Finding optimal policies in the $(S-1, S)$ lost sales inventory model with multiple demand classes

Willem van Jaarsveld* and Rommert Dekker

Econometric institute, Erasmus University Rotterdam,
P.O. Box 1738, 3000 DR Rotterdam, The Netherlands

Report Econometric Institute EI 2009-14

Juli 2009

Abstract

This paper examines the algorithms proposed in the literature for finding good critical level policies in the $(S-1, S)$ lost sales inventory model with multiple demand classes. Our main result is that we establish guaranteed optimality for two of these algorithms. This result is extended to different resupply assumptions, such as a single server queue. As a corollary, we provide an alternative proof of the optimality of critical level policies among the class of all policies.

Keywords: Inventory; Multiple demand classes; Customer differentiation; Rationing; Lost sales; Stochastic dynamic programming.

* Corresponding author: E-mail: vanjaarsveld@ese.eur.nl

1 Introduction

In many inventory systems, customers belong to different classes, for instance differing in their willingness to pay for fast delivery of their orders. In order to increase their profits, some companies provide different customer classes with different levels of service. This can be achieved by using inventory rationing, a concept in which inventory is withheld from less demanding, lower profit customer classes to preserve it for future, more critical demands. A related concept is a critical level policy, in which each customer class is assigned a critical level. When stock is below the critical level assigned to a particular customer class, the stock is withheld from that customer class and preserved for more important customer classes.

The problem of multiple demand classes was first described by Veinott (1965), who also introduced critical level policies. Topkis (1968) shows optimality of critical level policies for a system with generally distributed demand, periodic review and zero leadtime, in which case the critical levels depend on the time until the next review. Ha (1997) considers critical level policies in a make-to-stock system with lost demand, under a Poisson demand assumption. The production decision is an integral part of the model. He established optimality of critical levels and shows that demands of the highest criticality should always be satisfied. Furthermore, he shows that a base stock policy is optimal for managing production. This work was extended to the back-ordering case by de Véricourt, Karaesmen and Dallery (2002).

Dekker, Hill, Kleijn and Teunter (2002) consider the optimization of the critical levels and the base stock level for a problem with independent leadtimes. They derive expressions for the costs of a given critical level base stock policy. Subsequently, they derive bounds for the base stock level S on the basis of which the optimal critical level policy can be found, by solving the optimization problem for each possible S by explicit enumeration. Explicit enumeration is prohibitively slow for problems with many demand classes and large S. Therefore, Dekker et al. (2002) propose a fast approach to find good critical levels for which optimality is not guaranteed. For the case of two demand classes, Melchiors, Dekker and Kleijn (2000) extend this work to fixed quantity ordering. Deshpande, Cohen and Donohue (2003) consider a similar model, but with back-ordering of unsatisfied demand. The order in which back-ordered demands are satisfied leads to additional complications.

Continuing along the lines of Dekker et al. (2002), Kranenburg and van Houtum consider optimization of the critical levels and the base-stock level. Similarly to Dekker et al. (2002), the problem is split up into a number of sub-problems for fixed S. Kranenburg and van Houtum propose three algorithms for solving these sub-problems. In an extensive numerical experiment, they find that these algorithms are much faster (in the order of 200-1000 times as fast for problems with 2 to 5 demand classes) than complete enumeration. Moreover, the algorithms appear to find optimal solutions. Based on this, they conjecture without proof that the algorithms are optimal for all possible instances.

This paper examines the algorithms proposed by Kranenburg and van Houtum (2007) for finding good critical level policies in the $(S-1,S)$ lost sales inventory model with multiple demand classes. These algorithms resemble local search algorithms; for a precise description we refer to Section 4, or to the mentioned article. A question arising from their contribution is whether these algorithms can get stuck in a local optimum. We will answer this question negatively; we prove that the algorithms result in optimal solutions. This is a surprising result, because non-randomized local searches are known to get stuck in local optima in many other problems. We extend this result to a make-to-stock queue in which a base-stock level is fixed and we search for the optimal critical levels. As a corollary we establish the optimality of critical level policies, recovering and strengthening a result that was essentially derived by Miller (1969). To obtain the results, we rely on general theory on undiscounted Markov decision problems to derive results regarding the structure of the bias of "locally optimal" critical level policies. Ultimately, we show that the bias of such policies solve the optimality equations.

Kranenburg and van Houtum argue that there is a need for fast and accurate algorithms, and they show that their algorithms are fast. Our main contribution is that these algorithms can now be used in certainty that optimal solutions will be obtained. Furthermore, we show that the same general theory used for establishing structural results in many inventory models can also be used to devise fast special purpose algorithms for finding the optimal policy in inventory models. Lastly, we show that critical levels are optimal among the class of all policies for the model we consider.

The remainder of this paper is organized as follows. The model is formulated as a Markov decision process in Section 2. We then restate some general results from Markov decision theory in Section 3. The optimality of the algorithms is proved in Section 4. Some extensions are discussed in Section 5. Section 6 concludes.

