of any monotone partition $\chi = \{X_1, \ldots, X_L\}$ does not increase by shifting the highest indexed demand item in any of its families to the next family (i.e., by transferring the highest indexed demand item of some family X_l to X_{l+1}, $1 \leqslant l \leqslant L$).

Theorem 5 in Anily and Federgruen establishes that the partitioning problems in (9) are extremal indeed because the $f_T(\cdot,\cdot)$ function is concave in both arguments and the $C_T(\cdot,\cdot)$ function has antitone differences; see Lemma 3. (Concavity of $f_T(\tilde{h}, m)$ in $\tilde{h}$ and $c_T(\cdot,\cdot)$ having antitone differences guarantees that a monotone optimal partition exists. The additional concavity property of $f_T(\cdot,\cdot)$ with respect to m establishes the extremality of the partitioning problem.) The importance of this characterization follows from the fact that the partition $\chi^* = (X_1^*, \ldots, X_{L^*}^*)$ defined by (5) is optimal whenever a partitioning problem is extremal (see Theorem 1 in Anily and Federgruen).

We conclude with the next lemma.

Lemma 4. *The partition χ^* defined by* (10) *achieves the minimum in* (9) *and hence in* (4) *for all $T_0 > 0$!*

Let

$$H_l^* = \sum_{i \in X_l^*} h_i; \quad m_l^* = |X_l^*|; \quad \tau_l' = \tau'(H_l^*, m_l^*);$$

$$\tau_l = \tau(H_l^*, m_l^*) \quad (l = 1, \ldots, L^*).$$

In view of Lemma 4, (4) simplifies to

$$\underline{V} = \inf_{T_0 > 0} \left\{ K_0 / T_0 + \sum_{l=1}^{L^*} c_{T_0}(H_l^*, m_l^*) \right\}. \tag{11}$$

It remains to be shown that there exists a unique value T_0^* achieving the minimum in (11) and that this value can be computed easily. The optimal corresponding replenishment intervals $T_1^*, \ldots, T_L^*$ follow easily; see the definition of the function $c_{T_0}(\cdot,\cdot)$ and Lemma 2. Both results follow from a simple generalization of the corresponding results in Roundy (1985) and Queyranne (1987) for systems without production capacity constraints.

One easily verifies from (11) and Lemma 2 that $\underline{V}$ may be written in the form

$$\underline{V} = \inf_{T_0 > 0} \left\{ K_0/T_0 + \sum_{l=1}^{L^*} \alpha_l(T_0)/T_0 + \beta_l(T_0) + \gamma_l(T_0)T_0 \right\}$$

with $\alpha_l(\cdot), \beta_l(\cdot), \gamma_l(\cdot)$ piecewise constant functions ($l = 1, \ldots, L^*$) and that the functions $[\alpha_l(T)/T + \beta_l(T) + \gamma_l(T)T]$ ($l = 1, \ldots, L^*$) are convex and differentiable everywhere with the possible exception of the point $T = b/m_l^*$.

Thus, $\underline{V}$ is the infimum (over $T > 0$) of a *strictly* convex function $U(T)$ which is differentiable everywhere with the possible exception of the (at most L^*) points in $\{b/m_l^*: l = 1, \ldots, L^*\}$ and with $\lim_{T\downarrow 0} U(T) = \lim_{T\to\infty} U(T) = \infty$. Moreover, $U(\cdot)$ is of the form $[\alpha(T)/T + \beta(T) + \gamma(T)T]$ with $\alpha(T), \beta(T), \gamma(T)$ piecewise constant, changing values only when T crosses one of the at most $(2L^* + 3)$ values in $Z \stackrel{\text{def}}{=} \{\tau_l', l = 1, \ldots, L^*\} \cup \{\tau_l: l = 1, \ldots, L^*\} \cup \{b/m_l^*: l = 1, \ldots, L^*\}$. We conclude that the infimum in (11) is achieved for a unique value T_0^* which may be obtained by the following procedure.

T_0^*-Finder

Step 1. Rank the (at most $(2L^* + 3)$) values in Z in increasing order.

Step 2. Proceeding in this order, compute for each point in Z first the left and then the right derivative of $U(\cdot)$ until for some $z^+ \in Z$ a nonnegative (derivative) value is found. Let $\bar{z}$ be the preceding value in Z for which the left or right derivative was evaluated. (The very first left derivative can easily be shown to be negative.) If $\bar{z} = z^+$, then $T_0^* = \bar{z} = z^+$; otherwise $T_0^* = (\alpha/\gamma)^{1/2} \in [\bar{z}, z^+]$, where α and γ represent the constant values of $\alpha(\cdot)$ and $\gamma(\cdot)$ on the interval $[\bar{z}, z^+]$.

Note that Step 1 requires $0(L^* \log L^*)$ operations; Step 2 involves at most $(2L^* + 3)$ elementary evaluations. Queyranne has pointed out that the values in Z do not need to be ranked up front. He shows that T_0^* may be found in $0(L^*)$ operations only, by employing a linear-time median finding algorithm.

