Available online at www.sciencedirect.com

SCIENCE DIRECT

ELSEVIER

Discrete Applied Mathematics 145 (2004) 22–39

DISCRETE APPLIED MATHEMATICS

www.elsevier.com/locate/dam

# An asymptotic 98.5%-effective lower bound on fixed partition policies for the inventory-routing problem☆

Shoshana Anily[a], Julien Bramel[b]

[a] *Faculty of Management, Tel-Aviv University, Tel-Aviv 69978, Israel*
[b] *406 Uris Hall, Columbia University, New York, NY 10027, USA*

Received 13 December 2000; received in revised form 21 November 2002; accepted 19 September 2003

## Abstract

We consider the Inventory-Routing Problem where $n$ geographically dispersed retailers must be supplied by a central facility. The retailers experience demand for a product at a deterministic rate and incur holding costs for keeping inventory. Distribution is performed by a fleet of capacitated vehicles. The objective is to minimize the average transportation and inventory costs per unit time over the infinite horizon. In this paper, we focus on the set of fixed partition policies. In a fixed partition policy, the retailers are partitioned into disjoint and collectively exhaustive sets. Each set of retailers is served independently of the others and at its optimal replenishment rate. We derive a deterministic ($O(n)$) lower bound on the cost of the optimal fixed partition policy. A probabilistic analysis of the performance of this bound demonstrates that it is asymptotically 98.5%-effective. That is, as the number of retailers increases, the lower bound is very close to the cost of the optimal fixed partition policy.



*Keywords:* Inventory routing problem; Bin packing problem; Asymptotic probabilistic analysis

## 0. Introduction

We consider the Inventory-Routing Problem (IRP), also called the One Warehouse Multi-Retailer Distribution Problem. In this problem, retailers face constant retailer-specific demand rates for a single product. At each retailer a linear holding cost per unit time is incurred for items held in inventory. We assume one warehouse serves as the supply point for the retailers. Initially stationed at the warehouse and available to perform delivery is a fleet of vehicles of limited capacity. The objective is to determine a long-term *integrated* replenishment strategy (i.e., inventory rules, delivery quantities and routing patterns) allowing all retailers to meet their demands while minimizing long-run average system-wide transportation and inventory costs.

Managing such a system, which is common in vendor-managed inventory (VMI) situations, requires a careful coordination of deliveries to balance transportation and inventory charges. The problem is complicated by the fact that delivery costs include the total distance traveled, which requires solving Traveling Salesman Problems (TSP) defined by the retailers assigned to a route. In addition to the computational complexity, even if an optimal policy could be efficiently determined, the policy might be very complex and therefore not easily implemented. Hence there is a need to design and efficiently identify cost-effective policies that are of a simple nature. This has prompted research on this problem that concentrates mainly on a specific policy class called *partitioning policies*. These are characterized by a set of routes

☆ This research was supported by an internal grant from the Columbia Graduate School of Business and NSF CAREER Award DMI-97-02596.

*E-mail addresses:* anily@post.tau.ac.il (S. Anily), jdb8@columbia.edu (J. Bramel).

0166-218X/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.dam.2003.09.005

with the following properties: (1) each route visits a subset of retailers at equidistant time intervals and (2) each route is responsible for replenishing a certain fraction (which is retailer dependent) of the demand of each of the retailers it visits.

The literature distinguishes between two versions of the partitioning problem: the *split* and the *unsplit* demand case, see Anily and Bramel [2]. In the split demand case, a retailer may be served by a number of routes, whereas in the unsplit demand case, each retailer must be served on a single route. I.e., in the unsplit demand case, a route either serves all the demand of a given retailer or it does not serve (visit) it at all. A policy for the unsplit demand case therefore consists of a partition of the retailers into disjoint and collectively exhaustive sets where each set is served on a separate route. A policy for the unsplit demand case is called a *fixed partition policy* (FPP). Clearly, any feasible FPP is also a feasible solution for the split demand case but not vice versa. Even though both versions of the problem are NP-Hard, the split demand case is usually perceived as simpler since it allows for more flexibility in combining deliveries on a route so that the vehicles' capacity is better exploited.

Anily and Federgruen [4,5] and Anily [1] design partitioning policies for the split demand case using *region partitioning schemes*. Specifically, as a first step each retailer is replaced by unit-demand retailers. Then using region partitioning the unit-demand retailers are grouped into clusters that are served at optimal replenishment rates. These simple polynomial-time schemes are shown to construct policies that are asymptotically optimal within this class.

Gallego and Simchi-Levi [11] analyze a particular policy for the unsplit demand case (an FPP) called *direct shipping*. According to this policy each retailer is served *individually* (on a separate route) and at its optimal replenishment rate. Thus no attempt is made to combine shipments of different retailers into a single route. The authors show that this policy, naive as it is, provides a solution whose cost is within 6% of the minimal policy cost as long as the optimal replenishment quantity of *each* retailer (independent of the others) is at least 71% of the vehicle capacity.

In [7], Bramel and Simchi-Levi develop a heuristic to design an effective FPP. Their heuristic is based on formulating the partitioning problem as a location problem, called a *capacitated concentrator location problem*, with conveniently chosen costs and parameters. The solution to the location problem suggests an FPP which is shown, in a series of computational experiments, to be cost-effective.

A number of other approaches and analyses of this problem have appeared, e.g., Herer and Roundy [12] consider power-of-two policies for a version of the problem with uncapacitated vehicles. Vishwanathan and Mathur [16] develop power-of-two policies for a similar problem with multiple products. See [2] for a recent survey of bounds, heuristics and policy classes for the IRP.

Fixed partition policies are of interest for two important reasons: they have simple structure and can be very cost-effective. Their simplicity comes from the following facts: each retailer is served at equidistant intervals, each shipment to the same retailer is of the same size, and only a small number of routes (system-wide) need to be driven. Besides the computational results of [7], the cost-effectiveness of FPPs was shown in a rigorous manner by Chan et al. [8]. They consider the FPP class and show that the best FPP is within 41% of the minimal policy cost asymptotically. Also, as demonstrated in [8], under certain conditions (described below), there exists an FPP that is *asymptotically optimal* with respect to *all* feasible policies.

For a firm implementing an FPP, it is important to know the policy's effectiveness. Thus a lower bound is required. Chan et al. [8] develop a lower bound on the average cost of *any* feasible policy, and Anily and Federgruen [5] derive a lower bound over *all partitioning policies*. Clearly, both lower bounds may also serve as a lower bound on the cost of the optimal FPP. However, these comparisons may not be the most appropriate since they involve comparisons with policies which cannot or would not be implemented even if known. Therefore, it is important to determine the quality of a proposed FPP *within* the FPP class, thus providing a comparison among similar policies (or at least policies of similar complexity, implementation-wise). The lower bound of [8] provides some information about the effectiveness of FPPs within the FPP class. They show that, for certain instances, the lower bound over all feasible policies is asymptotically tight with the best FPP, but generally, the best FPP does not exceed $\sqrt{2} \cdot 100\%$ of the lower bound asymptotically. In addition, this bound is tight (see [6]).

In this paper, we develop a deterministic, linear-time lower bound on the cost of the best FPP, which depends only on the problem's parameters and a few characteristics of the joint probability distribution of the retailer's locations and demand rates. This bound is proven in Section 2.1 to be asymptotically 98.5%-effective. This means that as the number of retailers increases, the lower bound becomes at least 98.5% of the cost of the optimal FPP. The convergence rate of the lower bound to the optimal FPP solution is highly dependent on the joint p.d.f. of the retailer's locations and demand rates. For practical purposes, it will be interesting to investigate in future research the convergence rate of the proposed lower bound for various forms of joint p.d.f. Such results are important for the assessment of the suitability of the proposed lower bound for a given problem size.

The lower bound that we propose is a sort of *cost allocation* scheme: a cost is allocated to each retailer and for any set of retailers (that can be feasibly served together), the sum of the cost allocations over the set is a lower bound on the long-term average cost of serving the set in any FPP. This lower bound can then be used to assess the effectiveness of any FPP. As explained in the sequel, the lower bound value depends on a certain value characteristic of the demand distribution of the retailers. This value is not always available or easily computable. Still, as it turns out in Anily and Bramel [3], the existence proof of this lower bound and its proven asymptotic optimality gap are extremely valuable in the probabilistic analysis of the region partitioning scheme proposed in [3] for designing a good FPP. This is most notable, in view of the fact that the construction of an FPP in *polynomial-time* with an ex ante bound on its optimality gap has eluded researchers for many years. The only exception is the result on direct shipping of [11] which only holds in certain cases. The main reason for the relatively slow progress in this direction is due to the lack of a good lower bound on the cost of policies in the class. The lower bound presented here has made it possible to revisit the issue. Indeed, in Anily and Bramel [3] the authors develop a polynomial-time region partitioning procedure that constructs an FPP whose cost is asymptotically within 1.5% of the lower bound derived in this paper and thus within 1.5% of the best FPP.

In the next section we describe the notation and the probabilistic assumptions used to derive the asymptotic effectiveness. In Section 2, we develop the lower bound and in Section 2.1 we prove that it is asymptotically 98.5%-effective. In Section 3, we give some concluding remarks.

## 1. Preliminaries

Let $N=\{1,2,\dots,n\}$ denote the set of retailers. We denote by $w_i$ the *demand rate* (in units of product per unit time) at retailer $i\in N$. An unlimited number of vehicles of limited capacity $Q$ are available to perform the distribution. These vehicles are initially stationed at the *depot*, which could be a warehouse, distribution center or manufacturing facility. The cost of a delivery (loading the vehicle, traveling to a number of retailers, unloading at each retailer and returning to the depot) is assumed to consist of a fixed cost $c\geqslant 0$ plus a term that is proportional to the distance the vehicle travels. Without loss of generality, we assume the cost per unit distance is 1.

Each retailer incurs a holding cost $h$ per unit of product held in inventory per unit of time, independent of the location. To avoid situations where retailers are served at excessive frequencies and for analytical tractability, our model assumes (as in [1,4,5,8]) that there is a constraint on the maximum delivery *frequency* to each retailer. I.e., the inventory of a retailer may not be replenished more than $f$ times per time unit. More specifically, within any time interval of length $t>0$ units, a retailer cannot get more than $t\cdot f$ deliveries.