2 The model

We consider the model studied earlier by Dekker et al. (2002) and Kranenburg and van Houtum (2007). They use minimization of the long term average cost as optimality criterion. To comply with the convention used in Puterman (1994), we will interpret the costs as negative rewards and use maximization of the long term average reward as optimality criterion. Clearly, these two formulations are equivalent.

Demands for a part are classified according to criticality. Let J be the set of demand classes ($|J| \geq 1$). For each class $j \in J$, demands occur according to a Poisson process with rate $m_j > 0$. If an item is not delivered to class j upon request, the demand is lost and a penalty cost $p_j > 0$ is to be paid, which will be interpreted as a negative reward. Classes are numbered $1, 2, \ldots, |J|$ such that $p_1 \geq p_2 \geq \ldots \geq p_{|J|}$. The item is stored in a single stock location, and stock for the item is controlled by an order-up-to-S policy. We denote the state of the system by $k \in \{0, \ldots, S\}$, where k denotes the number of items on order. The heuristics of which we will prove optimality find critical levels for fixed S. We also assume fixed S, but for optimization purposes S can be enumerated in a separate loop using the bounds derived by Dekker et al. (2002).

Kranenburg and van Houtum (2007, Remark 1) make the important observation that under linear holding costs in the amount of stock on hand, we can assume without loss of generality that holding costs are also charged for items in replenishment. Under this assumption, the holding costs do not depend on the control of the system for fixed S, and can be omitted when considering optimization of the critical levels.

We assume i.i.d. exponential leadtimes. In Section 5 we show how to extend this assumption to the assumption of i.i.d. general leadtimes, as long as the control of the system is restricted to be of a certain type. We denote the rate by which new parts arrive in state k by $\nu_k = kL^{-1}$, where L is the expected leadtime. For convenience of notation, we include $\nu_0 = 0$ in this definition.

In order to model the problem as a Markov decision problem, we consider more general policies than the critical level policies to which Kranenburg and van Houtum (2007) restrict their attention. We let A_k be the set of Markovian deterministic decision rules in state k.

Each decision rule $a \in A_k$ prescribes which demand classes to accept and which to reject in state k. For $k < S$, each $a \in A_k$ is denoted as a subset of the set of demand classes J. E.g. if $a = \{1, 3, 4\}$ is selected as the decision rule in state k, then this denotes that under rule a demand classes $1, 3$ and 4 are accepted and other demand classes are rejected in state k. Thus, A_k is isomorphic with the powerset $\mathcal{P}(J)$ of J. In state S all demands are necessarily rejected. A_S thus consists only of the empty set. A Markovian deterministic stationary policy consists of a decision rule $a \in A_k$ for each state k. A policy will be denoted by $d = (d(0), \dots, d(S)) \in A_0 \times \dots \times A_S = D^{\mathrm{MD}}$. We will consider only stationary policies, a restriction that we will motivate in the following.

Because the time intervals between successive events are exponential, the problem can be modelled as a continuous time Markov decision process. Under the assumption that the control is only changed when transitions occur (a weak condition that can still be weakened), uniformization can be applied and the model can be transformed into a discrete-time Markov decision process which is equivalent in terms of long term average reward (see e.g. Puterman (1994, Section 11.5.3)). We will apply this transformation, and work with the transformed model. Under conditions valid for this discrete time model, Puterman (1994, Theorem 8.4.5) shows that there exists a stationary deterministic average optimal policy, which motivates our restriction to policies of this type.

The states of the transformed model are the same as the states of the original model. For a complete description of the discrete time model we further need the rewards and transition probabilities in state k under decision $a \in A_k$. After transforming the model, the transition probabilities can be found to be equal to

$$p(i|k,a) = \begin{cases} \hat{c}^{-1} \sum_{j \in a} m_j & i = k+1,\ k \neq S, \\ \hat{c}^{-1} \left(\nu_S - \nu_k + \sum_{j \in J \setminus a} m_j \right) & i = k, \\ \hat{c}^{-1} \nu_k & i = k-1,\ k \neq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The reward vector becomes

$$r(k,a) = -\hat{c}^{-1} \sum_{j \in J \setminus a} p_j m_j. \tag{2}$$

In the previous, we used the uniformization constant

$$\hat{c} = \nu_S + \sum_{j \in J} m_j. \tag{3}$$

By definition, $J \setminus a$ denotes the elements contained in J, but not in a; it thus denotes the demand classes which are declined under decision a. We denote the transition matrix under policy d by P_d, it has $p(i|k, d(k))$ as its (k,i)th entry. The reward vector for this policy will be denoted by r_d, it has $r(k, d(k))$ as its kth entry. Note that the model has $S+1$ states, so the transition matrix for any policy d is $(S+1)$ by $(S+1)$ and the reward vector has $S+1$ elements.