3. OBTAINING A FEASIBLE POWER-OF-TWO POLICY

As pointed out, the vector T^* of replenishment intervals may fail to be a power-of-two vector and may thus fail to be implementable. Below we exhibit a simple rounding procedure that transforms T^* into a power-of-two vector T^H, such that $C_{\chi^*}(T^H)$ exceeds $\underline{V}$ by a few percentage points only. First, partition the index set $\{1, \ldots, L^*\}$ into the following five sets:

$$G = \{l \mid 1 \leqslant l \leqslant L^*, T_0^* < \tau_l' \leqslant b/m_l^*\}$$

$$E = \{l \mid 1 \leqslant l \leqslant L, T_0^* \leqslant b/m_l^* \text{ and } \tau_l' \leqslant T_0^* \leqslant \tau_l\}$$

$$S = \{l \mid 1 \leqslant l \leqslant L^*, \tau_l < T_0^* \text{ and } \tau_l \leqslant b/m_l^*\}$$

$$I_1 = \{l \mid 1 \leqslant l \leqslant L^*, T_0^* \leqslant b/m_l^* \text{ and } b/m_l^* < \tau_l'\}$$

$$I_2 = \{l \mid 1 \leqslant l \leqslant L^*, T_0^* > b/m_l^* \text{ and } b/m_l^* < \tau_l\}.$$

Thus, if family $l \in G$, E or S then the optimal replenishment interval T_l^* is given by τ_l' (strictly bigger than T_0^* but no larger than b/m_l^*), T_0^* (strictly smaller than b/m_l^*), and τ_l (strictly smaller than T_0^* and no larger than b/m_l^*), respectively; see Lemma 2. Similarly, if $l \in I_1$ or I_2, then the optimal replenishment interval T_l^* is equal to its upper bound (b/m_l^*) and larger and strictly smaller than T_0^*, respectively. (Note that if $b = \infty$, $I_1 = I_2 = \varnothing$.)

The following rounding procedure generates a feasible power-of-two vector of replenishment intervals T^H which is order-preserving with respect to T^* in all but at most one component, i.e., the elements in the triples $(T_l^H, T_0^H, b/m_l^*)$ are ranked exactly like the elements in $(T_l^*, T_0^*, b/m_l^*)$ for all but at most one value of $l = 1, \ldots, L^*$. (See the Appendix for a more precise definition of order-preserving vectors.) These order-preserving characteristics help us to establish that the average cost of this policy, implemented with the collection of families χ^*, comes within a few percentage points of $\underline{V}$.

Rounding Procedure

Step 0. $\lambda := b/M^*$.

Step 1. For $l = 0, \ldots, L^*$, determine the unique integer t such that $T_l^* \in \lambda[2^{t-1/2}, 2^{t+1/2})$; $T_l^H := \lambda 2^t$. (Note $t = \lceil \log_2 T_l^* - \log_2(\sqrt{2}\lambda) \rceil$.)

Step 2. For $l = 1, \ldots, \hat{l}$ if $T_l^H > b/m_l^*$, then $T_l^H := 0.5 T_l^H$.

(Step 2 is required because for $l \leqslant \hat{l}$, (b/m_l^*) may fail to be a power-of-two times λ.)

Theorem 1

a. *The Rounding Procedure generates a feasible power-of-two policy T^H.*

b. $C_{\chi^*}(T^H)/V^* \leqslant C_{\chi^*}(T^H)/\underline{V} \leqslant 1.061 + 2\hat{l}M^*/N$.

Proof

a. T^H is clearly a power-of-two vector. For $l > \hat{l}$, $b/m_l^* = b/M^* = \lambda$; thus, since $T_l^* \leqslant b/m_l^*$, $T_l^H \leqslant b/m_l^*$, and hence, is feasible as well. For $l \leqslant \hat{l}$, let $T_l^{H'}$ be the value assigned to T_l^H in Step 1 and note that $T_l^{H'} \leqslant \sqrt{2}\, T_l^* \leqslant \sqrt{2}\, b/m_l^*$. Thus for $l \leqslant \hat{l}$, feasibility is preserved by Step 2.

b. Assume that $\hat{l} \geqslant 1$. In case $\hat{l} = 0$, the proof simplifies in an obvious way. Note that the components of T^H are order-preserving (with respect to T^*) with the possible exception of the first $\hat{l}$ components. Also, in view of Step 1, $1/\sqrt{2} \leqslant T_l^H/T_l^* \leqslant \sqrt{2}$ for all $l > \hat{l}$. Since the test in Step 2 is only satisfied if $T_l^H/T_l^* > 1$, we have $0.5 \leqslant T_l^H/T_l^* \leqslant \sqrt{2}$ for $l \leqslant \hat{l}$.