A fixed partition policy is specified by a partition of $N$ along with *replenishment rates* for each set in the partition. A partition of $N$ is any collection of non-empty sets $\{S_1,S_2,\dots,S_m\}$ such that $S_k\cap S_\ell=\emptyset$, for all $1\leqslant k,\ell\leqslant m$ $(k\neq\ell)$, and $\bigcup_{k=1}^m S_k=N$. Note a set $S\subseteq N$ can be *feasibly served* in an FPP if $w(S)\stackrel{\text{def}}{=}\sum_{i\in S}w_i\leqslant Qf$. That is, under our assumptions a set of retailers with total demand rate greater than $Qf$ cannot be feasibly served on a single route in an FPP. To see this note that the maximum delivery rate to a retailer has a vehicle deliver a full load ($Q$ units) every $1/f$ units of time, i.e., a rate of $Qf$ units per unit time. Given a partition $\{S_1,S_2,\dots,S_m\}$ we call it *feasible* if $w(S_k)\leqslant Qf$ for each $k=1,2,\dots,m$.

Given a set $S\subseteq N$, let $L^*(S)$ denote the length of an optimal traveling salesman tour through the depot and the retailers of $S$. For any set $S\subseteq N$ with $w(S)\leqslant Qf$, we define

$$z(S)\stackrel{\text{def}}{=}\min\left\{\frac{L^*(S)+c}{t}+\frac{1}{2}htw(S):\frac{1}{f}\leqslant t\leqslant\frac{Q}{w(S)}\right\}.$$

The function $z(S)$ gives the cost per unit time of serving $S$ in an FPP. The value of $t$ providing the minimum in $z(S)$ is the *optimal replenishment interval* of $S$. It follows that the cost of the FPP defined by the feasible partition $\{S_1,S_2,\dots,S_m\}$ is given by $\sum_{i=1}^m z(S_i)$. Since for fixed $n$ the number of feasible partitions is finite and the number of different sequences with which the retailers in each set can be visited is finite, the set of FPPs that need be considered for optimality is also finite (assuming that each set is ordered at its optimal replenishment rate). We therefore let $Z_n^{\text{FP}}$ denote the cost of an optimal fixed partition policy for a problem with $n$ retailers.

We propose a quite standard probabilistic model (the same one as in [8] and very similar to the one in [1,4,5]). We assume the retailers and the depot are located on the Euclidean plane. Without loss of generality, we place the depot at the origin. We assume the locations and demand rates of retailers are drawn from a joint distribution $\Phi$. We denote by $\psi$ the marginal distribution of retailer demand rates, and without loss of generality assume the range of feasible rates is $(0,Qf]$. Let $\beta$ denote the *average* demand rate (as a proportion of $Qf$), thus $\beta\in(0,1]$ and $\mathsf{E}[w]=\beta Qf$. Denote by $\mu$ the marginal distribution of retailer locations, and assume $\mu$ has compact support $\mathscr{A}\subset\mathbb{R}^2$. We denote by $\|x\|$ the Euclidean distance between $x\in\mathbb{R}^2$ and the depot. We denote by $d_i$ the distance retailer $i\in N$ is from the depot, i.e., if retailer $i$ is located at $x_i\in\mathscr{A}$, then $d_i\stackrel{\text{def}}{=}\|x_i\|$.

In [5], Anily and Federgruen show that an asymptotically optimal policy (within the class of region partitioning policies) for the IRP consists of partitioning the unit demand rate retailers into the minimum possible number of sets (routes).

Thus, we can expect that a good fixed partition policy is one that consists of a small number of feasible sets. Therefore the well-known Bin Packing Problem (BPP) will play a crucial role in our analysis.

The BPP is defined by a set of *item sizes* and bins of fixed capacity. The problem is to assign all the items into the minimum number of bins in such a way that in no bin is the capacity exceeded. Excellent surveys of this problem appear in [9,10]. In our context, we are concerned with packing the *demand rates* of a subset of retailers into the minimum number of bins of capacity $Qf$. I.e., a set of retailers $S$ makes up a feasible bin (route) if and only if $w(S) \leqslant Qf$.

An important bin-packing result concerns the number of bins required as the number of items increases. Let $b_m^*$ be the number of bins used in an optimal solution to the BPP defined on $m$ items of size $\{w_1, w_2, \ldots, w_m\}$. Using Kingman's theory of subadditive processes [13] (see also [15]), it is possible to show that there exists a constant $\gamma \in (0, 1]$, dependent only on $\psi$ (the distribution of demand sizes) and the bin capacity, such that $\gamma \stackrel{\text{def}}{=} \lim_{m\to\infty} b_m^*/m$ almost surely. This $\gamma$ is called the *bin-packing constant associated with* $\psi$, and note that $\beta \leqslant \gamma < 2\beta$. We define $\alpha = \gamma/\beta$ as the *packing efficiency* of $\psi$, and note that $\alpha \in [1, 2)$. We say a distribution with $\alpha = 1$ allows *perfect packing*, i.e., as the number of items increases the amount of wasted space in an optimal packing becomes a negligible fraction of the total space. On the other hand, a distribution with $\alpha \approx 2$ has almost all bins in an optimal packing only slightly more than half-full.

To assess the strength of our lower bound, we use the following measure:

**Definition 1.1.** A lower bound $\underline{Z}_n$ on $Z_n$ is *asymptotically $\eta$-effective* if

$$\underline{\lim}_{n\to\infty} \frac{1}{n} \underline{Z}_n \geqslant \eta \cdot \overline{\lim}_{n\to\infty} \frac{1}{n} Z_n \quad \text{(a.s.)}.$$

In particular, if $\underline{Z}_n$ is an asymptotically $\eta$-effective lower bound on $Z_n^{\text{FP}}$ then as the number of retailers increases, the cost of the lower bound approaches at least $\eta$ times the cost of the best FPP. A 100%-effective lower bound is an *asymptotically tight* lower bound.

As we have noted, Chan et al. [8] analyze FPPs using an identical probabilistic model to ours. Their main result is that fixed partition policies are asymptotically at most $\sqrt{\alpha}$ times the minimal policy cost (among *all* feasible policies). This is done as follows. They present the following lower bound:

$$B_n = \sum_{i=1}^{n} w_i \left( \frac{2d_i + c}{Q} + \frac{h}{2f} \right)$$

and show that $B_n$ is a lower bound on *any* feasible policy. They also construct an FPP (using an exponential-time grid partitioning scheme) with cost $Z_n^C \geqslant Z_n^{\text{FP}}$ that satisfies

$$\overline{\lim}_{n\to\infty} \frac{1}{n} Z_n^C \leqslant \sqrt{\alpha} \cdot \underline{\lim}_{n\to\infty} \frac{1}{n} B_n \quad \text{(a.s.)}.$$

This result implies that the best FPP is asymptotically within a factor of $\sqrt{\alpha}$ of the best feasible policy. Since $\alpha < 2$, this result implies that fixed partition policies are generally effective, in particular, no worse than $\sqrt{2}$ times an optimal policy's cost (asymptotically). Their effectiveness in fact depends on $\alpha$, which is a function of specific characteristics of the distribution of demand rates, the vehicle capacity and the frequency constraint. An interesting special case is $\alpha = 1$, i.e., when the demand rate distribution allows perfect packing. In this case, the results of [8] demonstrate that an optimal FPP is *asymptotically optimal* with respect to all feasible policies.

As can be noted from [8], the lower bound expression $B_n$ does not incorporate any measure related to the packing constant $\gamma$ or the packing efficiency $\alpha$. This is a result of the fact that $B_n$ is a lower bound on *all* feasible policies. The objective of this paper is to obtain a better lower bound for fixed partition policies by taking into account the inherent bin-packing problem and thus relating the lower bound to $\gamma$ (or $\alpha$). Indeed, we obtain a new lower bound with respect to fixed partition policies which is at least as large as the lower bound proposed by Chan et al. [8]. Our lower bound depends, in addition, on the bin-packing constant $\gamma$ associated with $\psi$. In the special case that $\psi$ allows perfect packing, i.e., $\gamma = \beta$, our lower bound coincides with the lower bound $B_n$ derived in [8]. Coffman and Lueker [10] describe some distributions for which the value of $\gamma$ is known. As we show in the sequel, even if the bin-packing constant associated with $\psi$ in an IRP instance is unknown, our technique usually allows to obtain a better (higher) lower bound than $B_n$ on the best FPP. For that sake we need to use another important concept from the bin packing literature called *dual feasible functions*. These have played a key role in a number of important bin-packing results; details can be found in Coffman and Lueker [10].

**Definition 1.2.** A function $u : [0, Qf] \to [0, 1]$ is *dual feasible for* $\psi$ if for all sets of items $S$ generated from $\psi$ with $w(S) \leqslant Qf$ we have $\sum_{i\in S} u(w_i) \leqslant 1$.

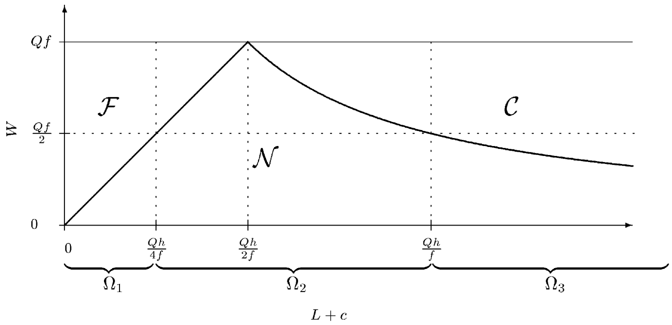


Fig. 1. A diagram of $z(L,W)$.