3 Existing theory

Our proof relies on a number of results in undiscounted Markov decision theory. These results hold for unichain, finite state Markov decision problems with finite decision sets and, consequently, bounded rewards. Note that the model we consider fulfills these conditions. The model is unichain by noting that state 0 (no orders outstanding) can be reached from any state in a finite number of steps, under any policy.

We start by defining a function that will enable us to efficiently denote the results that we need. Let $g \in \mathbb{R}$ and let h be a real-valued vector in $S+1$ dimensions. Define

$$B_d(g, h) = r_d - ge + (P_d - I)h \tag{4}$$

where I is the identity matrix and e is the vector with all entries equal to 1, both of appropriate dimension. This definition is similar to the definition of $B(g, h)$ in Puterman (1994, (8.4.3)), except that it does not include the maximum over all decisions $d \in D$ and therefore it depends on d.

When a policy $d \in D^{\text{MD}}$ is fixed, the model reduces to a Markov reward process. For the model under consideration, this Markov reward process induces a unique long term average reward g_d and a bias vector h_d. These quantities satisfy the a relation that will be exposed in the following lemma.

Lemma 1. *For a given policy $d \in D^{MD}$, the Markov decision problem reduces to a Markov reward process with transition matrix P_d and reward vector r_d. The average expected reward g_d and bias $\{h_d\}_{k=0}^{S}$ of this unichain Markov reward process satisfy*

$$B_d(g_d, h_d) = 0. \tag{5}$$

Furthermore, this equation determines g_d uniquely, and h_d up to an overall constant.

Proof. The result is a slight reformulation of Corollary 8.2.7 of Puterman (1994) and the remarks following it. □

Now, we will establish a link between the reward of two policies. To this end, we will need the limiting matrix which we will discuss here first. The results we state here can be found in Puterman (1994, Appendix A.4). Let P_d^* denote the limiting matrix associated with P_d

$$P_d^* = \lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{N} P_d^{t-1}.$$

Denote the (k, i)th element of this matrix by $p_d^*(i|k)$. For unichain Markov reward processes, this matrix has equal rows, and its elements are given by

$$p_d^*(i|k) = p_d^*(i)$$

where $p_d^*(i)$ is the long term fraction of time that the system is in state i under policy d. For recurrent states under policy d, $p_d^*(i) > 0$. Because $p_d^*(i|k)$ does not depend on the initial state k, the long term average expected reward does not depend on the initial state either. This is reflected by the fact that the average expected reward vector has equal elements. It is given by $g_d e = P_d^* r_d$. P_d^* satisfies $P_d^* P_d = P_d^*$. Note also that in a finite state space P_d^* is stochastic, so $P_d^* e = e$. These two equations can be used to find the steady state probabilities. Another approach to finding the steady state probabilities is by using a queueing theory argument, as is done by Kranenburg and van Houtum (2007). They subsequently use $g_d e = P_d^* r_d$, or equivalently

$$g_d = \sum_{j=0}^{S} p_d^*(j) r_d(j)$$

to find the long term average reward associated with policy d.

The following result uses the limiting matrix to establish a link between the average reward of two policies. It will be pivotal in proving a key property of the bias of the policies found by the algorithms of which we will prove the optimality.

Lemma 2. *Let $d \in D^{\mathrm{MD}}$ and let g_d and h_d be the gain and bias associated with d. Let d' denote another policy ($\in D^{\mathrm{MD}}$) with associated average expected reward $g_{d'}$. Let $P^*_{d'}$ denote the limiting matrix associated with $P_{d'}$. Then we have*

$$g_{d'}e = g_d e + P^*_{d'} B_{d'}(g_d, h_d).$$

Proof. We adapt the proof of Proposition 8.6.1 of Puterman (1994). We know that $g_{d'}e = P^*_{d'} r_{d'}$. We add and subtract $g_d e$ at the right hand side of this equation. Now, we note that $P^*_{d'}(P_{d'} - I) = 0$ and $P^*_{d'} e = e$, and obtain

$$g_{d'}e = g_d e + P^*_{d'} \left(r_{d'} - g_d e + (P_{d'} - I) h_d \right).$$

The result can be easily recognized using (4). □

The next lemma gives conditions under which a policy is optimal.

Lemma 3. *Let $d \in D^{MD}$ and let g_d and h_d be the gain and bias associated with d. If*

$$\max_{d' \in D^{MD}} B_{d'}(g_d, h_d) = 0 \tag{6}$$

then g_d is the optimal average expected reward, and d is an optimal policy attaining this reward.

Proof. g_d is the optimal reward by Puterman (1994, Theorem 8.4.1 c). Now, note that

$$B_d(g_d, h_d) = 0$$

by Lemma 1, which means that d attains the maximum in (6). We now apply Puterman's (1994) Theorem 8.4.4 to conclude optimality of d. □

4 Optimality of the algorithms

The final policies obtained when applying Algorithm 1 and 2 proposed in Section 5 of Kranenburg and van Houtum (2007) have a number of properties, which we formalize as follows. For ease of reference, we list the algorithms before Theorem 1.