To prove the optimality gap, we first derive a (crude) bound for the *cost* term of the first $\hat{l}$ families in $C_{\chi^*}(T^*) = \underline{V}$.

It follows from Lemma 2 that $c_{T_0^*}(H, m)$ is nondecreasing in (H/m) and m. Since χ^* is monotone, we have $H_1^*/m_1^* \leqslant H_2^*/m_2^* \leqslant \ldots \leqslant H_{L^*}^*/m_{L^*}^*$ and, of course, $m_1^* \leqslant m_2^* \leqslant \ldots \leqslant m_L^*$. Thus, since $L^* \geqslant N/M^*$

$$\sum_{l=1}^{\hat{l}} c_{T_0^*}(H_l^*, m_l^*) \leqslant \frac{\hat{l}}{L^*} \sum_{l=1}^{L^*} c_{T_0^*}(H_l^*, m_l^*) = \frac{\hat{l}}{L^*}\, \underline{V} \leqslant \left(\frac{\hat{l}M^*}{N}\right) \underline{V}. \tag{12}$$

It follows from Lemma A1 that numbers μ_0^*, μ_l^* ($l \notin E$) may be computed such that for any *fully* order-preserving policy T

$$C_{\chi^*}(T) = \frac{\mu_0^*}{2}\left(\frac{T_0}{T_0^*} + \frac{T_0^*}{T_0}\right) + \sum_{l \in G \cup S} \frac{\mu_l^*}{2}\left(\frac{T_l}{T_l^*} + \frac{T_l^*}{T_l}\right) + \sum_{l \in I_1 \cup I_2} \mu_l^*.$$

Since T_l^H may fail to be order-preserving and invoking the bounds on the ratios T_l^H/T_l^* ($l = 1, \ldots, L^*$), we obtain

$$\begin{aligned}
C_{\chi^*}(T^H) &= \sum_{l=1}^{\hat{l}} \left[\frac{K_l}{T_l^H} + \frac{1}{2} H_l^* T_l^H + \frac{1}{2} m_l^* h_0 \max(T_l^H, T_0^H)\right] \\
&\quad + \mu_0^*/2\left[\frac{T_0}{T_0^*} + \frac{T_0^*}{T_0}\right] + \sum_{l \in G \cup S \setminus \{1, \ldots, \hat{l}\}} \mu_l^*/2\left(\frac{T_l^H}{T_l^*} + \frac{T_l^*}{T_l^H}\right) \\
&\quad + \sum_{l \in I_1 \cup I_2 \setminus \{1, \ldots, \hat{l}\}} \mu_l^* \\
&\leqslant \sum_{l=1}^{\hat{l}} \left[\frac{2K_l}{T_l^*} + \sqrt{2}\left(\frac{1}{2} H_l^* T_l^*\right) + \sqrt{2}\left(\frac{1}{2} m_l^* h_0\right) \max(T_l^*, T_0^*)\right] \\
&\quad + \left(\sum_{l \notin (E \cup \{1, \ldots, \hat{l}\})} \mu_l^*\right) \max\left\{\frac{1}{2}\left(\frac{1}{x} + x\right) : 2^{-1/2} \leqslant x \leqslant 2^{1/2}\right\} \\
&\leqslant 2 \sum_{l=1}^{\hat{l}} c_{T_0^*}(H_l^*, m_l^*) + 1.061\underline{V} \\
&\leqslant \underline{V}\left[1.06 + \frac{2\hat{l}M^*}{N}\right].
\end{aligned}$$

The following conclusions may be drawn with respect to the worst case optimality gaps. Let $\overline{V} = C_{\chi^*}(T^H)$.

Corollary 1

a. *If* $b = \infty$, $(\overline{V} - V^*)/V^* \leqslant 0.061$.

b. *If* $b < \infty$, *L is variable and N is a multiple of* M^*, $(\overline{V} - V^*)/V^* \leqslant 0.061$.

c. *If* $b < \infty$, *L is variable and N is not a multiple of* M^*, $(\overline{V} - V^*)/V^* \leqslant 0.061 + 2M^*/N$.

d. *If* $b < \infty$, *and L is fixed*, $(\overline{V} - V^*)/V^* \leqslant 0.061 + 2[\alpha(M^*/(M^* - 1)) + 1]M^*/N$ *where* $\alpha \overset{\text{def}}{=} L - N/M^*$.

Proof. Parts a, b and c follow from Theorem 1 and $\hat{l} = 0$, 0 and 1, respectively.

d. $$\hat{l} \leqslant \left\lfloor \frac{(LM^* - N)}{(M^* - 1)} \right\rfloor + 1 \leqslant \alpha\left(\frac{M^*}{M^* - 1}\right) + 1; \text{ see (5)}.$$

Note that the upper bounds for the worst case optimality gaps in Cases c and d, even though comparatively small, have been obtained by rather crude bounding arguments. For example, the proof of Theorem 1 is based on an increase of the cost of the first $\hat{l}$ families by a factor of two (!) when replacing their replenishment intervals T_l^* by T_l^H, and, in practice, a considerably smaller increase should be expected. When L is fixed, $\alpha = L - N/M^*$ denotes the excess number of families above the minimal (and indeed optimal; see (10)). Note that for fixed α, the upper bound for the worst case optimality gap rapidly decreases to 0.061.

