MATHEMATICS OF OPERATIONS RESEARCH
Vol. 17, No. 4, November 1992
Printed in U.S.A.

SIMPLE POWER-OF-TWO POLICIES ARE CLOSE TO OPTIMAL IN A GENERAL CLASS OF PRODUCTION / DISTRIBUTION NETWORKS WITH GENERAL JOINT SETUP COSTS*

A. FEDERGRUEN, M. QUEYRANNE AND YU-SHENG ZHENG

We consider a production/distribution network represented by a general directed acyclic network. Each node is associated with a specific "product" or item at a given location and/or production stage. An arc (i, j) indicates that item i is used to "produce" item j. External demands may occur at any of the network's nodes. These demands occur continuously at item specific constant rates. Components may be assembled in any given proportions.

The cost structure consists of inventory carrying, and variable and fixed production/distribution costs. The latter depend, at any given replenishment epoch, on the specific set of items being replenished, according to an arbitrary set function merely assumed to be monotone and submodular.

We show that a simply structured, so-called *power-of-two* policy is guaranteed to come within 2% of a lower bound for the minimum cost. Under a power-of-two policy, all items are replenished at constant intervals and only when their inventory drops to zero; moreover, these replenishment intervals are all power-of-two multiples of a common base planning period. The above results generalize those of Roundy (1986).

1. Notation and preliminary results. Consider a production/distribution network represented by a general directed acyclic network $G = (N, A)$, with node set N and arc set A. Each node is associated with a specific "product", where a "product" represents a specific in-process or finished item, at a given physical location and/or production stage. With this general "product" definition, a directed arc (i, j) between a pair of nodes $i, j \in N$ indicates that product i is used to "produce" product j. The network is assumed to be acyclic to exclude circuits of products, each of which is consumed in producing its successor.

External demands may occur for any of the items, i.e., at any of the nodes in the network. These demands occur continuously at (item specific) constant rates. Components may be assembled in any given proportions. No backlogging is allowed. The cost structure consists of inventory carrying, variable and fixed production/distribution costs. Inventory carrying costs are incurred at constant rates per unit and per unit of time. Variable production costs are proportional to the production volumes. The value of these cost components is thus constant under any reasonable replenishment strategy, i.e., any strategy under which the items' long-run average production rates equal the corresponding demand rates—and may hence be ignored for the purpose of identifying optimal policies. The *fixed* production/distribution cost at any given replenishment epoch depends on the specific set of products being replenished

*Received October 5, 1989; revised May 7, 1991.
AMS 1980 subject classification. Primary: 90B30. Secondary: 90C25, 90C39.
IAOR 1973 subject classification. Main: Production. Cross references: Networks.
OR/MS Index 1978 subject classification. Primary: 346 Inventory/Production/Policies. Secondary: 111 Dynamic Programming.
Key words. Production networks, replenishment policies, power-of-two policy.

0364-765X/92/1704/0951/$01.25

according to an arbitrary (setup cost) set function merely assumed to be *monotone* and *submodular*:

(*monotonicity*) the fixed cost of a joint replenishment does not decrease by the inclusion of additional items;

(*submodularity*) the incremental fixed cost due to the addition of an extra item to a given set of replenishment activities is no larger than if the same item were added to a subset of these activities.

When an order is placed for a product, it is delivered instantaneously. Our objective is to minimize long-run average system-wide costs.

The structure of an optimal strategy may be exceedingly difficult even for the simplest of all multi-item models, e.g., the so-called "(first order interaction) joint replenishment problem" (see, e.g., Jackson et al. 1985 and Roundy 1985) in which a number of distinct products are interrelated only through a single joint setup cost incurred for any replenishment, independent of the specific products involved. The same is true in models with *separable* cost structures but physical interdependence between nodes in the network, such as single-item one-warehouse multiple-retailer systems. The complexity of fully optimal strategies makes them unattractive even if they could be computed efficiently.

We may, however, without loss of optimality, restrict ourselves to "*zero inventory ordering* policies" under which each product's inventory level equals zero at *any one* of its specified replenishment epochs. (For discrete-time models this is shown in Veinott 1969. See Zheng 1987 and Federgruen et al. 1989 for a simpler proof for our continuous time model; cf. also Schwarz 1973 for a proof for one-warehouse multiple-retailer systems with separable costs.)

In this paper, we show that a *simply structured*, so-called *power-of-two* policy is guaranteed to come within 2% of a lower bound for the minimum cost. A power-of-two policy applies zero-inventory ordering and prescribes for each product i a replenishment interval t_i such that a replenishment for this product occurs at times $0, t_i, 2t_i, 3t_i, \ldots$. Moreover, all products' replenishment intervals are chosen as powers-of-two multiples of a common base planning period. (This base planning period is sometimes predetermined but may be varied continuously in other settings. In the former case, the above worst case optimality gap is 6% instead of 2%.)

The above generalize earlier results of this type, in particular the seminal contribution of Roundy (1986) for the case of *separable* setup costs, or *joint* replenishment costs specified by a so-called "*family model*". (The family model represents a special case of submodular cost structures; see Federgruen and Zheng 1988b for a discussion of several alternative classes of submodular cost structures which cannot, or cannot efficiently, be represented by a family model.) Federgruen and Zheng (1988a) obtained the above results for "joint replenishment models" in which the interdependence between the different products is confined to joint (submodular) replenishment cost structures, without any technical or physical interdependencies (i.e., the corresponding network consists of a number of unconnected nodes, one for each item). The general model addressed in this paper was proposed and developed first by Queyranne (1985) and subsequently by Zheng (1987). The central observation, that, in inventory models with linear holding costs and constant demand rates, a cost increase of at most 6% results when restricting oneself to power-of-two multiples of a given base period, goes back to Brown (1978) (who demonstrated this for the single-item Economic Order Quantity model).