When $\psi$ is clear from the context, we will call $u$ simply *dual feasible*. For example, the function $u(w)=w/Qf$ for $w\in[0,Qf]$ is always dual feasible. We call this function the *trivial* dual feasible function. For integers $k\geqslant 2$, the function $u(w)=(1/(k-1))\lceil wk/Qf-1\rceil$ is also dual feasible, where $\lceil x\rceil$ is the smallest integer greater than or equal to $x$. A number of other examples are presented in [14]. Dual feasible functions have a number of useful properties some of which we present next.

**Property 1.3.** 1. *If $u$ is dual feasible for $\psi$, then $\gamma\geqslant\int u(w)\,d\psi(w)\equiv \mathsf{E}[u(w)]$.*
2. *There exists a dual feasible function $u^*$ such that $\gamma=\int u^*(w)\,d\psi(w)=\mathsf{E}[u^*(w)]$.*

**Proof.** 1. follows since $\sum_{i=1}^n u(w_i)$ is a lower bound on the number of bins required to pack $\{w_1,w_2,\ldots,w_n\}$. Dividing by $n$, taking the limit and using the strong law of large numbers proves the result. For 2, see Theorem 5.8 of [10], or [15]. □

## 2. A lower bound on fixed partition policies

Given a set $S$ with $W\stackrel{\text{def}}{=}w(S)\leqslant Qf$ and a route through $S$ and the depot of length $L$, we define the function:

$$z(L,W)\stackrel{\text{def}}{=}\min\left\{\frac{L+c}{t}+\frac{1}{2}htW:\frac{1}{f}\leqslant t\leqslant\frac{Q}{W}\right\}.$$

Let $t^*$ denote the value of $t$ achieving the minimum in $z(L,W)$. The function $z(L,W)$ gives the cost per unit time of dispatching one vehicle every $t^*\geqslant 1/f$ units of time to serve $S$. The vehicle leaves the depot carrying $Wt^*\leqslant Q$ units of product. If $t^*=1/f$ we say the frequency constraint is *active*, while if $t^*=Q/W$ we say the capacity constraint is *active*.

To describe our lower bound, we begin by analyzing $z(L,W)$.

**Property 2.1.** *The following can be easily shown*:

1. *For fixed $W$, $z(L,W)$ is strictly increasing in $L\geqslant 0$.*
2. *For fixed $L$ and $0\leqslant W\leqslant Qf$, $z(L,W)$ is strictly increasing in $W$ and, in particular, $z(L,W)\leqslant z(L,Qf)=(L+c)f+hQ/2$.*
3. *For fixed $L\geqslant 0$, $z(L,W)$ is concave in $W$. In addition, the derivative of $z(L,W)$ with respect to $W$ is continuous.*
4. *For any $L\geqslant 0$, $W>0$ and $\varepsilon\geqslant 0$: $z(L+\varepsilon,W)\leqslant z(L,W)+f\varepsilon$.*

To help with the exposition that follows, consider Fig. 1 below.

In Fig. 1, we consider the function $z(L,W)$ by plotting $L+c$ on the $x$-axis and $W$ on the $y$-axis. In this figure, we distinguish three important regions labeled $\mathscr{F}$, $\mathscr{C}$ and $\mathscr{N}$ depending on the value of $t^*$. We consider each of these regions in turn.

- *Region $\mathscr{F}$*: In this region, $t^*=1/f$, i.e., the *frequency* constraint is active in $z(L,W)$. Thus $z(L,W)=(L+c)f+hW/(2f)$. Note that in this region we have $W\geqslant 2(L+c)f^2/h$.
- *Region $\mathscr{C}$*: In this region, $t^*=Q/W$, i.e., the *capacity* constraint is active in $z(L,W)$. Thus $z(L,W)=(L+c)W/Q+hQ/2$. Note that in this region we have $W\geqslant Q^2h/(2(L+c))$.

- *Region* $\mathcal{N}$: In this region, $t^* = \sqrt{2(L+c)/hW}$, i.e., *neither* frequency nor capacity constraints are active in $z(L,W)$. Thus $z(L,W) = \sqrt{2hW(L+c)}$.

As can be observed from Fig. 1, it will be convenient to partition the $(L+c)$-axis into the sets $\Omega_1$, $\Omega_2$ and $\Omega_3$, where

$$\Omega_1 = \left[0, \frac{Qh}{4f}\right), \quad \Omega_2 = \left[\frac{Qh}{4f}, \frac{Qh}{f}\right) \quad \text{and} \quad \Omega_3 = \left[\frac{Qh}{f}, +\infty\right).$$

These sets naturally define three *rings* around the depot.

We start by constructing a lower bound on $z(L,W)$ *for fixed* $L$ and for $W > Qf/2$. Specifically, for fixed $L$, we seek a function $\underline{z}(L,W)$ that is linear in $W$ and satisfies $\underline{z}(L,W) \leqslant z(L,W)$ for all $W \in (Qf/2, Qf]$. Below we describe how this is used in the lower bound. We consider three separate intervals for $L$:

- $L+c \in \Omega_1$. Since $z(L,W)$ is linear in $W$, for $W \in (Qf/2, Qf]$, we simply set $\underline{z}(L,W) = z(L,W) = (L+c)f + hW/(2f)$.
- $L+c \in \Omega_3$. Since $z(L,W)$ is linear in $W$, for $W \in (Qf/2, Qf]$, we simply set $\underline{z}(L,W) = z(L,W) = (L+c)W/Q + hQ/2$.
- $L+c \in \Omega_2$. Here we let $\underline{z}(L,W)$ be the line that connects the points of $z(L,W)$ at $W = Qf/2$ and $W = Qf$. Since $z(L,W)$ is concave in $W$ (Property 2.1, part 3), we have $\underline{z}(L,W) \leqslant z(L,W)$ for all $W \in [Qf/2, Qf]$. Since $z(L, Qf/2) = \sqrt{(L+c)hQf}$ and $z(L, Qf) = (L+c)f + hQ/2$, the line has slope:

$$m_L = \frac{2}{Qf}((L+c)f + hQ/2 - \sqrt{(L+c)hQf}).$$

Let $q(L)$ denote the intercept of this line, where

$$q(L) \stackrel{\text{def}}{=} (L+c)f + hQ/2 - Qfm_L = 2\sqrt{(L+c)hQf} - [(L+c)f + hQ/2]. \tag{1}$$

We define

$$p(L) \stackrel{\text{def}}{=} Qfm_L = 2(L+c)f + hQ - 2\sqrt{(L+c)hQf} \tag{2}$$

and therefore $\underline{z}(L,W) = p(L)W/Qf + q(L)$.

In each of the three regions, $\underline{z}(L,W)$ is linear in $W$ and provides a lower bound on $z(L,W)$, as long as $W > Qf/2$. For the next step in our lower bound, consider a set $S$ of retailers with $w(S) = W \in (Qf/2, Qf]$ and $L = L^*(S)$. We assume $L$ is fixed and $\underline{z}(L,W) = aW + b$ (for some $a$ and $b$ dependent on $L$). From the derivation above it is clear that $\underline{z}(L,W)$ is a lower bound on the cost (per unit time) of serving $S$ in an FPP. This cost can be interpreted as a charge of $a$ per unit of demand rate in $S$ and a fixed charge of $b$ for the set itself. We *allocate* a portion of this lower bound to each retailer in $S$, in such a way that the sum of the allocations over the retailers in $S$ does not exceed $\underline{z}(L,W)$. This is where the concept of dual feasible functions is useful. Let $u$ be dual feasible for $\psi$. We charge $aw_i + bu(w_i)$ to retailer $i \in S$. Since $u$ is dual feasible, summing these charges over all retailers of $S$, the total charge is at most $aW + b = \underline{z}(L,W)$. Thus the allocation provides a lower bound on $\underline{z}(L,W)$ and therefore also a lower bound on $z(L,W)$. We now have a charge per retailer which we can use to bound the cost of any FPP. Unfortunately, it is not quite this simple since $a$ and $b$ depend on $L$, the length of the optimal traveling salesman tour through the retailers of $S$. We will therefore use a lower bound on $L$ in the cost allocation.

The cost allocation is formally done as follows. For any $d \geqslant 0$, $w \in [0, Qf]$ and dual feasible function $u$, define the function:

$$g_u(d,w) = \begin{cases} hw/(2f) + u(w)(2d+c)f & \text{if } 2d+c \in \Omega_1, \\ p(2d)w/Qf + u(w)q(2d) & \text{if } 2d+c \in \Omega_2, \\ (2d+c)w/Q + u(w)hQ/2 & \text{if } 2d+c \in \Omega_3. \end{cases}$$

We now show that the sum of the evaluations of the function $g_u$ over each retailer is a lower bound on the optimal FPP. To prove this we show a stronger result, namely that for any set $S \subseteq N$ with $w(S) \leqslant Qf$ (not just $w(S) > Qf/2$), the cost per unit time of serving the set is no less than the sum of the evaluations of $g$ over the retailers of $S$. The proof is given in the appendix.

**Lemma 2.2.** *Let $u$ be dual feasible for $\psi$. For any $S \subseteq N$ with $w(S) \leqslant Qf$, we have*

$$\sum_{i \in S} g_u(d_i, w_i) \leqslant z(S).$$

The following lower bound theorem is a direct result of Lemma 2.2.

**Theorem 2.3.** *For any $u$ dual feasible for $\psi$, we have*

$$\sum_{i=1}^{n} g_u(d_i, w_i) \leqslant Z_n^{\mathrm{FP}}.$$

We now derive the probabilistic (asymptotic) version of this lower bound. The proof follows directly from Theorem 2.3 and the strong law of large numbers.