We conclude with a description and discussion of the entire algorithm required to compute the lower bound $\underline{V}$, determine the collection of families χ^* and the power-of-two policy T^H.

Algorithm

Step 1. Rank the end-items $\{1, \ldots, n\}$ in ascending order of their (incremental) holding cost rates $\{h_i: i = 1, \ldots, n\}$. Use this list to generate the collection of demand items $X = \{x_1, \ldots, x_N\}$ (again numbered in ascending order of the holding cost rates).

Step 2. Determine χ^* in accordance with (10). Compute the numbers $\{H_l^*: l = 1, \ldots, L^*\}$ ($H_l^* = \sum_{i \in X_l^*} h_i$).

Step 3. Determine T_0^* by the linear-time implementation of the procedure T_0^*-finder; determine $\{T_l^*: l = 1, \ldots, L^*\}$ from (6) and (7).

Step 4. Determine T^H by the Rounding Procedure.

Complexity of the Algorithm

In Step 1, the ranking of the end-items requires $0(n \log n)$ operations and the creation of the list $X = \{x_1, \ldots, x_N\}$ an additional $0(N)$ operation. Step 2 requires $0(N)$ operations as well. As discussed, the T_0^*-finder procedure has a complexity $0(L^*) = 0(N)$, while $\{T_l^*: l = 1, \ldots, L^*\}$ may be computed in $0(L^*) = 0(N)$ elementary operations and evaluations of the square root function. Finally, Step 4 requires $0(L^*) = 0(N)$ operations and evaluations of the $\log_2(\cdot)$ function.

In summary, counting additions, multiplications, comparisons and evaluations of the square root and $\log_2(\cdot)$ function as elementary operations, we conclude that the complexity of the entire algorithm is $0(N \log N) = 0(n \log n + d_{\max} n)$.

Since a problem instance is specified by $0(n)$ input parameters (the end-items' demand rates and incremental holding cost rates, as well as a few cost and constant parameters), the algorithm is, strictly speaking, not fully polynomial in the usual *complexity theoretical* sense. We argue, however, that in practical applications $d_{\max}$ is relatively small, and for fixed values of $d_{\max}$ the algorithm is $0(n \log n)$ only!

We apply the algorithm to the following example.

Example. Consider a system with four end-items, i.e., $n = 4$. Their demand and echelon holding cost rates are given by

Product	1	2	3	4
Demand rate	5	7	4	10
Echelon holding	1	2	3	4

Assume that $M^* = 8$, L is variable, $b = 24$, $h_0 = 4$, $K_0 = 400$ and $K_1 = 100$.

Thus $X = \{x_1, \ldots, x_{26}\}$ where the demand-items $\{x_i, 1 \leqslant i \leqslant 5\}$ correspond to product 1, $\{x_i, 4 \leqslant i \leqslant 12\}$ correspond to product 2, $\{x_i, 13 \leqslant i \leqslant 16\}$ correspond to product 3 and $\{x_i, 17 \leqslant i \leqslant 26\}$ correspond to product 4. The optional partition is $\chi^* = \{X_1, X_2, X_3, X_4\}$, where $X_1^* = \{x_1, x_2\}$, $X_2^* = \{x_3, \ldots, x_{10}\}$, $X_3^* = \{x_{11}, \ldots, x_{18}\}$ and $X_4^* = \{x_{19}, \ldots, x_{26}\}$ resulting in $H_1^* = 2$, $H_2^* = 13$, $H_3^* = 24$, $H_4^* = 32$ and $m_1^* = 2$, $m_2^* = m_3^* = m_4^* = 8$.

As explained, the function $U(T)$, defined below (11), is differentiable everywhere except in the following points: $\tau_4' = 1.768$, $\tau_3' = 1.89$, $\tau_2' = 2.108$, $\tau_4 = 2.5$, $\tau_3 = 2.887$, $b/M^* = 3$, $\tau_2 = 3.922$, $\tau_1' = 4.472$, $\tau_1 = 10$ and $b/m_1 = 12$. (See (6) for the calculation of the τ' and τ values.) Following the T_0^*-finder procedure we conclude that $T_0^* = 3$. (At $T_0 = 3$ the left derivative of the function $U(\cdot)$ is negative and the right derivative is positive.) Moreover, according to the definition of the sets G, E, S, I_1 and I_2 we obtain that $G = \{1\}$, $E = \{2\}$ and $S = \{3, 4\}$. Thus $T_1^* = 4.472$, $T_2^* = 3$, $T_3^* = 2.887$, $T_4^* = 2.5$ and by substituting into $\underline{V}$ we get the lower bound $\underline{V} = 524.169$.