The above worst case optimality gap clearly motivates restricting oneself to power-of-two policies when searching for a close-to-optimal strategy. Moreover, while our paper does not address how an optimal power-of-two policy may be computed, several algorithms for this problem are indeed available. Queyranne's (1985) algorithm uses Maxwell and Muckstadt's (1985) divide-and-conquer method and requires at most $2|N| - 1$ calls to a general oracle for minimizing a submodular function (e.g., the ellipsoid method, Grötschel et al. 1981). Zheng (1987) describes two more efficient algorithms which consist of a limited number of polymatroid maximum flow computations. The only oracle required in these algorithms is one for polymatroid membership. In contrast with Queyranne's approach, this oracle can be specific to the submodular setup cost function being used. The networks in which these maximum flows need to be determined have, in addition to the usual upper bounds on individual arcs, additional capacity constraints for collections of arcs pointing to the network's sink. These maximum flow problems thus represent special cases of the *polymatroidal* network flow problems introduced by Hassin (1978, 1982) and Lawler and Martel (1982). See Hassin, Lawler and Martel, and Tardos et al. (1986) for efficient algorithms for this general class of maximum flow problems. The *number* of maximum flow computations is at most $2|N| - 1$ in the first of the two algorithms in Zheng (1987) and bounded by the number of *distinct* components in the optimal power-of-two vector in the second algorithm.

Earlier algorithms for special cases include the seminal papers of Maxwell and Muckstadt (1985) and Roundy (1986). The former address networks with *separable* costs, restricting themselves to *nested* power-of-two policies. (Under a nested policy, each node places an order each time one of its immediate predecessor nodes in the network does; the nestedness condition simplifies the analysis and computations significantly but, as pointed out in Roundy 1985, such policies may have a rather poor cost performance.) Roundy (1986) considers all (nested and nonnested) power-of-two policies under separable costs or nonseparable structures generated by the above-mentioned family model.

In §2 we specify the model, and formulate the problem of finding an optimal power-of-two policy as one of finding a saddlepoint of a nonlinear program with special integrality constraints implied by the power-of-two restrictions. In §3 we derive alternative formulations for the continuous relaxation of this program (obtained by relaxing the power-of-two integrality constraints). These are needed for the above described worst case performance analysis of power-of-two policies, which can be found in §4. In §5 we show that the same worst case optimality gaps arise when restricting oneself to *nested* policies only.

2. Notation and preliminary results. For each node $i \in N$, let $P(i)$ indicate the set of its immediate predecessors in the network, i.e., $P(i) = \{j \in N | (j, i) \in A\}$. ($P(i)$ represents the set of products which are used as inputs in the production of product i.) For each arc $(i, j) \in A$, λ_{ij} represents the number of units of product i required to produce one unit of product j. Let d_i represent the rate at which *external* demands for product i arise; h'_i denotes the cost per unit of time for carrying one unit of product i in inventory. The incremental holding cost rate for product i is given by

$$h_i = h'_i - \sum_{j \in P(i)} \lambda_{ji} h'_j.$$

These incremental holding cost rates are assumed to be nonnegative, i.e., $h_i \geqslant 0$, $i \in N$. Let $N_0 \overset{\text{def}}{=} \{i \in N\colon P(i) = \varnothing\}$ be the set of basic products (whose corresponding nodes have no predecessors). We assume that $h_i > 0$ for all $i \in N_0$. This assumption is without loss of generality: if $h_i = 0$ for some $i \in N_0$, infinitely large quantities of this product may be ordered from external sources with zero resulting inventory costs. Such a product may clearly be eliminated from the model.

With respect to the setup cost set function $K(\cdot)$ we assume:
(*Monotonicity*) $0 = K(\emptyset) \leqslant K(S) \leqslant K(T)$ if $S \subset T$.
(*Submodularity*) $K(S \cup \{i\}) - K(S) \geqslant K(T \cup \{i\}) - K(T)$ if $S \subset T$, $i \notin T$.

If $K(\cdot)$ fails to be monotone, replace $K(\cdot)$ by the monotone set function $\bar{K}$, defined by $\bar{K}(S) = \min\{K(T): T \supseteq S\}$, $S \subset N$. Note that $\bar{K}(S) \leqslant K(S)$, $S \subset N$; an expense of $\bar{K}(S)$ allows for the collection of products S, as well as possibly some others, to be replenished jointly. If $K(S) < K(T)$ for all $S \subset T$, the set function is called *strictly monotone*. The second assumption (submodularity) reflects *economies of scale* in the joint replenishment activities. Federgruen and Zheng (1988a) discuss several common setup cost structures under which the submodularity assumption is satisfied.

We assume that the system starts with zero inventory throughout. Finally it is assumed that when an order is placed for one or more products, it is delivered instantaneously. The planning problem becomes significantly harder under general nonzero leadtimes. See, however, Roundy (1986) for certain special cases which *are* tractable. Let C^* denote the minimum long-run average system-wide cost.

One of the advantages of power-of-two policies is the fact that relatively simple formulations may be obtained for their long-run average system-wide costs. For any power-of-two policy with replenishment interval vector t let $H[t]$ ($K[t]$) denote the long-run average system-wide holding (setup) costs. The problem of finding an optimal power-of-two policy is thus summarized by:

$$\min_{t>0} C[t] = K[t] + H[t]$$

$$\text{subject to } t_i = 2^{m_i} T_L \quad (m_i \text{ integer}), \qquad i \in N,$$

where T_L denotes the (fixed or variable) base planning period.