**Theorem 2.4.** *Assume $n$ retailer locations and demand rates are drawn from a joint distribution $\Phi(x,w)$, where $x \in \mathscr{A}$ and $w \in (0, Qf]$. Assume $u$ is dual feasible for $\psi \stackrel{\text{def}}{=} \int \Phi(x,\cdot)\,\mathrm{d}x$. Then*

$$\varliminf_{n\to\infty} \frac{1}{n} Z_n^{\mathrm{FP}} \geqslant \mathsf{E}[g_u] \stackrel{\text{def}}{=} \int g_u(\|x\|, w)\,\mathrm{d}\Phi(x,w) \quad \text{(a.s.)}.$$

Also:

**Corollary 2.5.** *Let $\mathscr{U}$ be the set of functions that are dual feasible for $\psi$, then:*

$$\varliminf_{n\to\infty} \frac{1}{n} Z_n^{\mathrm{FP}} \geqslant \underline{Z} \stackrel{\text{def}}{=} \sup_{u\in\mathscr{U}} \mathsf{E}[g_u] \quad \text{(a.s.)}.$$

Our next theorem characterizes the value of the lower bound $\underline{Z}$ under the probabilistic assumptions described in Section 1 and under the additional assumption that retailer locations and demand rates are independent. For this purpose, for $\ell \in \{1,2,3\}$, define $A_\ell \stackrel{\text{def}}{=} \{x \in \mathscr{A} : 2\|x\| + c \in \Omega_\ell\}$ and let $\bar{d}_\ell$ denote the average distance to the depot of retailers in ring $A_\ell$. The following result will be needed to prove the asymptotic effectiveness.

**Theorem 2.6.** *If retailer locations and demand rates are* independent, *then*

$$\begin{aligned}\varliminf_{n\to\infty} \frac{1}{n} Z_n^{\mathrm{FP}} &\geqslant \underline{Z}\\ &= \mu(A_1)[\beta hQ/2 + \gamma(2\bar{d}_1 + c)f]\\ &\quad + \mu(A_2)[(2\beta-\gamma)((2\bar{d}_2 + c)f + hQ/2) + 2\sqrt{hQf}(\gamma-\beta)\cdot \mathsf{E}[\sqrt{2d+c} \mid 2d + c \in \Omega_2]]\\ &\quad + \mu(A_3)[(2\bar{d}_3 + c)\beta f + \gamma hQ/2] \quad \text{(a.s.)}.\end{aligned} \tag{3}$$

**Proof.** Given $\psi$, Property 1.3, part 2 shows that there exists a dual feasible function $u^*$ with $\gamma = \int u^*(w)\,\mathrm{d}\psi(w)$. Use this particular $u^*$ to define $g_u$, i.e., $g_u = g_{u^*}$. Now consider separately the term associated with each ring. Each of the terms is derived from the definition of $g_u$, the strong law of large numbers and the definitions of $\beta$ and $\gamma$. $\square$

We now present a few examples to illustrate how the bound can be used. Given an instance with $n$ retailers, determining the largest possible lower bound (in Theorem 2.3) entails choosing the dual feasible function $u$ maximizing the bound. We refer the reader to [10,14] for additional examples of dual feasible functions to those used here. Generally, it is not possible to say which function $u$ will give the largest lower bound, however we can say that asymptotically the best dual feasible function to use in $g$ is $u^*$ from Property 1.3, part 2.

One special case is to consider using the trivial dual feasible function in $g$, i.e., $u(w) = w/Qf$ for $w \in (0, Qf]$. We get

$$g_u(d,w) = w\left(\frac{2d+c}{Q} + \frac{h}{2f}\right)$$

for all $d \geqslant 0$ and $w \in (0, Qf]$. In this case, the lower bound is exactly the lower bound derived by Chan et al. [8]. Interestingly, this lower bound is known (see [8]) to be a lower bound on all feasible policies, not just fixed partition policies. In general, this will not yield an asymptotically tight lower bound on fixed partition policies. In fact, the only case where the lower bound is asymptotically tight is when the trivial $u$ is the best possible dual feasible function ($u^*$ in Property 1.3, part 2). This is true if and only if $\psi$ allows perfect packing.

Fix $k \geqslant 2$ and consider the case where all retailer demand rates are greater than $Qf/k$, i.e., $\psi$ is zero over $[0, Qf/k]$. Then we can choose the dual feasible function $u(w) = (1/(k-1))\lceil kw/Qf - 1\rceil$ for $w > 0$. For example, if $\psi$ is the uniform distribution over $(Qf/k, Qf]$, then $\gamma = (k+1)/2k$. According to Property 1.3, part 2, this is asymptotically the best possible choice for $u$.

We now show that the lower bound $\underline{Z}$ is asymptotically 98.5%-effective within the class of fixed partition policies. This is done by constructing an FPP using grid partitioning and then showing that, asymptotically, the lower bound is at least 98.5% of the cost of this fixed partition policy.

From here on we assume that the retailer demand rates and retailer locations are *independent* of each other. We assume the retailer locations are generated according to a distribution $\mu$ with compact support $\mathscr{A} \subset \mathbb{R}^2$, and the retailer demand rates are generated according to a distribution $\psi$ on $(0, Qf]$.

The fixed partition policy we construct is called *Grid Partition* (GP) and its cost is denoted $Z_n^{\mathrm{GP}}$. A similar solution was constructed by Chan et al. [8]. We remark that this FPP cannot be constructed in polynomial-time since it requires solving a number of bin-packing problems to optimality. The construction is used solely to prove that our lower bound is asymptotically effective. Identifying, in polynomial-time, a provably effective FPP is a separate challenge which is addressed in [3].

The upper bound construction is based on a parameter $\theta \in [1,2)$. Later on, this number will be fixed but for now we will write it in this more general way. We partition $\mathscr{A}$ in a manner different from in the lower bound, and based on $\theta$. Note that in the notation below we often omit the dependence on $\theta$. Define

$$\Omega_1' = \left[0, \frac{Qh}{2f\theta}\right), \quad \Omega_2' = \left[\frac{Qh}{2f\theta}, \frac{Qh\theta}{2f}\right) \quad \text{and} \quad \Omega_3' = \left[\frac{Qh\theta}{2f}, +\infty\right).$$

Since $\theta \in [1,2)$ we see that $\Omega_1 \subset \Omega_1'$, $\Omega_2' \subset \Omega_2$, and $\Omega_3 \subset \Omega_3'$. These three sets naturally define three rings around the depot. Let $A_\ell' = \{x \in \mathscr{A} : 2\|x\| + c \in \Omega_\ell'\}$, for $\ell \in \{1,2,3\}$, denote these rings.

Select an $\varepsilon > 0$. The GP($\varepsilon$) policy is constructed as follows. Superimpose on $\mathscr{A}$ an infinite grid of side $\varepsilon/\sqrt{2}$ and let $R_\ell'$ be an index set of the square[1] subregions in $A_\ell'$ that have strictly positive measure under $\mu$. Let $\underline{\delta}_j$ denote the minimal distance between subregion $j \in R_\ell'$ and the depot. We consider each square subregion in turn. In each we determine an optimal packing of the retailer demands into bins of capacity $Qf$. Let $\{S_{\ell jk}\}$ denote the retailers in the $k$th bin of subregion $j \in R_\ell'$ (in ring $A_\ell'$). Then

$$Z_n^{\mathrm{FP}} \leqslant Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \sum_{\ell=1}^{3} \sum_{j \in R_\ell'} \sum_k z(L^*(S_{\ell jk}), w(S_{\ell jk})).$$

It is easy to see that if a set of retailers $S$ is fully contained in subregion $j$ then $L^*(S) \leqslant 2\underline{\delta}_j + (|S|+1)\varepsilon$. By Property 2.1, part 1:

$$Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \sum_{\ell=1}^{3} \sum_{j \in R_\ell'} \sum_k z(2\underline{\delta}_j + (|S_{\ell jk}|+1)\varepsilon, w(S_{\ell jk})).$$

Hence by Property 2.1, part 4:

$$Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \sum_{\ell=1}^{3} \sum_{j \in R_\ell'} \sum_k z(2\underline{\delta}_j, w(S_{\ell jk})) + 2nf\varepsilon. \tag{4}$$

In the next step, we construct an upper bound on $z(L,W)$. This upper bound is linear in $W$, for fixed $L$.

**Lemma 2.7.** *Fix $\theta \in [1,2)$. For any $L \geqslant 0$ and $W \in [0, Qf]$:*

$$z(L,W) \leqslant \bar{z}_\theta(L,W) \stackrel{\mathrm{def}}{=} \begin{cases} hW/(2f) + (L+c)f & \text{if } L+c \in \Omega_1', \\ W\sqrt{(L+c)h\theta/2Qf} + \sqrt{(L+c)hQf/2\theta} & \text{if } L+c \in \Omega_2', \\ (L+c)W/Q + hQ/2 & \text{if } L+c \in \Omega_3'. \end{cases}$$

**Proof.** The upper bound $\bar{z}_\theta(L,W)$ is linear in $W$ for fixed $L$. In $A_1 \subseteq A_1'$ and $A_3 \subseteq A_3'$, the upper bound is easy to see. That is, if $L+c \in \Omega_1'$ then from Fig. 1, either $(L+c, W) \in \mathscr{F}$ or $(L+c, W) \in \mathscr{N}$ and in both cases we have $z(L,W) \leqslant hW/(2f) + (L+c)f$. Similarly, if $L+c \in \Omega_3'$ then from Fig. 1, either $(L+c, W) \in \mathscr{C}$ or $(L+c, W) \in \mathscr{N}$ and in both cases we have $z(L,W) \leqslant (L+c)W/Q + hQ/2$.

[1] Note we will use the term "square" for simplicity, even though some of these subregions may not be square since they are intersections of squares with the rings or with the border of $\mathscr{A}$.

Region $A_2$ (which intersects $A'_1$, $A'_2$ and $A'_3$) requires a more careful analysis. Assume $L+c \in \Omega'_2$. In this region, $\bar{z}_\theta(L,W)$ simply represents the line that is tangent to $z(L,W)$ at the point $W = Qf/\theta$. Since for fixed $L$ the function $z(L,W)$ is concave in $W$, the tangent line is always above the function $z(L,W)$. To formally prove the inequality, we consider all three cases: $(L+c,W) \in \mathscr{F}$, $(L+c,W) \in \mathscr{N}$ and $(L+c,W) \in \mathscr{C}$.