Following the Rounding Procedure we obtain: $\lambda = 3$, $T_0^H = 3$, $T_1^H = 6$, $T_2^H = 3$, $T_3^H = 3$, $T_4^H = 3$. It is easy to verify that $C_{\chi^*}(T^H) = 527.5$ which is less than 0.64% above the lower bound value!

APPENDIX

Proof of Lemma 1

In this appendix, we give an outline of the proof of Lemma 1 by showing how the proof of Theorem 1 in Roundy (1985) needs to be modified. Let $\chi = (X_1, \ldots, X_L)$ be a feasible ordered partition of X and $\hat{T}$ be the unique vector of replenishment intervals with $\underline{U}(\chi) = C_\chi(\hat{T}) = \inf\{C_\chi(T): 0 \leqslant T_0; 0 \leqslant T_l \leqslant b/m_l\}$. The existence of this vector follows as in Section 2. In Section 3, we define family index sets G, S, E, I, and I_2 with respect to a specific partition χ^* and vector T^*. Redefine these sets as well as the numbers $\{\tau_l, \tau_l', H_l: l = 1, \ldots, L\}$ with respect to the partition χ and the vector $\hat{T}$.

A vector T *preserves the order* (of $\hat{T}$) if $T_l \geqslant T_0$ and $T_l \leqslant b/|X_l|$, $l \in G$; $T_l = T_0$ and $T_0 \leqslant b/|X_l|$, $l \in E$; $T_l \leqslant b/|X_l|$, $T_l \leqslant T_0$, $l \in S$; $T_l = b/|X_l|$ and $T_0 \leqslant b/|X_l|$, $l \in I_1$ and $T_l = b/|X_l|$, $T_0 \geqslant b/|X_l|$, $l \in I_2$. Note that any order-preserving policy is feasible.

The long-run average cost associated with the partition χ and any order-preserving vector T may be written in the form $C_\chi(T) = K/T_0 + HT_0 + \sum_{l \notin E}(K_1/T_l +$

$\overline{H}_l T_l)$ where

$$K = K_0 + |E| K_1$$

$$H = {}^1/_2 \sum_{l \in E} (H_l + m_l h_0) + {}^1/_2 \sum_{l \in S \cup I_2} m_l h_0$$

$$\overline{H}_l = \begin{cases} {}^1/_2 (H_l + m_l h_0) & l \in G \cup I_1 \\ {}^1/_2 H_l & l \in S \cup I_2. \end{cases}$$

Let

$$\mu_0 = 2(KH)^{0.5}$$

$$\mu_l = \begin{cases} 2(K_1 H_l)^{0.5} & l \in G \cup S \\ |X_l| K_1 / b + \overline{H}_l b / |X_l| & l \in I_1 \cup I_2. \end{cases}$$

Lemma A1

a. $\underline{U}(\chi) = \mu_0 + \sum_{l \notin E} \mu_l$.

b. $\hat{T}_0 = \sqrt{K/H}$; $\hat{T}_l = \sqrt{K_1/H_l} \quad l \in G \cup S$.

Proof. Note from (1) that $C_\chi(T)$ is convex in $T \in R_+^{L+1}$. The function $\hat{C}_\chi$: $R^{|G|+|S|+1} \to R$:

$$(T_0, \{T_l\}_{l\in G}, \{T_l\}_{l \in S})$$
$$\to \frac{K_0}{T_0} + \sum_{l \in E} \left\{ \frac{K_1}{T_0} + \frac{1}{2}(H_l + m_l h_0) T_0 \right\}$$
$$+ \sum_{l \in I_1 \cup I_2} \left\{ \frac{m_l K_1}{b} + \frac{H_l b}{2 m_l} + \frac{1}{2} m_l h_0 \max\left(T_0; \frac{b}{m_l} \right) \right\}$$
$$+ \sum_{l \in G \cup S} \left\{ \frac{K_1}{T_l} + \frac{1}{2} H_l T_l + \frac{1}{2} m_l h_0 \max(T_0; T_l) \right\}$$