Definition 1. *A policy d will be said to belong to the locally optimal critical level policies D^L if it has the following two properties*

i) d is of critical level type, viz, for each demand class $j \in J$ there exists a critical level $c_j \in \{0, \ldots, S\}$, such that demands of class j are accepted when $k < S - c_j$, and declined when $k \geq S - c_j$. So $j \in d(k)$ if and only if $k < S - c_j$. Furthermore, the critical levels are monotone in demand criticality, i.e. $i > j \Rightarrow c_i \geq c_j$. Note that these critical levels fully determine a policy, but that not every policy can be described by a set of critical levels.

ii) d is locally optimal, in the sense that a unit increase or decrease of any single critical level such that monotonicity is not violated does not result in an increase of the average expected reward.

In the following, we will use the lemmas from the previous section to establish the optimality of policies $d \in D^L$. First, we need to obtain a form for $B_d(g,h)$ specific for our model. It is straightforward, but it requires some precision and tenacity, to use (1), (2) and (3) to find the following expression for the kth element of $B_d(g,h)$ as defined in (4):

$$\begin{aligned}(B_d(g,h))(k) = &- g + \hat{c}^{-1}\Bigg(-\nu_k\left(h(k) - h(k-1)\right) \\ &- \sum_{j\in J\setminus d(k)} p_j m_j + \sum_{j\in d(k)} m_j\left(h(k+1) - h(k)\right)\Bigg).\end{aligned} \tag{7}$$

We have introduced the variables $h(-1) = 0$ and $h(S+1) = 0$ for convenience of notation, which necessarily have a pre-factor 0 since $d(S) = \emptyset$ and $\nu_0 = 0$. In the following lemma, we show that the Markov reward process induced by a locally optimal critical level policy has a bias with a certain structure.

Lemma 4. *Suppose $d \in D^L$. Let g_d and h_d be the gain and bias associated with d. Let $j \in J$ with associated critical level c_j be given.*

i) Suppose $c_j \neq S$. Then $h_d(S - c_j) - h_d(S - c_j - 1) \geq -p_j$.

ii) Suppose $c_j \neq 0$. Then $h_d(S - c_j + 1) - h_d(S - c_j) \leq -p_j$.

Proof. For *i)*, suppose first that d can be modified by increasing c_j by 1 without violating monotonicity. Call this modified policy d'. It differs from d only by a unit increase of c_j. d' thus only differs from d because it rejects demands of class j in state $S - c_j - 1$ instead of accepting them, viz,

$$(d'(0), \dots, d'(S - c_j - 1), \dots, d'(S)) = (d(0), \dots, d(S - c_j - 1) \setminus \{j\}, \dots, d(S)).$$

Using this observation, we can use (7) to show that

$$B_{d'}(g,h) = B_d(g,h) - \hat{e}_{S-c_j-1}\hat{c}^{-1} m_j (h(S - c_j) - h(S - c_j - 1) + p_j) \tag{8}$$

where $\hat{e}_{S-c_j-1}$ is the vector with 1 as its $(S - c_j - 1)$th entry, and zero for all other entries. Now, we apply Lemma 2, and in the second equality we use (8) and Lemma 1.

$$\begin{aligned} g_{d'}e &= g_d e + P^*_{d'} B_{d'}(g_d, h_d) \\ &= g_d e - P^*_{d'}(\hat{e}_{S-c_j-1}\hat{c}^{-1} m_j (h_d(S - c_j) - h_d(S - c_j - 1) + p_j)). \end{aligned}$$

Referring to the discussion regarding the limiting matrix P^*_d in Section 3 we conclude that

$$g_{d'}e = \left(g_d - p^*_{d'}(S - c_j - 1)\hat{c}^{-1} m_j (h_d(S - c_j) - h_d(S - c_j - 1) + p_j)\right) e. \tag{9}$$

$p^*_{d'}(S - c_j - 1)$ denotes the long term average fraction of time spent in state $S - c_j - 1$. It is strictly positive because demands for class j are accepted in class 0 trough $S - c_j - 2$ under policy d', from which we infer that $S - c_j - 1$ is recurrent. $m_j > 0$ by assumption. Since d' differs from d only in the unit decrease of a single critical level, we have $g_d - g_{d'} \geq 0$ by

$d \in D^L$. From (9), $h(S-c_j)-h(S-c_j-1)+p_j$ must be non-negative as well, from which the result immediately follows.