*Case* 1: $(L+c,W) \in \mathscr{F}$. Both $z(L,W) = (L+c)f + hW/(2f)$ and the claimed upper bound are linear in $W$, hence we compare slopes. Since in $\Omega'_2$ the slope of the claimed upper bound is no less than $h/(2f)$, we can make $W$ as small as possible, i.e., $W = 2f^2(L+c)/h$. Then $z(L,W) = 2(L+c)f$. The claimed upper bound is of the form $U\sqrt{\theta} + V/\sqrt{\theta}$ where

$$U = \frac{2f^2(L+c)}{h}\sqrt{\frac{(L+c)h}{2Qf}} \quad \text{and} \quad V = \sqrt{(L+c)hQf/2}. \tag{5}$$

Note that $\sqrt{UV} = (L+c)f$. If $\theta$ were variable, the minimum possible value of the claimed upper bound would be obtained at $\theta = V/U$. In that case, the claimed upper bound is $2\sqrt{UV} = 2(L+c)f$ and the result follows.

*Case* 2: $(L+c,W) \in \mathscr{N}$. We again assume $\theta$ is variable and set $\theta$ to make the claimed upper bound as small as possible (without affecting the value of $z(L,W)$). The best $\theta$ to choose is $\theta = Qf/W$ and then the claimed upper bound coincides with $z(L,W) = \sqrt{2(L+c)hW}$. The bound follows directly.

*Case* 3: $(L+c,W) \in \mathscr{C}$. Then $z(L,W) = (L+c)W/Q + hQ/2$. Again, both the claimed upper bound and $z(L,W)$ are linear in $W$, so we compare slopes. In $\Omega'_2$ the slope of the claimed upper bound is no less than $(L+c)/Q$, and hence we can make $W$ as small as possible, i.e., we set $W = hQ^2/(2(L+c))$. We then have $z(L,W) = hQ$ and the claimed upper bound is of the form $U'\sqrt{\theta} + V/\sqrt{\theta}$ where

$$U' = \frac{hQ^2}{2(L+c)}\sqrt{\frac{(L+c)h}{2Qf}}$$

and $V$ is as defined in (5). Note that $\sqrt{U'V} = hQ/2$. Therefore, we can set $\theta = V/U'$ to determine the minimum possible value of the claimed upper bound. In that case, the claimed upper bound is $2\sqrt{U'V} = hQ$ proving the result. □

Using this upper bound in (4) gives

$$Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \sum_{\ell=1}^{3} \sum_{j \in R'_\ell} \sum_k \bar{z}_\theta(2\underline{\delta}_j, w(S_{\ell jk})) + 2nf\varepsilon. \tag{6}$$

We now investigate the asymptotic cost of the FPP we have constructed. For this purpose, let $\bar{d}'_\ell$ denote the average distance to the depot for retailers in the ring $A'_\ell$, for $\ell \in \{1,2,3\}$.

**Lemma 2.8.**

$$\begin{aligned}\overline{\lim_{n\to\infty}} \frac{1}{n} Z_n^{\mathrm{GP}(\varepsilon)} &\leqslant \mu(A'_1)[(2\bar{d}'_1 + c)\gamma f + \beta hQ/2] + \mu(A'_2)\sqrt{2hQf\beta\gamma} \cdot \mathsf{E}(\sqrt{2d+c} \mid 2d+c \in \Omega'_2) \\ &\quad + \mu(A'_3)[(2\bar{d}'_3 + c)\beta f + \gamma hQ/2] + 2\varepsilon f \qquad \text{(a.s.)}.\end{aligned} \tag{7}$$

**Proof.** Below we use the upper bounding function $\bar{z}_\theta(L,W)$ defined in Lemma 2.7 with $\theta = \alpha$.

For $\ell \in \{1,2,3\}$, let $N_\ell$ denote the set of retailers in ring $A'_\ell$, and let $n_\ell = |N_\ell|$.

For $\ell \in \{1,2,3\}$, define

$$Z^{(\ell)} \stackrel{\text{def}}{=} \sum_{j \in R'_\ell} \sum_k \bar{z}_\alpha(2\underline{\delta}_j, w(S_{\ell jk})).$$

Thus, (6) can be rewritten:

$$Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \sum_{\ell=1}^{3} Z^{(\ell)} + 2nf\varepsilon.$$

Now dividing by $n$ and taking the limit results in (almost surely):

$$\begin{aligned}
\overline{\lim_{n\to\infty}}\ \frac{1}{n} Z_n^{\mathrm{GP}(\varepsilon)} &\leqslant \sum_{\ell=1}^{3} \overline{\lim_{n\to\infty}}\ \frac{1}{n} Z^{(\ell)} + 2\varepsilon f \\
&= \sum_{\ell=1}^{3} \overline{\lim_{n\to\infty}}\ \frac{n_\ell}{n}\frac{Z^{(\ell)}}{n_\ell} + 2\varepsilon f \\
&\leqslant \sum_{\ell=1}^{3} \left(\overline{\lim_{n\to\infty}}\ \frac{n_\ell}{n}\right)\left(\overline{\lim_{n\to\infty}}\ \frac{Z^{(\ell)}}{n_\ell}\right) + 2\varepsilon f \\
&= \sum_{\ell=1}^{3} \mu(A'_\ell)\ \overline{\lim_{n_\ell\to\infty}}\ \frac{Z^{(\ell)}}{n_\ell} + 2\varepsilon f \qquad (\text{since } n\to\infty \Rightarrow n_\ell\to\infty).
\end{aligned} \tag{8}$$

Recall $R'_\ell$ is the index set of the square subregions in ring $A'_\ell$, for $\ell\in\{1,2,3\}$. We use the term "square" even though the subregions on the boundaries may not be completely square. Also, we omit subregions that have no density (holes), i.e., subregions $S\subset\mathscr{A}$ for which $\mu(S)=0$. For the $j$th (square) subregion of $A'_\ell$, for $\ell\in\{1,2,3\}$, let $N_{\ell j}$ denote the set of retailers in the subregion, $n_{\ell j}=|N_{\ell j}|$ and $b^*_{\ell j}$ denote the number of bins created by GP($\varepsilon$) in the subregion. Note that for $\ell\in\{1,2,3\}$, $j\in R'_\ell$ implies that $2\underline{\delta}_j+c\in\Omega'_\ell$. We consider the three rings $A'_1$, $A'_2$ and $A'_3$ in turn.

In the analysis, each of the rings has in common a term of a specific form and therefore we analyze it here separately. Consider $\ell\in\{1,2,3\}$ and $\{t_j : j\in R'_\ell\}$ (arbitrary) then (almost surely):

$$\begin{aligned}
\overline{\lim_{n_\ell\to\infty}}\ \frac{1}{n_\ell}\sum_{j\in R'_\ell} t_j b^*_{\ell j} &\leqslant \sum_{j\in R'_\ell} t_j\ \overline{\lim_{n_\ell\to\infty}}\ \frac{b^*_{\ell j}}{n_\ell} \\
&= \sum_{j\in R'_\ell} t_j\ \overline{\lim_{n_\ell\to\infty}}\ \frac{n_{\ell j}}{n_\ell}\frac{b^*_{\ell j}}{n_{\ell j}} \\
&\leqslant \sum_{j\in R'_\ell} t_j\ \overline{\lim_{n_\ell\to\infty}}\ \frac{n_{\ell j}}{n_\ell}\ \overline{\lim_{n_\ell\to\infty}}\ \frac{b^*_{\ell j}}{n_{\ell j}} \\
&= \sum_{j\in R'_\ell} t_j\ \overline{\lim_{n_\ell\to\infty}}\ \frac{n_{\ell j}}{n_\ell}\ \overline{\lim_{n_{\ell j}\to\infty}}\ \frac{b^*_{\ell j}}{n_{\ell j}} \qquad (\text{since } n_\ell\to\infty \Rightarrow n_{\ell j}\to\infty) \\
&= \gamma\sum_{j\in R'_\ell} t_j\ \overline{\lim_{n_\ell\to\infty}}\ \frac{n_{\ell j}}{n_\ell} \qquad (\text{since } b^*_{\ell j}/n_{\ell j}\to\gamma \text{ a.s. as } n_{\ell j}\to\infty) \\
&= \gamma\sum_{j\in R'_\ell} t_j\mu(A'_{\ell j})/\mu(A'_\ell),
\end{aligned} \tag{9}$$

where $\mu(A)\stackrel{\text{def}}{=}\int_A\mu$, and $A'_{\ell j}$ corresponds to the $j$th subregion of $A'_\ell$.

Now consider subregion $A'_1$. We have

$$\begin{aligned}
Z^{(1)} &= \sum_{j\in R'_1}\sum_k \bar{z}_\alpha(2\underline{\delta}_j, w(S_{1jk})) \\
&= \sum_{j\in R'_1}\sum_k [hw(S_{1jk})/(2f) + (2\underline{\delta}_j + c)f] \\
&= \sum_{i\in N_1} hw_i/(2f) + f\sum_{j\in R'_1}(2\underline{\delta}_j + c)b^*_{1j}.
\end{aligned}$$

Thus, almost surely

$$\overline{\lim_{n_1\to\infty}} \frac{Z^{(1)}}{n_1} = h/(2f) \overline{\lim_{n_1\to\infty}} \frac{1}{n_1}\sum_{i\in N_1} w_i + f \overline{\lim_{n_1\to\infty}} \frac{1}{n_1}\sum_{j\in R_1'} (2\underline{\delta}_j + c)b_{1j}^*$$

$$\leqslant h\mathsf{E}(w)/(2f) + \gamma f \sum_{j\in R_1'} (2\underline{\delta}_j + c)\mu(A_{1j}')/\mu(A_1') \quad \text{(by (9))}$$

$$= \beta hQ/2 + \frac{\gamma f}{\mu(A_1')}\sum_{j\in R_1'} (2\underline{\delta}_j + c)\mu(A_{1j}')$$

$$\leqslant \beta hQ/2 + \gamma f(2\bar{d}_1' + c). \tag{10}$$

The last inequality follows since $\underline{\delta}_j \leqslant d_i$ for all $i \in R_{1j}'$, i.e., for all points $i$ falling within subregion $A_{1j}'$.