Let R denote the set of *routes* in the network G where a route $r = (i_1, \ldots, i_m)$ is a directed path in G starting at an arbitrary node and terminating at an end-product. For each route $r = (i_1, \ldots, i_m) \in R$ let $d_r = \lambda_{i_1 i_2} \cdots \lambda_{i_{m-1} i_m} d_{i_m}$ and $H_r = \frac{1}{2} h_{i_1} d_r$ denote its (induced) demand and holding cost rate, respectively. It is shown in Roundy (1986) (and Lemma 2.7 in Zheng 1987) that

$$H[t] = \sum_{r \in R} H_r\left(\max_{i \in r} t_i\right). \tag{1}$$

To derive an expression for $K[t]$ let $\alpha = (\alpha_1, \ldots, \alpha_n)$ be a permutation of the indices $N = \{1, 2, \ldots, n\}$ such that $t_{\alpha_1}^{-1} \geqslant t_{\alpha_2}^{-1} \geqslant \cdots \geqslant t_{\alpha_n}^{-1}$, i.e., the nodes' reorder frequencies are nonincreasing in this permutation. Observe that under t whenever node α_i is replenished, all of the nodes $\{\alpha_1, \ldots, \alpha_{i-1}\}$ are replenished as well. Only one of the following n sets of nodes is therefore replenished at any replenishment epoch:

$$\{\alpha_1\}, \{\alpha_1, \alpha_2\}, \ldots, \{\alpha_1, \ldots, \alpha_i\}, \ldots, \{\alpha_1, \ldots, \alpha_n\}.$$

The order frequency of the set $\{\alpha_1, \ldots, \alpha_i\}$ is $t_{\alpha_i}^{-1} - t_{\alpha_{i+1}}^{-1}$ (setting $t_{\alpha_{n+1}}^{-1} = 0$). Therefore

$$K[t] = \sum_{i \in N} K(\{\alpha_1, \ldots, \alpha_i\})\left(t_{\alpha_i}^{-1} - t_{\alpha_{i+1}}^{-1}\right).$$

By rearranging the terms in this summation, we obtain:

$$K[t] = \sum_{i \in N} \left[K(\{\alpha_1, \ldots, \alpha_i\}) - K(\{\alpha_1, \ldots, \alpha_{i-1}\}) \right] t_{\alpha_i}^{-1}. \tag{2a}$$

Let

$$K = \left\{ k \in R^n \colon \sum_{i \in S} k_i \leqslant K(s),\ S \subset N,\ \sum_{i \in N} k_i = K(N),\ k \geqslant 0 \right\}.$$

K is referred to as the *setup cost polyhedron*. Note that each vector $k \in K$ may be viewed as an *allocation* of the joint setup cost structure to the individual products, as in Roundy (1986) and Atkins and Iyogun (1987).

LEMMA 0.

$$K[t] = \max_{k \in K} \sum_{i \in N} k_i / t_i. \tag{2b}$$

PROOF. Since $K(\cdot)$ is monotone and submodular, and $K(\varnothing) = 0$, K is a so-called (base of a) *polymatroid*, see Edmonds (1970) or Welsh (1976). It is well known from Edmonds that, on this special type of polyhedron, any linear objective may be optimized by the *greedy procedure*. More specifically, the linear program to the right of (2b) has an optimal solution k^* with $k^*_{\alpha_i}$ given by the expression in square brackets in (2a), $i = 1, \ldots, N$. □

Substituting (1) and (2) into the objective function of (P) we obtain the mathematical program:

$$\min_{t, \tau} \max_{k \in K} \left\{ \sum_{i \in N} k_i / t_i + \sum_{r \in R} H_r \tau_r \right\} \tag{3)(P}$$

$$\text{subject to } \tau_r = \max_{i \in r} t_i, \qquad r \in R, \tag{4a}$$

$$t_i = 2^{m_i} T_L \quad (m_i \text{ integer}), \qquad i \in N. \tag{4b}$$

Since $H_r \geqslant 0$, the optimal objective value for (P) remains unchanged when relaxing the constraints (4a) to:

$$\tau_r \geqslant t_i, \quad \text{all } r \in R \text{ and } i \in r. \tag{4a$'$}$$

We redefine $t := (t, \tau)$ where $\tau = \{\tau_r \colon r \in R\}$.

3. The continuous relaxation of (*P*): alternative formulations. Except for the special integrality constraints (4b), the problem of finding an optimal power-of-two policy ((P)) can thus be formulated as one of finding a saddlepoint of a so-called convex-concave function in the vector pair (k, t) with both k and t to be chosen from separate polyhedra of special structure. For, let $\mathbf{N} = N \cup R$, $\tilde{T} = \{t \in R^{|\mathbf{N}|} \colon t > 0, t$ satisfies (4a′), (4b)}. Define the function $f(t, k) = \Sigma_{i \in N} k_i / t_i + \Sigma_{r \in R} H_r \tau_r$ and note that $f(t, k)$ is *convex-concave*, i.e., it is convex in t and linear and hence concave in k.

Problem (P) may then be written as

$$\min_{t \in \tilde{T}} \max_{k \in K} f(t, k). \tag{5}$$

See Rockafellar (1970, Part VII) for a treatment of general saddlepoint (minimax) problems of convex-concave functions.

In this section we consider the *continuous relaxation* of (P) obtained by relaxing the integrality constraints (4b). Thus, let $T = \{t \in R^{|N|}: t > 0, t \text{ satisfies (4a')}\}$:

$$\text{(6)(RP)} \qquad c^* = \min_{t \in T} \max_{k \in K} f(t, k).$$

Theorem 1 below shows that the optimal value of the *continuous* relaxation (RP) constitutes a lower bound for the minimum system-wide costs under *any* strategy (in addition to it obviously being a lower bound for the cost value of the best power-of-two policy). This result, while of interest by itself, also provides the foundation for the proof of our main result (Theorem 2), by establishing the existence of a power-of-two policy $\bar{t}^*$ with cost $C[\bar{t}^*] \leqslant 1.021c^*$ and hence $C[\bar{t}^*] \leqslant 1.021C^*$.