Now suppose that increasing c_j violates monotonicity. Then, let j' be the demand class with the least penalty cost, for which $c_{j'} = c_j$. It is easy to verify from the definitions that the critical level $c_{j'}$ can be increased without violating monotonicity. Now, apply the argument above for j'. We find that

$$h(S-c_{j'})-h(S-c_{j'}-1)+p_{j'} \geq 0$$

which directly implies the result since $p_{j'} \leq p_j$ and $c_{j'} = c_j$ by hypothesis.

The proof of *ii)* is similar. Suppose d' can be constructed from d by a unit decrease of c_j without violating monotonicity. Then d' differs from d because it accepts demands for class j in state $S-c_j$ instead of declining them. Thus

$$B_{d'}(g,h) = B_d(g,h) + \hat{e}_{S-c_j}\hat{c}^{-1}m_j(h(S-c_j+1)-h(S-c_j)+p_j).$$

Similarly as before

$$g_{d'}e = g_d e + p^*_{d'}(S-c_j)\hat{c}^{-1}m_j(h(S-c_j+1)-h(S-c_j)+p_j)e$$

from which the result follows readily. Suppose now that c_j cannot be decreased without violating monotonicity. Then, let j' be the demand class with the highest penalty cost, for which $c_{j'} = c_j$. c_j' can be increased without violating monotonicity, and we can proceed as before to conclude that the result continues to hold. □

Lemma 4 can be intuitively understood by using the interpretation of $h_d(k)-h_d(k-1)$ as the comparative advantage of being in state k instead of being in state $k-1$ under policy d.

In the following lemma, we prove that the bias of the Markov reward process induced by a locally optimal policy is concave and strictly decreasing in the number of outstanding orders.

Lemma 5. *Suppose $d \in D^L$. Let g_d and h_d be the gain and bias associated with d. Then*

i) For $k \in \{0,\dots,S-1\}$

$$h_d(k+1)-h_d(k) < 0$$

ii) For $k \in \{1,\dots,S-1\}$

$$h_d(k+1)-h_d(k) \leq h_d(k)-h_d(k-1)$$

Proof. We start by proving *i)* for $k=0$. From the definition of the critical levels we must either have a critical level c_j for which $S-c_j=0$, or all demands are accepted in state 0. In the first case, we apply *ii)* of Lemma 4 to conclude that $h_d(1)-h_d(0) \leq -p_j < 0$. In the latter case we note first that g_d and h_d solve (5) by Lemma 1, which implies that

$$0 = (B_d(g_d,h_d))(0).$$

By using (7) and by noting that $d(0) = J$ for the case under consideration this implies that

$$0 = \hat{c}^{-1} \sum_{j \in J} m_j \left(h_d(1) - h_d(0)\right) - g_d.$$

It is easy to see that under any policy there must be at least one recurrent state in which demands are declined. Therefore, g_d is strictly negative. Furthermore, $\hat{c} > 0$, $|J| \geq 1$ and $m_j > 0$. The result follows.

We now prove *ii)* for $k = 1$ (suppose $S > 0$). From the definition of the critical levels we either have a critical level c_j for which $S - c_j = 1$, or all demand classes accepted in state 0 are also accepted in state 1 and vice versa. The result immediately follows by combining *i)* and *ii)* of Lemma 4 in the former case. In the latter case we use again that g_d and h_d solve (5), from which it follows that

$$0 = (B_d(g_d, h_d))(1) - (B_d(g_d, h_d))(0).$$

since both terms on the right hand side are zero. Using $d(1) = d(0)$ for the case we are considering and (7) we find that this implies that

$$\hat{c}^{-1} \sum_{j \in d(0)} m_j \left(h_d(2) - 2h_d(1) + h_d(0)\right) = \hat{c}^{-1} \nu_1 \left(h_d(1) - h_d(0)\right).$$

The right hand side is strictly negative by *i)* for $k = 0$. Clearly, $d(0) = \emptyset$ contradicts negativity of the right hand side. We conclude that $d(0) \neq \emptyset$, and the result follows.

We now proceed by induction. Note that *i)* for k follows from *ii)* for k and *i)* for $k - 1$. To complete our inductive argument, it thus suffices to show that *ii)* for $k \in \{1, \ldots, S - 1\}$ follows from *i)* and *ii)* for $k - 1$.

Again, we either have a critical level c_j for which $S - c_j = k$, or the demands accepted in state k are also accepted in state $k - 1$ and vice versa. In the former case, the result follows immediately by combining *i)* and *ii)* of Lemma 4, so we do not need the induction hypothesis in this case. In the latter case, we have $d(k) = d(k - 1)$. Again

$$0 = (B_d(g_d, h_d))(k) - (B_d(g_d, h_d))(k - 1)$$

which holds by Lemma 1, implies for $k \in \{1, \ldots, S - 1\}$ that

$$\begin{aligned} &\sum_{j \in d(k)} m_j \left(h_d(k+1) - 2h_d(k) + h_d(k-1)\right) \\ &\qquad = \nu_k \left(h_d(k) - h_d(k-1)\right) - \nu_{k-1} \left(h_d(k-1) - h_d(k-2)\right). \end{aligned} \tag{10}$$