Now consider subregion $A_2'$. We have

$$Z^{(2)} = \sum_{j\in R_2'}\sum_k \bar{z}_\alpha(2\underline{\delta}_j, w(S_{2jk}))$$

$$= \sum_{j\in R_2'}\sum_k [w(S_{2jk})\sqrt{(2\underline{\delta}_j + c)h\alpha/2Qf} + \sqrt{(2\underline{\delta}_j + c)hQf/2\alpha}]$$

$$= \sum_{j\in R_2'} [w(N_{2j})\sqrt{(2\underline{\delta}_j + c)h\alpha/2Qf} + b_{2j}^*\sqrt{(2\underline{\delta}_j + c)hQf/2\alpha}]$$

$$\leqslant \sum_{i\in N_2} w_i\sqrt{(2d_i + c)h\alpha/2Qf} + \sum_{j\in R_2'} b_{2j}^*\sqrt{(2\underline{\delta}_j + c)hQf/2\alpha}.$$

Thus, almost surely

$$\overline{\lim_{n_2\to\infty}} \frac{Z^{(2)}}{n_2}$$

$$\leqslant \overline{\lim_{n_2\to\infty}} \frac{1}{n_2}\sum_{i\in N_2} w_i\sqrt{(2d_i + c)h\alpha/2Qf} + \sqrt{hQf/2\alpha}\, \overline{\lim_{n_2\to\infty}} \frac{1}{n_2}\sum_{j\in R_2'} b_{2j}^*\sqrt{2\underline{\delta}_j + c}$$

$$\leqslant \mathsf{E}(w)\sqrt{h\alpha/2Qf}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2') + \gamma\sqrt{hQf/2\alpha}\sum_{j\in R_2'} \sqrt{2\underline{\delta}_j + c}[\mu(A_{2j}')/\mu(A_2')] \quad \text{(by (9))}$$

$$\leqslant \mathsf{E}(w)\sqrt{h\alpha/2Qf}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2') + \gamma\sqrt{hQf/2\alpha}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2')$$

$$= \sqrt{\beta\gamma hQf/2}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2') + \sqrt{\beta\gamma hQf/2}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2')$$

$$= \sqrt{2\beta\gamma hQf}\cdot \mathsf{E}(\sqrt{2d+c}\,|\,2d + c \in \Omega_2'). \tag{11}$$

Now consider ring $A_3'$. We have

$$Z^{(3)} = \sum_{j\in R_3'}\sum_k \bar{z}_\alpha(2\underline{\delta}_j, w(S_{3jk}))$$

$$= \sum_{j\in R_3'}\sum_k [(2\underline{\delta}_j + c)w(S_{3jk})/Q + hQ/2]$$

$$\leqslant \sum_{i\in N_3} (2d_i + c)w_i/Q + (hQ/2)\sum_{j\in R_3'} b_{3j}^*.$$

Thus, almost surely

$$\overline{\lim_{n_3\to\infty}} \frac{Z^{(3)}}{n_3} \leqslant \overline{\lim_{n_3\to\infty}} \frac{1}{n_3}\sum_{i\in N_3}(2d_i+c)w_i/Q+(hQ/2)\overline{\lim_{n_3\to\infty}}\frac{1}{n_3}\sum_{j\in R_3'} b_{3j}^*$$

$$\leqslant (2\bar{d}_3'+c)\beta f + hQ\gamma/2\sum_{j\in R_3'}\mu(A_{3j}')/\mu(A_3') \quad \text{(by (9))}$$

$$= (2\bar{d}_3'+c)\beta f + hQ\gamma/2. \tag{12}$$

The result then follows by combining (10)–(12) with (8). □

We now relate this upper bound (7) to the lower bound derived in Section 2. To do this we will need to define two functions $\underline{c}$ and $\bar{c}$.

**Lemma 2.9** (Lower bound cost allocation). *Define*

$$\underline{c}(d) = \begin{cases} hQ\beta/2+(2d+c)f\gamma & \text{if } 2d+c\in\Omega_1,\\ (2\beta-\gamma)((2d+c)f+hQ/2)+2(\gamma-\beta)\sqrt{(2d+c)hQf} & \text{if } 2d+c\in\Omega_2,\\ (2d+c)f\beta+\gamma hQ/2 & \text{if } 2d+c\in\Omega_3. \end{cases}$$

*Then*

$$\overline{\lim_{n\to\infty}}\frac{1}{n}\sum_{i=1}^{n}\underline{c}(d_i) = \underline{Z} \leqslant \underline{\lim}_{n\to\infty}\frac{1}{n}Z_n^{\mathrm{FP}} \quad \text{(a.s.)}.$$

**Proof.** The result follows by using Corollary 2.5 and noting that $\overline{\lim}_{n\to\infty}(1/n)\sum_{i=1}^{n}\underline{c}(d_i)$ is, by the strong law of large numbers, almost surely equal to the right-hand side of (3). □

**Lemma 2.10** (Upper bound cost allocation). *Define*

$$\bar{c}(d) = \begin{cases} hQ\beta/2+(2d+c)f\gamma & \text{if } 2d+c\in\Omega_1',\\ \sqrt{2hQf\beta\gamma(2d+c)} & \text{if } 2d+c\in\Omega_2',\\ (2d+c)f\beta+\gamma hQ/2 & \text{if } 2d+c\in\Omega_3'. \end{cases}$$

*Then*

$$\overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{FP}} \leqslant \underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{c}(d_i) \quad \text{(a.s.)}.$$

**Proof.** By definition

$$\overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{FP}} \leqslant \overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{GP}(\varepsilon)}.$$

By the strong law of large numbers,

$$\underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{c}(d_i)+2\varepsilon f$$

is almost surely equal to the right-hand side of (7). Therefore

$$\overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{FP}} \leqslant \overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{c}(d_i)+2\varepsilon f \quad \text{(a.s.)},$$

holds for any $\varepsilon>0$. Letting $\varepsilon$ go to zero, we conclude that

$$\overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{FP}} \leqslant \underline{\lim}_{\varepsilon\to 0}\overline{\lim_{n\to\infty}}\frac{1}{n}Z_n^{\mathrm{GP}(\varepsilon)} \leqslant \underline{\lim}_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\bar{c}(d_i) \quad \text{(a.s.)}. \qquad \square$$

An analysis of the ratio of $\underline{c}$ to $\bar{c}$ allows us to derive the asymptotic effectiveness of the lower bound, as described next.

**Lemma 2.11.** *Let*

$$\eta \stackrel{\text{def}}{=} \inf\left\{\frac{\underline{c}(\|x\|)}{\bar{c}(\|x\|)} : x \in \mathscr{A}\right\}$$

*then*

$$\underline{Z} \geqslant \eta \cdot \overline{\lim_{n\to\infty}} \frac{1}{n} Z_n^{\mathrm{FP}} \quad \text{(a.s.)},$$

*i.e.,* $\underline{Z}$ *is an asymptotic* $\eta$*-effective lower bound on the average cost per retailer in the best FPP.*

**Proof.** First note that $\underline{c}(\|x\|) \geqslant \eta\bar{c}(\|x\|)$ for any $x \in \mathscr{A}$. The result then follows by using the definitions of $\bar{c}$ and $\underline{c}$ and Lemmas 2.9 and 2.10. □

**Theorem 2.12.** $\underline{Z}$ *is an asymptotic* 98.5%*-effective lower bound on the cost of the best FPP.*

**Proof.** By Lemma 2.11 we need only determine the value of $\eta$. For $2d + c \in \Omega_1 \cap \Omega_1' = \Omega_1$ and $2d + c \in \Omega_3 \cap \Omega_3' = \Omega_3$, the cost allocations are identical. We therefore consider $2d + c \in \Omega_2$. This is separated into three cases. In each case we use the substitutions $v = 2(2d + c)f/hQ$ (or equivalently $(2d + c)f = vhQ/2$) and $\alpha = \gamma/\beta$.

*Case* 1: $2d + c \in \Omega_2 \cap \Omega_1'$. Hence $2d + c \in [Qh/4f, Qh/(2f\alpha))$ and $\frac{1}{2} \leqslant v < 1/\alpha$. Taking the derivative of the ratio $\underline{c}/\bar{c}$ with respect to $v$ shows that the ratio is decreasing with $v$ and thus we replace $v$ with its upper bound of $1/\alpha$ (i.e., $2d + c = hQ/(2f\alpha)$). The ratio can then be written as

$$\frac{(2-\alpha)(1+\alpha) + 2(\alpha-1)\sqrt{2\alpha}}{2\alpha}. \tag{13}$$

This expression is minimized at $\alpha$ satisfying the equation $\alpha^2 - \sqrt{2\alpha}(\alpha + 1) + 2 = 0$. This is at $\alpha \approx 1.2599$ where the ratio is approximately 99.12%.

*Case* 2: $2d + c \in \Omega_2 \cap \Omega_2'$. Hence $2d + c \in [Qh/(2f\alpha), Qh\alpha/(2f)]$ and $1/\alpha \leqslant v \leqslant \alpha$. The ratio $\underline{c}/\bar{c}$ is minimized at $v = 1$ (or $2d + c = hQ/2f$). At this point the ratio is

$$(\sqrt{2} - 1)\,\frac{\alpha + \sqrt{2}}{\sqrt{\alpha}}. \tag{14}$$

This ratio is minimized at $\alpha = \sqrt{2}$ where its value is $(4 - 2\sqrt{2})/\sqrt{\sqrt{2}} \approx 98.52\%$.

*Case* 3: $2d + c \in \Omega_2 \cap \Omega_3'$. Hence $2d + c \in (Qh\alpha/(2f), Qh/f]$ and $\alpha < v \leqslant 2$. Taking the derivative of $\underline{c}/\bar{c}$ with respect to $v$ shows that the ratio is increasing in $v$ as long as $v \leqslant 2$. We therefore set $v = \alpha$. This corresponds to replacing $2d + c$ with its lower bound of $Qh\alpha/(2f)$. We get an expression which is minimized at $\alpha$ satisfying the same equation as in Case 1 and giving the same value (99.12%). □

It can be verified that the asymptotic effectiveness of the lower bound for a general distribution of demand rates with packing efficiency $\alpha$ is given by

$$(\sqrt{2} - 1)\frac{\alpha + \sqrt{2}}{\sqrt{\alpha}}. \tag{15}$$

That is, for any $\alpha \in [1, 2)$, the value in expression (13) is no less than the value in (14). In the following figure, the asymptotic effectiveness of the lower bound (15) is graphed as a function of $\alpha$ (Fig. 2).