To prepare the proof of this theorem, we derive an equivalent formulation of (RP). We first show that the minimum and maximum operators in (6) may be interchanged.

LEMMA 1. *Let Θ be a compact and convex subset of some Euclidean space and let f: $\Theta \times K \to R$ be a continuous function which is convex in $t \in \Theta$ and concave in $k \in K$. Then*

$$\min_{t \in \Theta} \max_{k \in K} f(t, k) = \max_{k \in K} \min_{t \in \Theta} f(t, k).$$

PROOF. See Rockafellar (1970). □

An interchange of the max- and min-operators in (6) is now simply justified.

LEMMA 2 (Minimax Theorem for the continuous relaxation of (P)). *Assume $K(\cdot)$ is strictly monotone.*

$$\text{(7)(RP)} \qquad c^* = \min_{t \in T} \max_{k \in K} \left\{ \sum_i k_i/t_i + \sum_r H_r \tau_r \right\}$$

$$= \max_{k \in K} \min_{t \in T} \left\{ \sum_i k_i/t_i + \sum_r H_r \tau_r \right\}.$$

PROOF. Let $\delta \stackrel{\text{def}}{=} \min\{K(N) - K(N \setminus \{a\}): a \in N\}$. Since $K(\cdot)$ is strictly monotone, we have $\delta > 0$. Note that for all $k \in K$ and $a \in N$, $k_a = K(N) - \sum_{i \neq a} k_i \geqslant K(N) - K(N \setminus \{a\}) \geqslant \delta$. Let $\overline{T}_L = T_L$ if T_L is fixed, and $\overline{T}_L$ be an arbitrary allowable value for T_L if T_L is variable. Defining $t^L \in T$ by $t_i^L = \overline{T}_L$ for all $i \in N$, we have $\tau_r^L = \overline{T}_L$ for all $r \in R$. Thus for all $k \in K$ we have

$$f(t^L, k) = K(N)/\overline{T}_L + H(N)\overline{T}_L \stackrel{\text{def}}{=} c^0,$$

where $H(N) \equiv \sum_{r \in R} H_r$. Let

$$t^- \stackrel{\text{def}}{=} \delta/c^0 \quad \text{and} \quad t^+ \stackrel{\text{def}}{=} c^0 \Big/ \left(\tfrac{1}{2} \min_{r \in R} d_r \min_{i \in N_0} h_i \right).$$

Define $\Theta = \{t: t^- 1 \leqslant t \leqslant t^+ 1, \text{(4a)}\}$ with $1 \in R^{|N|}$ a vector of ones. Since Θ is compact and convex and f has the properties required in Lemma 1 we have for some

$(t^*, k^*) \in \Theta \times K$:

$$\max_{k \in K} \min_{t \in \Theta} f(t,k) = \min_{t \in \Theta} \max_{k \in K} f(t,k) = f(t^*, k^*) \leqslant c^0 \tag{8}$$

where the inequality follows from the vector $\overline{T}_L 1 \in \Theta$. ($t^- \leqslant T_L$ since $\delta \leqslant K(N) \leqslant K(N) + H(N)\overline{T}_L^2 = c^0 \overline{T}_L$; $\overline{T}_L \leqslant t^+$ since

$$\tfrac{1}{2}\Big(\min_{r \in R} d_r\Big)\Big(\min_{i \in N_0} h_i\Big)\overline{T}_L \leqslant H(N)\overline{T}_L \leqslant c^0.\Big)$$

In view of (8), it suffices to show that

$$f(t,k) \geqslant c^0 \quad \text{for all } t \in T \setminus \Theta \text{ and } k \in K. \tag{9}$$

Thus, fix $t \in T \setminus \Theta$ and $k \in K$.

Observe that it suffices to consider the following two cases:

(1) $t_a = \min_{i \in N} t_i < t^-$ *for some* $a \in N$. Clearly,

$$f(t,k) \geqslant k_a / t_a \geqslant \delta / t^- = c^0. \tag{10a}$$

(2) $\max_{r \in R} \tau_r > t^+$. Note, in view of (4a′), that $\max \tau_r$ is achieved for some route l whose initial node $a \in N_0$. Hence,

$$f(t,k) \geqslant \sum_{r \in R} H_r \tau_r > H_l t^+ \geqslant \tfrac{1}{2}\Big(\min_{r \in R} d_r\Big)\Big(\min_{i \in N_0} h_i\Big) t^+ = c^0. \tag{10b}$$

Thus (9) follows from (10a) and (10b). □

Note that, for any $k \in K$, the (inner) minimization to the right of (7) is a *convex program* which may therefore be replaced by its dual. This allows us to replace the saddlepoint representation of (RP) by a pure maximization problem:

LEMMA 3. *Assume $K(\cdot)$ is strictly monotone.*

$$c^* = \max \sum_{i \in N} 2(k_i v_i)^{1/2}, \tag{11a}$$

$$\text{subject to } k \in K, \tag{11b}$$

$$H_r = \sum_{i \in r} x_{ri}, \qquad r \in R, \tag{11c}$$

$$\sum_{r: i \in r} x_{ri} = v_i, \qquad i \in N, \tag{11d}$$

$$x \geqslant 0, \quad v \geqslant 0. \tag{11e}$$

PROOF. For any $r \in R$ and $i \in R$, let x_{ri} be a Lagrange multiplier associated with constraint (4a′). Fix $k^0 \in K$, let (RP_{k^0}) denote the convex (inner) minimization problem to the right of (7) with $k_i = k_i^0$ $(i \in N)$ and let $c^*(k^0)$ denote its optimal value. The associated Lagrangian of (RP_{k^0}) is given by

$$L(x,t) = \sum_{i \in N} k_i^0 / t_i + \sum_{r \in R} H_r \tau_r + \sum_{r \in R} \sum_{i \in r} x_{ri}(t_i - \tau_r). \tag{12}$$