The right hand side of this equation can be shown to be equal to

$$\nu_{k-1} \left(h_d(k) - 2h_d(k-1) + h_d(k-2)\right) + \left(\nu_k - \nu_{k-1}\right) \left(h_d(k) - h_d(k-1)\right)$$

The first term is not positive by the induction hypothesis *ii)* for $k-1$, and the second term is strictly negative by induction hypothesis *i)* for $k-1$ and by $\nu_k - \nu_{k-1} > 0$. So, $d(k) = \emptyset$ leads to a contradiction, and we conclude that $d(k) \neq \emptyset$ and $h_d(k+1) - 2h_d(k) + h_d(k-1) \leq 0$. By induction, the result follows. □

In the following lemma, we use the results derived in the previous two lemmas to show that a policy d that is of locally optimal critical level type satisfies the optimality equations. Therefore, it is also globally optimal.

Lemma 6. *Let $d \in D^L$. Then d is an optimal policy, and the average expected reward associated with d is the optimal reward.*

Proof. Let g_d and h_d denote the average expected reward and bias of the Markov reward process induced by d. The hypotheses of Lemmas 4 and 5 are satisfied for h_d. To show that the hypothesis of Lemma 3 is satisfied we need to show that

$$\max_{d' \in D^{\mathrm{MD}}} B_{d'}(g_d, h_d)$$

equals the 0-vector. Since g_d and h_d satisfy (5) by Lemma 1, it is equivalent to show that for each $k \in \{0, \ldots, S\}$ the following expression

$$\max_{d' \in D^{\mathrm{MD}}} (B_{d'}(g_d, h_d))(k) - (B_d(g_d, h_d))(k) \tag{11}$$

equals 0. For $k = S$, this holds trivially since A_S only consists of one element ($\emptyset$), reflecting that all demands are necessarily lost in state S. Now consider the case $k < S$. Using (7) and remembering that D^{MD} is the Cartesian product of the decision sets A_k for the different states, it is straightforward to show that (11) is equivalent to

$$\max_{d'(k) \in A_k} \left(\sum_{j \in d'(k) \cap (J \setminus d(k))} m_j \left(h(k+1) - h(k) + p_j\right) \right.$$
$$\left. - \sum_{j \in (J \setminus d'(k)) \cap d(k)} m_j \left(h(k+1) - h(k) + p_j\right) \right). \tag{12}$$

where equal terms were cancelled. Note that $d'(k) \cap (J \setminus d(k))$ denotes the demands that are accepted under d' but declined under d in state k.

Take now an arbitrary demand class $j \in J \setminus d(k)$ that is declined under d in state k. We will show that $h(k+1) - h(k) + p_j$ is non-positive. d is of critical level type, so by definition 1 there exists a critical level c_j for demand class j. Since j is declined under d in state k, it is a matter of checking this definition to establish that the critical level c_j for j satisfies $S - c_j \leq k$. Note that this implies that $S - c_j \leq S - 1$. We thus can apply *ii)* of Lemma 4 to conclude that

$$h_d(S - c_j + 1) - h_d(S - c_j) \leq -p_j.$$

By applying *ii)* of Lemma 5 repeatedly and by using that $S - c_j \leq k$ we conclude that

$$h_d(k+1) - h(k) \leq h_d(S - c_j + 1) - h_d(S - c_j).$$

Combining the above equations yields the result. The first term in (12) is thus non-positive.

Take now an arbitrary demand $j \in d(k)$. It can be shown in a very similar manner as above that $h(k+1) - h(k) + p_j$ is nonnegative. c_j now satisfies $S - c_j > k$, implying $S - c_j > 0$. We then apply *i)* of Lemma 4, and continue as before.

When including the minus sign, the second term in (12) is thus non-positive as well. Therefore, the maximum is bounded from above by 0. Now, note that $d'(k) = d(k)$ attains the bound, from which we conclude that the maximum equals 0. We conclude that the hypothesis of Lemma 3 is satisfied. The result now immediately follows. □

We are now ready to prove the optimality of the algorithms proposed by Kranenburg and van Houtum (2007). For ease of reference, we restate the algorithms here, adapted where needed to our notation and the fact that we have used a reward model to align with Puterman (1994, Chapter 8). Kranenburg et al. show that it is never optimal to decline the most critical demand classes, which will be denoted by $\{1, \dots, j^c\}$ where $j^c = \max\{j \in J | p_1 = p_j\}$. The proposed algorithms are