(obtained by substituting $T_l = T_0$, $l \in E$ and $T_l = b/m_l$, $l \in I_1 \cup I_2$ in $C_\chi(T)$) is thus strictly convex as well. On the polyhedron $\Pi \overset{\text{def}}{=} \{(T_0, \{T_l\}_{l\in G}, \{T_l\}_{l \in S})$: $T_l \geqslant T > 0$, $l \in G$ and $0 < T_l \leqslant T$, $l \in S\}$ we have $\hat{C}_\chi = \psi(T_0, \{T_l\}_{l \in G}, \{T_l\}_{l\in S}) \overset{\text{def}}{=} K_0/T_0 + HT_0 + \sum_{l \in I_1 \cup I_2} (m_l K_1 / b + \overline{H}_l b / m_l) + \sum_{l \in G \cup S} (K_1 / T_l + \overline{H}_l T_l)$. The vector $(\hat{T}_0, \{\hat{T}_l\}_{l \in G}, \{\hat{T}_l\}_{l \in S})$ is the unique global minimum of $\hat{C}_\chi$ because the capacity constraints are not binding for $l \in G \cup S$. This vector is also an interior point of Π and hence must be a local (unconstrained) minimum of ψ. But the latter has only one unconstrained minimum with $T_0 = \sqrt{K/H}$ and $T_l = \sqrt{K_1/H_l}$ $l \in G \cup S$. Thus, $\hat{T}_0 = \sqrt{K/H}$ and $\hat{T}_l = \sqrt{K_1/H_l}$, $l \in G \cup S$. Part a of the lemma now follows by substitution into $\hat{C}_\chi$.

The proof of Lemma 1 is based on a modification of the proof of Theorem 1 in Roundy (1985). As in Roundy, we show that it is possible to allocate the costs incurred by an arbitrary policy using a given partition $\chi = \{X_1, \ldots, X_L\}$ to the individual families of finished demand-items and the intermediate product in such a way that:

- the average cost incurred by the policy is at least as large as the sum of the average costs allocated to the families and the intermediate product;
- the total average cost allocated to a family or the intermediate product is the average cost of a solution to a single facility problem;
- the sum of the costs of the optimal solutions to these single facility problems is at least as large as $\underline{U}(\chi)$.

To facilitate this cost allocation process we first define *reallocated holding cost rates*. Referring to the intermediate product as the 0th family of demand-items, let for each $l = 1, \ldots, L$ $K_l = K_1$ and define the reallocated holding cost rate $\overline{H}_l = K_l / \hat{T}_0^2$ for $l \in W = E \cup \{0\}$. Also define $\hat{\mu}_l = 2(K_l \overline{H}_l)^{0.5}$, $l \in W$. For $l \notin W$, $\overline{H}_l$ has already been defined and we set $\hat{\mu}_l = \mu_l$. The following lemma is an adaptation of Lemma 1 of Roundy (1985).

Lemma A2

a. $H = \sum_{l \in W} \overline{H}_l$.

b. $\underline{U}(\chi) = \sum_{l=0}^{L} \hat{\mu}_l$ *where* $\hat{\mu}_l = \min_{0 < x \leqslant b/|X_l|} \{K_l / x + \overline{H}_l x\}$, $l = 1, \ldots, L$ *and* $\hat{\mu}_0 = \min_{x > 0} \{K_0 / x + \overline{H}_0 x\}$.

c. ${}^1/_2 H_l \leqslant \overline{H}_l \leqslant {}^1/_2 H_l + {}^1/_2 m_l h_0$, $l = 1, \ldots, L$.

d. $\overline{H}_0 = \sum_{l=1}^{L} H^l$ *where* $H^l = {}^1/_2 H_l + {}^1/_2 m_l h_0 - \overline{H}_l$, $l = 1, \ldots, L$.

Proof

a. Since $K_l / \overline{H}_l = \hat{T}^2 = K/H$ for all $l \in W$, we have $H = K / \hat{T}_0^2 = \sum_{l \in W} K_l / \hat{T}_0^2 = \sum_{l \in W} \overline{H}_l$.

b. It follows from the definition of the sets G and S that

$$\hat{T}_l = \sqrt{K_l / \overline{H}_l} \in (0, b/|X_l|]$$

achieves $\min_{0 \leqslant x} \{K_l / x + \overline{H}_l x\}$ so that $\hat{\mu}_l = 2(K_l \overline{H}_l)^{0.5} = K_l / \hat{T}_l + \overline{H}_l \hat{T}_l = \min_{0 \leqslant x \leqslant b/|X_l|} \{K_l / x + \overline{H}_l x\}$, $l \in G \cup S$.

Similarly, for $l \in E$, $\hat{T}_l = \hat{T}_0 = \sqrt{K_l / \overline{H}_l} \in (0, b/|X_l|)$ achieves $\min_{0 \leqslant x} \{K_l / x + \overline{H}_l x\}$ so that $\hat{\mu}_l = \min_{0 \leqslant x \leqslant b/|X_l|} \{K_l / x + \overline{H}_l x\}$, $l \in E$ as well. If $l \in I_1$, $\sqrt{K_l / \overline{H}_l} = \sqrt{K_1 / (H_l + \frac{1}{2} m_l h_0)} = \tau_l' > b/|X_l|$ so that $\hat{\mu}_l = K_l |X_l| / b + \overline{H}_l b / 2 |X_l| = \min_{0 \leqslant x \leqslant b/|X_l|} \{K_l / x + \overline{H}_l x\}$, $l \in I_1$.