By definition, see Geoffrion (1971), the Lagrangian dual of (RP_{k^0}) can be written as

$$D = \sup_{x \geqslant 0} \inf_{t>0} L(x,t). \tag{13}$$

Letting

$$v_i = \sum_{r:\, i \in r} x_{ri}, \qquad i \in N, \tag{14}$$

and regrouping the terms in the second summation in (12), we have

$$L(x,t) = \sum_{i \in N} \left[k_i^0 / t_i + v_i t_i \right] + \sum_{r \in R} \left(H_r - \sum_{i \in r} x_{ri} \right) \tau_r.$$

If, for a given $x \geqslant 0$, $(H_r - \sum_{i \in r} x_{ri}) < 0$, then $\inf L(x,t) = \lim_{t \to \infty} L(x,t) = -\infty$; likewise, if $(H_r - \sum_{i \in r} x_{ri}) \geqslant 0$, $\inf_{\tau_r > 0} (H_r - \sum_{i \in r} x_{ri}) \tau_r = 0$. To achieve the supremum in (13), we may thus restrict ourselves to vectors x for which

$$H_r \geqslant \sum_{i \in r} x_{ri}, \qquad r \in R. \tag{15}$$

Imposing these additional constraints and substituting (14), we may rewrite (13) as:

$$\begin{aligned} D &= \sup_{\{x \geqslant 0,\, v \geqslant 0:\, (14), (15)\}} \inf_{t > 0} \sum_{i \in N} \left[k_i^0 / t_i + v_i t_i \right] \\ &= \sup_{\{x \geqslant 0,\, v \geqslant 0:\, (14), (15)\}} \sum_{i \in N} 2 \left(k_i^0 v_i \right)^{1/2}. \end{aligned} \tag{16}$$

Strong duality, i.e., $c^*(k^0) = D$, and the existence of a pair (x^*, v^*) achieving the supremum in (16) all follow from (RP_{k^0}) being a *stable convex* program, see Theorem 3 in Geoffrion (1971). Stability may be verified by Slater's condition, i.e., there exists a vector $t > 0$, with $\tau_r > T_i$ for all $r \in R$ and $i \in r$. (Let $\tau_r = 2$, all $r \in R$ and $T_i = 1$, all $i \in N$.) Finally note that in an optimal solution (x^*, v^*), (15) must be satisfied as *equalities* and may thus be replaced by (11c). The lemma thus follows from (16) and the subsequent observations. □

Note that each vector $v \in R^n$ which satisfies (11c) and (11d) may be viewed as a vector of (re)allocated holding cost rates, as in Roundy (1985, 1986). The first of the two algorithms in Zheng (1987), discussed in the introduction, finds an optimal solution (t^*, k^*) for (RP) as well as a corresponding optimal dual solution (x^*, v^*), see Lemma 3.

4. A lower bound theorem and worst case analysis of power-of-two policies. In this section we show that the optimal value of the continuous relaxation (RP) of (P) constitutes a lower bound for the *minimum* system-wide average cost under *any* strategy. This allows us to demonstrate that the average cost of an optimal power-of-two policy comes within 6% or 2% of this lower bound depending upon whether the base planning period T_L is fixed or variable.

We first need the following preliminaries: each unit of each product may clearly be assigned to a specific route and it is thus possible to distinguish between different units of the same product according to the specific routes they have been assigned to. Consider a given policy, time $t > 0$, and route $r = (i_1, \ldots, i_m)$. Define, as in Roundy (1986), route r's echelon inventory E_r^t as the total number of units of product i_1, which are held in stock somewhere along the route r at time t (perhaps as components of more advanced products) and which have been specified to follow route r, measured in multiples of $\frac{1}{2}d_r$, i.e., as twice the number of time units of demand for product r's (unique) end item which this inventory is capable of supporting. (Recall d_r is the induced demand rate for route r, see §2 for a precise definition.) For each $i \in N$ and $t > 0$, define

$$\eta_i^t = \min\{E_r^t \colon r \in R_i\} \tag{17}$$

where R_i is the set of routes starting at node i.

THEOREM 1 (The Lower Bound Theorem). *The optimal value c^* of* (RP) *is a lower bound for the average cost of any feasible policy over any finite horizon.*

PROOF. Assume first that $K(\cdot)$ is strictly monotone so that Lemmas 2 and 3 apply. Let t^* achieve the minimum in (RP), see Lemma 2. It follows from Lemma 3 that a vector $k^* \in K$, and vectors x^*, v^* exist which satisfy (11c), (11d), (11e) and with $\sum_{i \in N} 2(k_i^* v_i^*)^{1/2} = c^*$.

For $\tau > 0$ and a feasible policy π, let c be the total cost incurred by the policy during the time interval $[0, \tau)$. We show that $c \geqslant c^* \tau$. We evaluate the total setup costs and holding costs separately.