> Algorithm 1. Keep c_j, $j \in J$, $j \leq j^c$ always fixed at 0. Start with an arbitrary choice for c_j, $j \in J$, $j > j^c$, that satisfies monotonicity. Define the neighborhood as all policies that still satisfy the monotonicity constraint and that have critical levels that differ at most one from the corresponding critical levels in the original policy. If the reward of the cheapest neighbor is strictly larger than the reward of the current solution, then select this neighbor and set this policy as the current solution, and repeat the process of evaluating all neighbors for this new policy. Otherwise, stop and take the current solution as the solution found by the algorithm.
>
> Algorithm 2. Keep c_j, $j \in J$, $j \leq j^c$ always fixed at 0. Start with an arbitrary choice for c_j, $j \in J$, $j > j^c$, that satisfies monotonicity. For $j = |J|$, find $c_j \in \{c_{j-1}, \dots, c_{j+1}\}$ with the highest reward, at fixed values of the other critical levels, and change c_i accordingly (define $c_{|J|+1} = S$). When the reward for the current solution ties with the best alternative, keep the current solution. Repeat this optimization for one critical level at a time for $j = |J| - 1$ down to $j^c + 1$. After that, optimize again for $j = |J|$. Continue this iterative process until for none of the j-values ($> j^c$) a strict improvement is found. This is the solution found by the algorithm.

The following theorem establishes the optimality of Algorithms 1 and 2.

Theorem 1. *Algorithms 1 and 2 converge in a finite number of steps. When they terminate, the final solution is optimal among the class of Markovian deterministic policies in general, and in particular among the class of critical level policies.*

Proof. We show that the policy found upon termination of the above algorithms belongs to D^L. Then Lemma 6 guarantees optimality of this policy. A policy d^t found upon termination of either of these algorithms is clearly of critical level type. Also, for both algorithms, decreasing or increasing a single critical level for a demand class $j > j^c$ does not increase the average expected reward because this would contradict the termination of the algorithm.

In order for d^t to belong to D^L, it remains to check that a unit increase in the critical level c_{j^c} associated with j^c decreases the expected reward. But this is precisely what is shown for any policy in Kranenburg and van Houtum (2007, Lemma 2) in order to establish that the optimal critical levels for demand classes $j \leq j^c$ are 0, which motivated them to keep these critical levels fixed at 0 in the first place. We conclude that $d^t \in D^L$. The final solution is thus optimal. To conclude that the algorithms converge in a finite number of steps, note that a solution that was visited cannot be visited again because that would contradict that the rewards are strictly increasing. Because there are only a finite number of critical level combinations, the algorithms must converge in a finite number of steps. □

Note that Lemma 6 can serve as the basis to define other local search based algorithms which are guaranteed to be optimal. We could for instance adapt Algorithm 2 by decreasing the neighborhood to unit increases or decreases in the critical levels.

The following corollary is interesting in our opinion because of the manner in which it is proven.

Corollary 1. *A monotone critical level policy is optimal for the problem we consider. For the most critical demand classes $j \leq j^c$ the optimal critical level is equal to* 0*.*

Proof. The result follows immediately from Theorem 1, and the fact that Markovian deterministic policies dominate in the model. □

By Kranenburg's (2007) observation with respect to the holding cost, early work by Miller (1969) becomes applicable for this model. Miller considers a queueing system with n servers with equal, exponential service rate and controlled admissions. The reward incurred differs across different customers, which arrive following a Poisson process. His objective is to maximize the long term average reward. Depending on the number of servers that are occupied, the gatekeeper may decide to reject customers to save capacity for more critical customers. Because Kranenburg and van Houtum show the holding costs can be assumed to be fixed for fixed S, it is not hard to see that Miller's model is equivalent to the model considered here.

In terms of the model considered here, Miller shows that critical levels are optimal (even though he does not use the concept of critical level policies), and that demands of the highest criticality are always accepted. This result differs from the result derived by Ha (1997), e.g. because Ha's model assumes a make-to-stock environment, more general holding costs and it includes discounted models.

5 Extensions

General leadtimes

Our model assumes i.i.d. exponential leadtimes. Most results obtained in this paper can be extended to the case of generally distributed i.i.d. leadtimes considered by Kranenburg and van Houtum (2007), as long as we restrict the decision to accept or reject demands to depend only upon the criticality of the demand and the number of parts on stock (Note that Kranenburg and van Houtum (2007) assume that the control of the system is of critical level type, which imposes an even stronger restriction). The steady state distribution of outstanding orders and consequently the long term expected reward of such a policy do not depend upon the distribution of the leadtime. This can for instance be shown by a queueing theory argument of the type that is employed in Kranenburg and van Houtum (2007), or by the arguments employed in Dekker et al. (2002). Therefore, a policy that is optimal in the exponential case is also optimal for the general leadtime case, but only within this restricted class of policies. Therefore, the algorithms continue to find the optimal critical level policy among the class of critical level policies.

Note that imposing the control to depend only upon the number of outstanding orders is a true restriction for general leadtimes, as information about outstanding orders may improve the quality of stock control. Ha (2000) delves deeper into this question by considering the optimal control for Erlang distributed production times in a make-to-stock environment. Because of the special properties of this distribution, the size of the state spaces remains manageable. Teunter and Klein Haneveld (2008) consider general leadtimes in an (s, Q) system. The complexity of the analysis is kept within bounds by using the approximative assumption that only the costs up until the arrival of the next replenishment order are relevant.