A similar argument verifies Part b for $l \in I_2$. Obviously, $\hat{\mu}_0 = 2\sqrt{K_0 \overline{H}_0} = \min_{0 \leqslant x} \{K_l / x + \overline{H}_0 x\}$.

Part c is immediate from the definition of $\overline{H}_l$ and $\tau_l' \leqslant \hat{T}_0 \leqslant \tau_l$, $l \in E$. Finally, in view of Part a

$$\overline{H}_0 = H - \sum_{l \in E} \overline{H}_l = \sum_{l \in E} \left[{}^1/_2 H_l + {}^1/_2 m_l h_0 - \overline{H}_l \right] + \sum_{l \in S \cup I_2} {}^1/_2 m_l h_0 = \sum_{l=1}^{L} H^l.$$

Proof of Lemma 1

Choose an arbitrarily feasible infinite-horizon policy for the partition χ and let $c(t')$ be the average cost incurred by this policy over the finite interval $[0, t')$. We show that $\liminf_{t' \to \infty} c(t') \geqslant \underline{U}(\chi)$.

Let $J_l(t')$ be the number of replenishments of family l in $[0, t')$. Since the policy is feasible we must have

$$\limsup_{t \to \infty} t / J_l(t) \leqslant b / \mid X_l \mid. \qquad (*)$$

Following the proof of Theorem 1 in Roundy (1985) and employing the identities in Parts c and d of Lemma A2, one verifies that

$$c(t') \geqslant \sum_{l=0}^{L} \left[K_l \frac{J_l(t')}{t'} + \overline{H}_l \frac{t'}{J_l(t')} \right].$$

Thus, in view of (*), there exists a sequence $\{t_k\}_{k=1}^{\infty}$ with $\lim_{k \to \infty} c(t_k) = \liminf_{t \to \infty} c(t)$ and $x_l \stackrel{\text{def}}{=} \lim_{k \to \infty} t_k / J_l(t_k)$ exists for all $l = 1, \ldots, L$. Thus

$$\begin{aligned} \liminf_{t \to \infty} c(t) &\geqslant \sum_{l=0}^{L} \left[\frac{K_l}{x_l} + \overline{H}_l x_l \right] \geqslant \inf_{x > 0} \left[\frac{K_0}{x} + \overline{H}_0 x \right] \\ &\quad + \sum_{l=1}^{L} \inf_{0 \leqslant x \leqslant b / \mid X_l \mid} \left[\frac{K_l}{x} + \overline{H}_l x \right] \\ &= \mu_0 + \sum_{l=1}^{L} \mu_l = \underline{U}(\chi); \end{aligned}$$

see Lemma A2, Part b.

ACKNOWLEDGMENT

We are indebted to the Associate Editor and the referees for their thorough reading and careful editing of a previous version of this paper. The research of the first author has been supported in part by NSERC grant A4802. The research of the second author has been partially supported by NSF grant ECS-8604409 as well as a grant by the Faculty Research Fund of the Graduate School of Business, Columbia University.

REFERENCES

ANILY, S. 1987. Integrating Inventory Control and Transportation Planning. Ph.D. Dissertation, Columbia University, New York.

ANILY, S., AND A. FEDERGRUEN. 1990. Structured Partitioning Problems. *Opns. Res.* **39**, 132–149.

BAKER, K. R., P. S. DIXON, M. J. MAGAZINE AND E. A. SILVER. 1978. An Algorithm for the Dynamic Lot-Size Problem With Time-Varying Production Capacity Constraints. *Mgmt. Sci.* **24,** 1710–1720.

BARANY, I., T. J. VAN ROY AND L. A. WOLSEY. 1984. Strong Formulations for Multi-Item Capacitated Lot Sizing. *Mgmt. Sci.* **30,** 1255–1261.

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

BITRAN, G. R., AND H. H. YANASSE. 1982. Computational Complexity of the Capacitated Lot Size Problem. *Mgmt. Sci.* **28,** 1174–1186.

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

CHAKRAVARTY, A. K., J. B. ORLIN AND U. G. ROTHBLUM. 1982. A Partitioning Problem With Additive Objective With an Application to Optimal Inventory Groupings for Joint Replenishment. *Opns. Res.* **30,** 1018–1020.

CHAKRAVARTY, A., J. ORLIN AND U. G. ROTHBLUM. 1985. Consecutive Optimizators for a Partitioning Problem With Applications to Optimal Inventory Groupings for Joint Replenishment. *Opns. Res.* **33,** 820–834.

EPPEN, G., AND L. SCHRAGE. 1981. Centralized Ordering Policies in a Multi-Warehouse System With Leadtimes and Random Demand. In *Multi-Level Production / Inventory Control Systems: Theory and Practice*, L. Schwarz (ed.). North-Holland, Amsterdam, 51–67.

FEDERGRUEN, A., AND Y. ZHENG. 1988. The Joint Replenishment Problem With General Joint Cost Structures: General Solution Methods and Performance Bounds. *Opns. Res.* (to appear).