The total setup cost K^{tot}. Let $J(S)$ be the number of times in $[0, \tau)$ that an order is placed specifically for the set of items $S \subseteq N$. The number of times product i is ordered in $[0, \tau)$ is then given by $J_i = \sum_{S:\, i \in S} J(S)$. Thus, since $k^* \in K$,

$$\begin{aligned} K^{\mathrm{tot}} &= \sum_{S \subseteq N} J(S) K(S) \geqslant \sum_{S \subseteq N} J(S) \sum_{i \in S} k_i^* = \sum_{i \in N} k_i^* \sum_{S:\, i \in S} J(S) \\ &= \sum_{i \in N} k_i^* J_i. \end{aligned}$$

The total holding cost H^{tot}. Following the discussion in the introduction, we assume without loss of generality that π applies zero inventory ordering. The rate at which the total holding cost accumulates at time t is then given by

$$\begin{aligned} \sum_{r \in R} H_r E_r^t &= \sum_{r \in R} \sum_{i \in r} x_{ri}^* E_r^t = \sum_{i \in N} \sum_{r:\, i \in r} x_{ri}^* E_r^t \\ &\geqslant \sum_{i \in N} \sum_{r:\, i \in r} x_{ri}^* \eta_i^t = \sum_{i \in N} v_i^* \eta_i^t. \end{aligned}$$

(The first and last equalities follow from (11c) and (11d), respectfully. To verify the inequality note that for any route $r = (i_1, \ldots, i, \ldots, i_m) \ni i$, $\eta_i^t \leqslant E_{r'}^t \leqslant E_r^t$ where $r' = (i, \ldots, i_m)$.) We conclude that

$$H^{\mathrm{tot}} \geqslant \sum_{i \in N} \int_0^\tau v_i^* \eta_i^t \, dt.$$

Therefore

$$c = K^{\mathrm{tot}} + H^{\mathrm{tot}} \geqslant \sum_{i \in N} \left\{ k_i^* J_i + \int_0^\tau v_i^* \eta_i^t \, dt \right\}. \tag{18}$$

For any $i \in N$, note that $\{\eta_i^t,\ t \geqslant 0\}$ decreases at a constant rate of 2 and increases only at replenishment epochs of product i. (The latter follows from (17) and the observation that E_r^t increases only at replenishment epochs of product i for any route r which starts at product i.) For any $i \in N$, the expression within curled brackets in (18) may thus be interpreted as the total setup and holding costs in $[0, \tau)$ in a single-item model with constant demand rate 2, setup cost k_i^*, holding cost rate v_i^*, initial inventory zero and under an arbitrary policy prescribing J_i setups in $[0, \tau)$. It follows from Carr and Howe (1962) that the expression is minimized by ordering $2\tau/J_i$ units every τ/J_i time units. We conclude that

$$c \geqslant \sum_{i \in N} \left(k_i^* J_i + v_i^* \tau^2 / J_i\right) \geqslant \sum_{i \in N} 2(k_i^* v_i^*)^{1/2} \tau = c^* \tau.$$

If $K(\cdot)$ fails to be strictly monotone, define for all $\epsilon > 0$, the perturbed cost function $K_\epsilon(\cdot)$ by $K_\epsilon(S) = K(S) + \epsilon|S|$ for all $S \subseteq N$. (One easily verifies that $K_\epsilon(\cdot)$ is strictly monotone, and continues to be submodular.) Let C_ϵ^* denote the minimum system wide cost and c_ϵ^* the optimal values of (RP) under the setup cost function $K_\epsilon(\cdot)$. By the above proof, we have $c_\epsilon^* \leqslant C_\epsilon^*$ for all $\epsilon > 0$. Moreover, $c^* = \lim_{\epsilon \downarrow 0} c_\epsilon^*$ and $C^* = \lim_{\epsilon \downarrow 0} C_\epsilon^*$. Thus, $c^* \leqslant C^*$. □

The proof of the above theorem bears similarity to that in Roundy (1986) for the model considered ibid.

As in prior analyses of this type (Maxwell and Muckstadt 1985, Roundy 1986, Federgruen and Zheng 1988a), the worst case analysis of power-of-two policies is completed by showing that any optimal replenishment vector t for (RP) may be replaced by a feasible power-of-two vector $\bar{t}^*$ whose (long-run average) cost value is at most 6% (2%) higher than c^* when the base planning period T_L is fixed (variable).

In case T_L is fixed, it suffices to replace each component i of t^* by a power-of-two multiple of T_L which is closest in a relative sense, i.e., such that $\max\{t_i^*/\bar{t}_i^*, \bar{t}_i^*/t_i^*\}$ is closest to one.

If T_L may be varied, its optimal value may be found by a simple $O(n \log n)$ search procedure, see Roundy (1985). We are now ready for our main result:

THEOREM 2 (Worst case analysis of power-of-two policies). (a) *Assume T_L is fixed. There exists a power-of-two policy $\bar{t}^*$ whose cost value* $C[\bar{t}^*] \leqslant 1.061c^* \leqslant 1.061C^*$. (b) *Assume T_L is variable. There exists a power-of-two policy $\bar{t}^*$ whose cost value* $C[\bar{t}^*] \leqslant 1.021c^* \leqslant 1.021C^*$.

PROOF. (a) See appendix. The proof constructs a power-of-two vector $\bar{t}^*$ with

$$C[\bar{t}^*] \leqslant 1.061c^*. \tag{19}$$

Hence, by Theorem 1, $C[\bar{t}^*] \leqslant 1.061C^*$. Part (b) follows from part (a) and Roundy (1986). □

The proof of Theorem 2(a) is a variant of that of Theorem 5 in Federgruen and Zheng (1988a) and the 94% theorem in Roundy (1986).

We note that, if $K(\cdot)$ is strictly monotone, the power-of-two policy generated by applying the above rounding procedure to an optimal solution of (RP) is in fact *optimal* among power-of-two policies, i.e., it solves (P). See Zheng (1987) for a proof.

5. Nested policies. As pointed out in the introduction, it is considerably simpler to compute optimal *nested* power-of-two policies instead of optimal general power-of-two policies. (On the other hand, as pointed out by Roundy 1985, and referred to in the introduction, all nested policies may be significantly more expensive than other feasible policies.) In this section we show that the 2% or 6% worst case optimality gaps for power-of-two policies are preserved when restricting oneself to nested, but otherwise general (stationary and nonstationary) policies. In other words, best *nested* power-of-two policy comes within 2% or 6% of the minimum cost $\bar{c}^*$ among all *nested* (stationary and nonstationary) policies.