We now show that the asymptotic bound is the tightest possible.

**Theorem 2.13.** *If* $c < hQ/2f$, *then the asymptotic effectiveness bound proven in Theorem* 2.12 *is the tightest possible.*

**Proof.** We prove this by showing that for any value of $\alpha \in [1, 2)$, we can construct an example where the ratio of the lower bound to the best possible fixed partition policy is as given by (15). Consider an example with $n$ retailers that are located at various points on a ring at a distance exactly $hQ/(4f) - c/2 > 0$ from the depot. Assume each retailer demand rate is equal to $Qf/\alpha$, and thus $\gamma = 1$ and $\beta = 1/\alpha$. For this example, since $Qf/\alpha > Qf/2$ there is only one possible partition: each retailer is served separately. In the optimal FPP each of these sets is served at its optimal replenishment rate. Note that for all sets (retailers) neither frequency nor capacity constraint is active. The cost of the policy is $hQ/\sqrt{\alpha}$ per retailer, therefore the total cost is $n \cdot hQ/\sqrt{\alpha}$. By setting $u(w) = 1$, for all $w > Qf/2$ (and $u(w) = 0$ otherwise), the lower bound is $hQ(\sqrt{2} - 1)(\sqrt{2}/\alpha + 1)$ per retailer or $nhQ(\sqrt{2} - 1)(\sqrt{2}/\alpha + 1)$ in total. The ratio of these two quantities is exactly the expression in (15). □

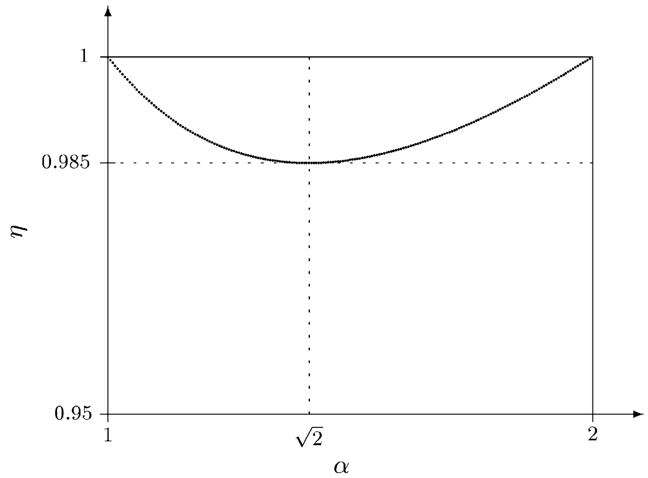


Fig. 2. The asymptotic effectiveness of the lower bound $\underline{Z}$ as a function of $\alpha$.

We note that if $c \geqslant hQ/2f$ then the asymptotic effectiveness of the lower bound is at least 98.5%.

## 3. Conclusion

For an appropriately chosen $u$, the lower bound $\sum_{i\in N} g_u(d_i, w_i)$ converges to 98.5% of the cost of the best FPP. This bound can be used to gauge the effectiveness of any fixed partition policy. In a companion paper [3], we actually construct an FPP using a region partitioning scheme and show, using the lower bound developed here, that its cost is within 1.5% of the best FPP (asymptotically).

We conclude with an additional remark concerning the lower bound above. It can also be used to further understand the results of Chan et al. [8] concerning the asymptotic effectiveness of fixed partition policies in general. In particular, when the demand rates are "difficult to pack" (i.e., $\alpha \approx 2$), Chan et al. show that the best FPP is no more than $\sqrt{2}$ times their lower bound (asymptotically). Unfortunately, it is not clear whether the gap comes from their specific choice of FPP and another FPP exists that comes closer to their lower bound, or their lower bound is weak within the class of FPPs. We have shown here that the *latter* is the case. In fact, for the case of $\alpha \approx 2$, the fixed partition policy they construct using GP($\varepsilon$) is very nearly asymptotically optimal (as $\varepsilon \to 0$) within the class of fixed partition policies (though it takes exponential-time to construct it). This can be deduced from the fact that in this case our lower bound is essentially asymptotically tight, since $\eta \to 1$ as $\alpha \to 2$.

In our analysis of the upper bound, we assume that the retailers in each subregion are assigned separately to routes by using an optimal solution to the bin-packing problem. However, an interesting question is whether the quality of the FPP constructed is sensitive to the specific optimal packing, in case the optimal bin-packing solution is not unique. Moreover, assume that for a given subregion there exists two optimal packings, where in one all bins are assigned the same weight and in the other there is an imbalance among the bins. Which packing is expected to give a better upper bound? These are open questions for future research. We provide here an example which demonstrates that the quality of the solution can be significantly affected by the specific choice of the optimal packing.

*An example*. For any integers $n_1, n_2$ with $n_1/n_2 \in [1,2)$, the example has $n_1(2n_2 - n_1 + 1)$ retailers that are located at a single point that is a distance exactly $hQ/(4f) - c/2 > 0$ from the depot. The demand rate is equal to $Qf/2 + \delta$ for $n_1$ of the retailers and $(n_2Qf - n_1(Qf/2+\delta))/n_1(2n_2 - n_1)$ for the rest of them. In this case, $\gamma = (2n_2 - n_1 + 1)^{-1}$, $\beta = \gamma n_2/n_1$ and $\alpha = n_1/n_2$. By setting $u(w) = 1$ for all $w > Qf/2$ (and $u(w) = 0$ otherwise), the lower bound is $n_1 hQ(\sqrt{2} - 1)(\sqrt{2}n_2/n_1 + 1)$ in total.

We now construct an optimal bin packing solution. This puts each retailer with demand rate $Qf/2 + \delta$ with $2n_2 - n_1$ retailers with demand rate $(n_2Qf - n_1(Qf/2 + \delta))/n_1(2n_2 - n_1)$ together in one bin. Each of these subsets of retailers is then served at its optimal replenishment rate. Note that for all these subsets of retailers neither frequency nor capacity constraint is active. The cost of the policy is $hQ/\sqrt{n_1/n_2}$ for each set of retailers. Therefore the total cost is $n_1 hQ/\sqrt{n_1/n_2}$. The ratio of the lower bound to this quantity is exactly the expression in (15).

On the other hand, put each retailer with demand rate $Qf/2+\delta$ into one bin and then $n_1$ of the retailers with demand rate $(n_2Qf - n_1(Qf/2+\delta))/n_1(2n_2-n_1)$ to each of the $2n_2-n_1$ of these $n_1$ bins is also an optimal bin-packing solution. Note that for all these subsets of retailers neither frequency nor capacity constraint is active. By serving these subsets at their optimal replenishment rates, the total cost of the policy is $2(n_1-n_2)2\sqrt{hQ/2f[h/2(Qf/2+\delta)]} + (2n_2-n_1)2\sqrt{hQ/2f[h/2(Qf/2-(2(n_2-n_1)/(2n_2-n_1))\delta)]}$ that goes to the lower bound as $\delta \to 0$.

## Acknowledgements

The authors would like to thank several anonymous referees for their useful comments, including the example just stated.

## Appendix A.

**Property A.** *The function $g_u(d,\cdot)$ is non-decreasing in $d \geqslant 0$.*

**Proof.** This can be easily verified algebraically. □

**Lemma 2.2.** *Let $u$ be dual feasible for $\psi$. For any $S \subseteq N$ with $w(S) \leqslant Qf$, we have*

$$\sum_{i\in S}^{n} g_u(d_i,w_i) \leqslant z(S).$$

**Proof.** Consider a set $S$ served on a route. Let $L \stackrel{\text{def}}{=} L^*(S)$, $W \stackrel{\text{def}}{=} w(S)$ and $K \stackrel{\text{def}}{=} L+c$. Define $m \stackrel{\text{def}}{=} \lfloor Qf/W \rfloor$, i.e., $Qf/(m+1) < W \leqslant Qf/m$. Note that since $u$ is dual feasible:

$$W \leqslant \frac{Qf}{m} \Rightarrow \sum_{i\in S} u(w_i) \leqslant \frac{1}{m}. \tag{A.1}$$

This follows since $m$ *copies* of $S$ correspond to a feasible bin, and therefore must satisfy the feasibility constraint, i.e., $m\sum_{i\in S} u(w_i) \leqslant 1$.

We consider three cases:

*Case* 1: $K \in \Omega_1$. Then

$$\begin{aligned}
\sum_{i\in S} g_u(d_i,w_i) &= \sum_{i\in S} [hw_i/(2f) + u(w_i)(2d_i+c)f] \\
&\leqslant \sum_{i\in S} [hw_i/(2f) + u(w_i)Kf] \quad (\text{since } 2d_i \leqslant L,\ \forall i \in S) \\
&= hW/(2f) + Kf\sum_{i\in S} u(w_i) \\
&\leqslant hW/(2f) + Kf/m,
\end{aligned}$$

where the last inequality follows from (A.1). We consider two subcases:

*Subcase* 1.1: $(K,W) \in \mathscr{F}$. Since $z(S) = hW/(2f) + Kf$, the result follows since $m \geqslant 1$.

*Subcase* 1.2: $(K,W) \in \mathscr{N}$, i.e., $W < 2Kf^2/h$. Since $z(S) = \sqrt{2hWK}$, we need only show that

$$hW/(2f) + Kf/m \leqslant \sqrt{2hWK}$$

for $m \geqslant 2$ (see Fig. 1, $K \in \Omega_1$ and $(K,W) \in \mathscr{N}$ implies $W \leqslant Qf/2$). Again, considering the inequality as a function of $W$, we see that the inequality is satisfied as long as $W$ is between the two numbers:

$$\frac{2Kf^2}{h}\left(1 \pm \sqrt{1-\frac{1}{m}}\right)^2.$$

Since $W < 2Kf^2/h$, $W$ is clearly less than the larger of the two numbers. We must therefore check that

$$\frac{Qf}{m+1} \geqslant \frac{2Kf^2}{h}\left(1-\sqrt{1-\frac{1}{m}}\right)^2.$$

We can set $K$ to its largest possible value in $\Omega_1$ (i.e., $K = Qh/4f$). This becomes

$$\frac{2}{m+1} \geqslant \left(1-\sqrt{1-\frac{1}{m}}\right)^2, \tag{A.2}$$

which holds for any $m \geqslant 2$.