FEDERGRUEN, A., AND P. ZIPKIN. 1984a. Allocation Policies and Cost Approximations for Multi-Location Inventory Systems. *Naval Res. Logist. Quart.* **31,** 97–130.

FEDERGRUEN, A., AND P. ZIPKIN. 1984b. Approximations of Dynamic Multi-Location Production and Inventory Problems. *Mgmt. Sci.* **30,** 69–84.

FEDERGRUEN, A., AND P. ZIPKIN. 1984c. Computational Issues in an Infinite Horizon, Multi-Echelon Inventory Model. *Opns. Res.* **32,** 218–836.

FEDERGRUEN, A., AND P. ZIPKIN. 1986a. An Inventory Model With Limited Production Capacity and Uncertain Demands I: The Average Cost Criterion. *Math. Opns. Res.* **11,** 193–207.

FEDERGRUEN, A., AND P. ZIPKIN. 1986b. An Inventory Model With Limited Production Capacity and Uncertain Demands II: The Discounted Cost Criterion. *Math. Opns. Res.* **11,** 208–215.

FEDERGRUEN, A., AND P. ZIPKIN. 1988. Allocation Policies and Cost Approximation for Two-Echelon Inventory Systems With Centralized Stock (in preparation).

FLORIAN, M., AND M. KLEIN. 1971. Deterministic Production Planning With Concave Costs and Capacity Constraints. *Mgmt. Sci.* **18,** 18–20.

FLORIAN, M., J. K. LENSTRA AND A. H. G. RINNOOY KAN. 1980. Deterministic Production Planning: Algorithms and Complexity. *Mgmt. Sci.* **26,** 669–679.

GOYAL, S. K. 1973. Determination of Economic Packaging Frequency for Items Jointly Replenished. *Mgmt. Sci.* **20,** 232–235.

GOYAL, S. K. 1974a. Optimal Ordering Policy for a Multi-Item, Single Supplier System. *Opns. Res.* **25,** 293–298.

GOYAL, S. K. 1974b. Determination of Optimum Packaging Frequency of Items Jointly Replenished. *Mgmt. Sci.* **21,** 436–443.

GOYAL, S. K., AND A. S. BELTON. 1979. On 'A Simple Method of Determining Order Quantities in Joint Replenishments Under Deterministic Demand'. *Mgmt. Sci.* **26,** 604.

GRAVES, S. 1979. On the Deterministic Demand Multi-Product Single Machine Lot-Scheduling Problem. *Mgmt. Sci.* **25,** 276–280.

JACKSON, P., W. L. MAXWELL AND J. A. MUCKSTADT. 1985. The Joint Replenishment Problem With Power-of-Two Intervals. *IIE Trans.* **17,** 25–32.

JACKSON, P., W. L. MAXWELL AND J. A. MUCKSTADT. 1988. Determining Optimal Reorder Intervals in Capacitated Production-Distribution Systems. *Mgmt. Sci.* **35,** 938–958.

MAXWELL, W., AND J. MUCKSTADT. 1985. Establishing Consistent and Realistic Reorder Intervals in Production-Distribution Systems. *Opns. Res.* **33,** 1316–1341.

MUCKSTADT, J. 1985. Planning Component Delivery Intervals in Constrained Assembly Systems. In *Multi-Stage Production Planning and Inventory Control*, S. Axsäter, C. Schneeweiss and E. Silver (eds.). Springer Verlag, Berlin.

MUCKSTADT, J., AND R. ROUNDY. 1987. One-Warehouse, Multi-Retailer Distribution Systems. *Mgmt. Sci.* **33,** 1416–1430.

MUCKSTADT, J., AND R. ROUNDY. 1988. Analysis of Multi-Stage Production Systems. Technical Report No. 806, School of O.R. and I.E., Cornell University, Ithaca, New York.

NOCTURNE, D. J. 1972. Economic Ordering Frequency for Several Items Jointly Replenished. *Mgmt. Sci.* **19,** 1093–1096.

QUEYRANNE, M. 1987. Finding 94% Effective Policies in Linear Time for Some Production/Inventory Systems. Faculty of Commerce Working Paper, University of British Columbia, Vancouver.

ROUNDY, R. 1985. 98%-Effective Integer-Ratio-Lot-Sizing for One-Warehouse-Multi-Retailer-Systems. *Mgmt. Sci.* **31,** 1416–1430.

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

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

SHU, F. T. 1971. Economic Ordering Frequency for Two Items Jointly Replenished. *Mgmt. Sci.* **17,** B406–B410.

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

WIJNGAARD, J. 1972. An Inventory Problem With Constrained Order Capacity. TH-Report 72-WSK-63, Eindhoven University of Technology, Eindhoven, The Netherlands.

ZHENG, Y. 1987. Replenishment Strategies for Production/Distribution Networks With General Joint Setup Costs. Ph.D. Dissertation, Columbia University, New York.