The long-run average holding cost $H[t]$ of a *nested* power-of-two policy t is given by a simple linear function of $t = (t_1, \dots, t_n)$; i.e., there exist numbers $\{H_i\colon i \in N\}$ such that $H[t] = \sum_{i \in N} H_i t_i$ for *all* power-of-two vectors t; see Maxwell and Muckstadt (1985). This greatly simplifies the formulation of the objective function of (P) at the expense of additional constraints:

$$(20) \qquad t_i \geqslant t_j, \quad \text{all } (i,j) \in A,$$

to enforce the restriction to nested (power-of-two) policies. The problem of determining an *optimal nested* power-of-two policy may thus be formulated as:

$$(21)(\text{PN}) \qquad \min_{t>0} \max_{k \in K} \left\{ \sum_{i \in N} [k_i/t_i + H_i t_i] \right\},$$

$$(22) \qquad \text{subject to } t_i \geqslant t_j, \quad \text{all } (i,j) \in A,$$

$$(23) \qquad t_i = 2^{m_i} T_L \quad (m_i \text{ integer}), \qquad i \in N.$$

Let $T_1 = \{t > 0\colon (22)\}$.

Following the analysis of §3, one easily verifies that the *continuous* relaxation of (PN), i.e., the problem of finding the best nested power-of-two policy (obtained by relaxing (23)), may be written as:

$$(24)(\text{RPN}) \qquad \bar{c}^* = \min_{t \in T_1} \max_{k \in K} \left\{ \sum_{i \in N} [k_i/t_i + H_i t_i] \right\}$$

$$= \max_{k \in K} \min_{t \in T_1} \left\{ \sum_{i \in N} [k_i/t_i + H_i t_i] \right\}.$$

THEOREM 3 (Lower bound theorem for nested policies). *The minimum value $\bar{c}^*$ of* (RPN) *is a lower bound for the average cost of any* (*stationary or nonstationary*) *feasible nested policy over any finite horizon.*

PROOF. The proof is analogous to that of Theorem 1. For each $i \in N$, and $t > 0$, let E_i^t denote the echelon inventory of product i. It is easily verified that, under any *nested* policy, the rate at which holding costs are incurred at time t is given by $\sum_{i \in N} H_i E_i^t$. Let (t^*, k^*) be a saddle point of (RPN). Fix $\tau > 0$ and a feasible nested policy. Adopting the notation in the proof of Theorem 2 one concludes that

$$c \stackrel{\text{def}}{=} K^{\text{tot}} + H^{\text{tot}} \geqslant \sum_{i \in N} \left\{ k_i^* J_i + \int_0^\tau H_i E_i^t \, dt \right\} \geqslant \sum_{i \in N} \left\{ k_i^* J_i + H_i \tau^2 / J_i \right\}$$

where the second inequality as well as $K^{\text{tot}} \geqslant \sum_{i \in N} k_i^* J_i$ are verified as in the proof of Theorem 2. Since the policy is nested we have $J_i \leqslant J_j$ for all $(i,j) \in A$. Let $t_i = \tau / J_i$ $(i \in N)$. We conclude that

$$c \geqslant \min \left\{ \sum_{i \in N} \{k_i^*/t_i + H_i t_i\} \tau \colon t_i \geqslant t_j \text{ for all } (i,j) \in A \right\} = \bar{c}^* \tau. \quad \square$$

By straightforward adaptation of the analysis in §4 we conclude:

COROLLARY 1 (Worst case analysis of nested power-of-two policies). *There exists a nested power-of-two policy* t^* *whose cost value*

$$C[t^*] \leqslant 1.021\bar{c}^* \leqslant 1.021\bar{C}^* \quad \text{in case } T_L \text{ is variable}, \quad \text{and}$$

$$C[t^*] \leqslant 1.061\bar{c}^* \leqslant 1.061\bar{C}^* \quad \text{in case } T_L \text{ is fixed}.$$

Appendix. Proof of Theorem 2. We assume that $K(\cdot)$ is *strictly* monotone so that Lemma 2 applies. If it is not, the proof can be amended by considering a sequence of perturbed setup cost functions, as in the proof of Theorem 1. In view of Lemma 2, let t^* denote an optimal solution to (RP) and $\bar{t}^*$ the power-of-two vector obtained by the above rounding procedure. Let $(\alpha_1, \ldots, \alpha_n)$ be a permutation of the node indices such that $t_{\alpha_1}^{*-1} \geqslant t_{\alpha_2}^{*-1} \geqslant \cdots \geqslant t_{\alpha_n}^{*-1}$ Note that the components of $\bar{t}^*$ may be ranked in the same way. Assume t^* takes on M distinct values $t^*(1) < \cdots < t^*(M)$ and let $\bar{t}^*(1) \leqslant \cdots \leqslant \bar{t}^*(M)$ denote their rounded power-of-two values. For all $l = 1, \ldots, M$ let $N_l = |\{i \in N: t_i^* = t^*(l)\}|$;

$$\bar{K}_l = K(\{\alpha_1, \ldots, \alpha_{N_1 + \cdots + N_l}\}) - K(\{\alpha_1, \ldots, \alpha_{N_l + \cdots + N_{l-1}}\})$$

$$= \sum_{i = N_1 + \cdots + N_{l-1} + 1}^{N_1 + \cdots + N_l} [K(\{\alpha_1, \ldots, \alpha_i\}) - K(\{\alpha_1, \ldots, \alpha_{i-1}\})],$$

$$\bar{H}_l = \sum_{\{r \in R:\, \tau_r = t(l)\}} H_r.$$

It follows from Lemma 0 that

$$c^* = \sum_{l=1}^{M} \left[\bar{H}_l t^*(l) + \bar{K}_l / t^*(l)\right] \quad \text{and} \quad C[\bar{t}^*] = \sum_{l=1}^{M} \left[\bar{H}_l \bar{t}^*(l) + \bar{K}_l / \bar{t}^*(l)\right].$$