*Case* 2: $K \in \Omega_2$. We use the shorthand notation: $p \equiv p(L)$ and $q \equiv q(L)$. Since $2d_i \leqslant L$ and by Property A:

$$\begin{aligned}\sum_{i\in S} g_u(d_i, w_i) &\leqslant \sum_{i\in S} g_u(L/2, w_i)\\ &= \sum_{i\in S}[pw_i/Qf + u(w_i)q]\\ &= pW/Qf + q\sum_{i\in S} u(w_i)\\ &\leqslant pW/Qf + q/m,\end{aligned}$$

where the last inequality follows from (A.1).

We consider three subcases.

*Subcase* 2.1: $(K, W) \in \mathscr{C}$, and thus $m=1$. Since $z(S)=KW/Q+hQ/2$ we need only show that: $KW/Q+hQ/2 \geqslant pW/Qf+q$. Note that each is linear in $W$ hence we can compare slopes. Since $p/Qf \geqslant K/Q$ we make $W$ as large as possible (equal to $Qf$) and see that the bound does hold (since $p+q \leqslant Kf + hQ/2$, see (1) and (2)).

*Subcase* 2.2: $(K, W) \in \mathscr{N}$. Since $z(S) = \sqrt{2hWK}$ we need only show that

$$pW/Qf + q/m \leqslant \sqrt{2hWK}. \tag{A.3}$$

We consider two cases: $m = 1$ and $m \geqslant 2$.

If $m=1$, then $W > Qf/2$. By construction, $pW/Qf+q$ is a lower bound on the function $z(L, W)$ for any $W \in (Qf/2, Qf]$ and for $L + c = K \in \Omega_2$. Therefore it is a lower bound on the strictly concave part of the function.

If $m \geqslant 2$, then we consider (A.3) and see that the inequality is satisfied as long as $W$ is between the two numbers:

$$\frac{Q^2f^2hK}{2p^2}\left(1 \pm \sqrt{1\frac{2pq}{mQfhK}}\right)^2.$$

We first consider the larger of the two numbers. We perform the substitution $a = Kf + hQ/2$ and $b = \sqrt{KhQf}$. Then $p = 2(a-b)$ and $q = 2b - a$. We must show that

$$\frac{Qf}{m} \leqslant \frac{Qfb^2}{8(a-b)^2}\left(1+\sqrt{1-\frac{4(a-b)(2b-a)}{mb^2}}\right)^2. \tag{A.4}$$

We then use the substitution $x = a/b$, and note that in $\Omega_2$ we have $\sqrt{2} \leqslant x \leqslant \frac{3}{2}$. Inequality (A.4) then becomes

$$\frac{1}{m} \leqslant \frac{1}{8(x-1)^2}\left(1+\sqrt{1-\frac{4(x-1)(2-x)}{m}}\right)^2. \tag{A.5}$$

Or equivalently

$$\sqrt{\frac{8}{m}}(x-1) \leqslant 1+\sqrt{1-\frac{4(x-1)(2-x)}{m}}.$$

Then

$$\left(\sqrt{\frac{8}{m}}(x-1)-1\right)^2 \leqslant 1-\frac{4(x-1)(2-x)}{m},$$

which is equivalent to

$$x^2 - x(1+\sqrt{2m}) + \sqrt{2m} \leqslant 0.$$

This inequality is satisfied as long as $x$ is between the numbers 1 and $\sqrt{2m}$ which is the case for $m \geqslant 2$.

We now consider the smaller of the two numbers. Using the same substitutions and following (A.5), we must show that

$$\frac{1}{m+1} \geqslant \frac{1}{8(x-1)^2}\left(1-\sqrt{1-\frac{4(x-1)(2-x)}{m}}\right)^2.$$

Or, equivalently

$$\sqrt{\frac{8}{m+1}}(x-1) \geqslant 1-\sqrt{1-\frac{4(x-1)(2-x)}{m}}.$$

Then

$$x^2(m-1) + x(3-m-m\sqrt{2(m+1)}) + m\sqrt{2(m+1)} - 2 \leqslant 0.$$

We need only check that the inequality holds for the extreme values of $x$, i.e., $x=\sqrt{2}$ and $x=\frac{3}{2}$. This is the true for all $m \geqslant 2$.

*Suubcase* 2.3: $(K,W) \in \mathscr{F}$, and thus $m=1$. Since $z(S)=Kf+hW/(2f)$, we need only show that: $pW/Qf+q \leqslant Kf+hW/(2f)$. Each is linear in $W$ hence we compare slopes. Since one can verify that $p/Qf \geqslant h/2f$, we make $W$ as large as possible (equal to $Qf$) and, using the fact that $p+q \leqslant Kf+hQ/2$, we see that the bound does hold.

*Case* 3: $K \in \Omega_3$. Then

$$\begin{aligned}\sum_{i\in S} g(d_i,w_i) &= \sum_{i\in S}[(2d_i+c)w_i/Q + u(w_i)hQ/2]\\ &\leqslant \sum_{i\in S}[Kw_i/Q + u(w_i)hQ/2] \quad (\text{since } 2d_i \leqslant L,\ \forall i \in S)\\ &= KW/Q + hQ\sum_{i\in S}u(w_i)/2\\ &\leqslant KW/Q + hQ/(2m),\end{aligned}$$

where the last inequality follows from (A.1). We consider two subcases:

*Subcase* 3.1: $(K,W) \in \mathscr{C}$. Since $z(S)=KW/Q+hQ/2$, the result follows since $m \geqslant 1$.

*Subcase* 3.2: $(K,W) \in \mathscr{N}$, i.e., $WK < Q^2h/2$ and $m \geqslant 2$. Then $z(S)=\sqrt{2hWK}$. We therefore need to show that

$$KW/Q + hQ/(2m) \leqslant \sqrt{2hWK},$$

for $m \geqslant 2$. If we consider this inequality as a function of $W$, it is satisfied as long as $W$ is between the two numbers, i.e.:

$$\frac{Q^2h}{2K}\left(1 \pm \sqrt{1-\frac{1}{m}}\right)^2.$$

Since $W < Q^2h/(2K)$, $W$ is clearly less than the larger of the two numbers. We must therefore check that the lower bound on $W$ is at least the smaller of the two numbers:

$$\frac{Qf}{m+1} \geqslant \frac{Q^2h}{2K}\left(1-\sqrt{1-\frac{1}{m}}\right)^2.$$

We can assume $K$ is at its smallest value in $\Omega_3$ (i.e., $K=Qh/f$). This results in

$$\frac{2}{m+1} \geqslant \left(1-\sqrt{1-\frac{1}{m}}\right)^2,$$

which is exactly (A.2) and was verified to be true for all $m \geqslant 2$ in Subcase 1.2. □

## References


[1] S. Anily, The general multi-retailer EOQ problem with vehicle routing costs, European J. Oper. Res. 79 (1994) 451–473.

[2] S. Anily, J. Bramel, Vehicle routing and the supply chain, in: M. Magazine, R. Ganeshan, S. Tayur (Eds.), Quantitative Models for Supply Chain Management, Kluwer, Dordrecht, 1998.

[3] S. Anily, J. Bramel, A polynomial region partitioning scheme for fixed partition policies for the inventory-routing problem, Working Paper, Columbia Graduate School of Business, 1999.

[4] S. Anily, A. Federgruen, A class of Euclidean routing problems with general route cost functions, Math. Oper. Res. 15 (1990) 268–285.

[5] S. Anily, A. Federgruen, One warehouse multiple retailer systems with vehicle routing costs, Manage. Sci. 36 (1990) 92–114.

[6] J. Bramel, A note on the tightness of the asymptotic effectiveness bounds for fixed partition policies, Working Paper, Columbia Graduate School of Business, 1998.

[7] J. Bramel, D. Simchi-Levi, A location based heuristic for general routing problems, Oper. Res. 43 (1995) 649–660.

[8] L.M.A. Chan, A. Federgruen, D. Simchi-Levi, Probabilistic analyses and practical algorithms for inventory-routing models, Oper. Res. 46 (1998) 96–106.

[9] E.G. Coffman Jr., M.R. Garey, D.S. Johnson, Approximation algorithms for bin-packing—an updated survey, in: G. Ausiello, M. Lucertini, P. Serafini (Eds.), Algorithm Design for Computer System Design, Springer, Berlin, 1984, pp. 49–106.

[10] E.G. Coffman Jr., G.S. Lueker, Probabilistic Analysis of Packing and Partitioning Algorithms, Wiley, New York, 1991.

[11] G. Gallego, D. Simchi-Levi, On the effectiveness of direct shipping strategy for the one warehouse multi-retailer R-systems, Manage. Sci. 36 (1990) 240–243.

[12] Y.T. Herer, R. Roundy, Heuristics for a one-warehouse multiretailer distribution problem with performance bounds, Oper. Res. 45 (1997) 102–115.

[13] J.F.C. Kingman, Subadditive Processes, in: Lecture Notes in Mathematics, Vol. 539, Springer, Berlin, 1976.

[14] G.S. Lueker, Bin packing with items uniformly distributed over intervals $[a, b]$, in: Proceedings of the 24th Annual Symposium on Foundations of Computer Science, 1983, pp. 289–297.

[15] W.T. Rhee, M. Talagrand, Martingale inequalities and $\mathcal{NP}$-complete problems, Math. Oper. Res. 12 (1987) 177–181.

[16] S. Viswanathan, K. Mathur, Integrating routing and inventory decisions in one warehouse multi-retailer multi-product distribution systems, Manage. Sci. 43 (1997) 294–312.