$t^*(l)$ is clearly the unique minimand of the EOQ cost function $\{\bar{H}_l x + \bar{K}_l / x: x > 0\}$ $(l = 1, \ldots, M)$. (If $t^*(l)$ is bigger (smaller) than this minimum, it could be decreased (increased) to a new value $t'(l)$, resulting in a new feasible solution t' of (RP) with $\max_{k \in K} f(t', k) < c^*$, contradicting the optimality of t^*.) Thus, c^* and $C[\bar{t}^*]$ represents the sum of M independent EOQ cost functions evaluated at the minimizing intervals $t^*(l)$ $(l = 1, \ldots, M)$ and intervals $\bar{t}^*(l)$ with $1/\sqrt{2} \leqslant \bar{t}^*(l)/t^*(l) \leqslant \sqrt{2}$, respectively. It is thus easily verified and well known that each of the terms in $C[\bar{t}^*]$ is at most 6.1% larger than the corresponding term in c^*. Hence $C[\bar{t}^*] \leqslant 1.061 c^* \leqslant 1.061 C^*$, by Theorem 1. □

References

Atkins, D. and Iyogun, P. (1987). A Lower Bond for a Class of Inventory/Production Problems. *Oper Res. Lett.* **6** 62.

Brown, R. G. (1978). Inventory Control. *Handbook of Operations Research* **2** 173–212.

Carr, C. R. and Howe, C. W. (1962). Optimal Service Policies and Finite Time Horizons. *Management Sci.* **9** 126–140.

Edmonds, J. (1970). Submodular Functions, Matroids and Certain Polyhedra. in *Combinatorial Structures and Their Applications*. R. Guy et al. (Eds.), Gordon and Breach, New York, 69–87.

Federgruen, A. and Zheng, Y. S. (1988a). The Joint Replenishment Problem with General Joint Cost Structures: General Solution Methods and Performance Bounds. *Oper. Res.* (to appear).

Federgruen, A. and Zheng, Y. S. (1988b). Minimizing Submodular Set Functions: Efficient Algorithms for Special Structures with Applications to Joint Replenishment Problems. Columbia University, Graduate School of Business, Working Paper, New York.

_____, Queyranne, M. and Zheng, Y. S. (1989). Simple Power of Two Policies Are Close to Optimal in a General Class of Production/Distribution Networks with General Joint Setup Costs. (unabridged version of this paper).

Geoffrion, A. (1971). Duality in Nonlinear Programming: A Simplified Applications Oriented Development. *SIAM Rev.* **13**, **1** 1–37.

Grötschel, M., Lovasz, L. and Schrijver, A. (1981). The Ellipsoid Method and Its Consequences in Combinatorial Optimization. *Combinatorica* **1** 169–197.

Hassin, R. (1978). On Network Flows. Ph.D. Dissertation, Yale University.

_____. (1982). Minimum Cost Flow in the Set-Constraints. *Networks* **12** 1–21.

Jackson, P. L., Maxwell, W. L. and Muckstadt, J. A. (1985). The Joint Replenishment Problem with a Powers of Two Restriction. *IIE Trans.* **17** 25–32.

Lawler, E. L. (1982). Generalization of the Polymatroidal Network Flow Model. Report BW 158, Mathematisch Centrum, Amsterdam.

_____ and Martel, C. U. (1982). Computing Maximal "Polymatroidal" Network Flows. *Math. Oper. Res.* **7** 334–346.

Maxwell, W. L. and Muckstadt, J. A. (1985). Establishing Consistent and Realistic Reorder Intervals in Production-Distribution Systems. *Oper. Res.* **33** 1316–1341.

Queyranne, M. (1985). A Polynomial-Time Submodular Extension to Roundy's 98% Effective Heuristic for Production/Inventory Systems. Working Paper no. 1136, Faculty of Commerce and Business Administration, University of British Columbia, Vancouver.

Rockafellar, R. T. (1970). *Convex Analysis*. Princeton University Press, Princeton, NJ.

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

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

Schwarz, L. B. (1973). A Simple Continuous Review Deterministic One-Warehouse N-Retailer Inventory Problem. *Management Sci.* **19** 555–566.

Tardos, E., Tovey, C. and Trick, M. (1986). Layered Augmented Path Algorithms. *Math. Oper. Res.* **11** 362–370.

Veinott, A. F. (1969). Minimum Concave-Cost Solution of Leontief Substitution Models of Multifacility Inventory Systems. *Oper. Res.* **17** 262–291.

Welsh, D. (1976). *Matroid Theory*. Academic Press, London.

Zheng, Y.-S. (1987). Replenishment Strategies for Production/Distribution Networks with General Joint Setup Costs. Ph.D. Dissertation, Columbia University, New York.

FEDERGRUEN: GRADUATE SCHOOL OF BUSINESS, COLUMBIA UNIVERSITY, NEW YORK, NEW YORK 10027

QUEYRANNE: FACULTY OF COMMERCE AND BUSINESS ADMINISTRATION, UNIVERSITY OF BRITISH COLUMBIA, VANCOUVER, BRITISH COLUMBIA, CANADA V6T 1Z2

ZHENG: THE WHARTON SCHOOL, UNIVERSITY OF PENNSYLVANIA, PHILADELPHIA, PENNSYLVANIA 19104