Dependent leadtimes

Before, we have assumed i.i.d. exponentially distributed leadtimes. This is equivalent to stating that the orders are served in a queue with S identical servers with rate L^{-1}. The problem of inventory rationing however also arises in other settings. Make-to-stock, equivalent with a single server queue, is assessed by Ha (1997). Other examples include queues with a number of servers larger than 1, but smaller than S.

Before, we had $\nu_k = L^{-1}k$. We now assume general $\nu_k > 0$, but such that $\nu_{k+1} \geq \nu_k$. This includes the examples mentioned above. The reader can verify that the only properties of ν_k that were used until Lemma 6 were the properties $\nu_k > 0$ (for instance, to establish that the model is unichain), and $\nu_{k+1} > \nu_k$ (in the inductive argument in the proof of Lemma 5). It requires only minor modification to this proof to allow for $\nu_{k+1} = \nu_k$.

Lemma 7. *The results stated in Lemma 5 remain valid for general ν_k, as long as $\nu_{k+1} \geq \nu_k$ and $\nu_k > 0$.*

Proof. All results, except the last inductive argument, remain valid without modification. In the last inductive argument, a possible issue occurs when $\nu_k = \nu_{k-1}$; we can no longer conclude strict positivity of the right hand side of (10), only non-negativity remains. Note that this still suffices to establish the required result in case $d(k) \neq \emptyset$. However, $d(k) = \emptyset$ no longer leads to contradiction.

Therefore, we consider the case $d(k) = \emptyset$ separately. Note that this implies that $d(k+1) = \emptyset$ as well. From Lemma 1 we have

$$0 =(B_d(g_d, h_d))(k+1) - (B_d(g_d, h_d))(k)$$

from which it follows that

$$0 = \nu_{k+1}\left(h_d(k+1) - h_d(k)\right) - \nu_k\left(h_d(k) - h_d(k-1)\right).$$

The result immediately follows since $\nu_{k+1} \geq \nu_k$ and $h_d(k) - h_d(k-1)$ is negative by the induction hypothesis. □

Thus, under the assumptions in this section, Lemmas 4, 5, 6 remain valid. Theorem 1 and its corollary remain valid, except that Kranenburg and van Houtum's Lemma 2 no longer holds. We thus need to consider changing the critical levels for the most critical demand classes in the search algorithms, and we can no longer keep them fixed at 0.

Note furthermore, that we implicitly assume that the holding cost does not depend on the rationing decision for fixed S. For the original model, Kranenburg and van Houtum's observation ensures that this assumption can be made without severe restrictions. Their observation is however not valid for the extended model, and assuming fixed holding costs for fixed S is more restrictive in those cases. It is valid in practical situations in case the holding costs are also incurred for parts that are in on order, for instance for repairable components and other closed loop supply chains.

6 Conclusions

We established optimality of 2 of the 3 algorithms proposed by Kranenburg and van Houtum (2007). We strengthened this result to include resupply conditions other than the one considered by Kranenburg and van Houtum. In the process, we recovered the result by Miller (1969), strengthening it by allowing for more general resupply assumptions.

References

de Véricourt, F., Karaesmen, F. and Dallery, Y.: 2002, Optimal stock allocation for a capacitated supply system, *Management Science* **48**, 1486–1501.

Dekker, R., Hill, R., Kleijn, M. and Teunter, R.: 2002, On the $(S-1, S)$ lost sales inventory model with priority demand classes, *Naval research logistics* **49**, 593–610.

Deshpande, V., Cohen, M. and Donohue, K.: 2003, A threshold inventory rationing policy for service-differentiated demand classes, *Management Science* **49**, 683–703.

Ha, A. Y.: 1997, Inventory rationing in a make-to-stock production system with several demand classes and lost sales, *Management Science* **43**, 1093–1103.

Ha, A. Y.: 2000, Stock rationing in an $M/E_k/1$ make to stock queue, *Management Science* **46**, 77–87.

Kranenburg, A. and van Houtum, G.: 2007, Cost optimization in the $(S-1, S)$ lost sales inventory model with multiple demand classes, *Operations research letters* **35**, 493–502.

Melchiors, P., Dekker, R. and Kleijn, M. J.: 2000, Inventory rationing in an (s, Q) inventory model with lost sales and two demand classes, *Journal of the operational research society* **51**, 111–122.

Miller, B.: 1969, A queueing reward system with several customer classes, *Management science* **16**, 234–245.

Puterman, M. L.: 1994, *Markov decision processes, discrete stochastic dynamic programming*, John Wiley and Sons, Inc. , New York, NY, USA.

Teunter, R. H. and Klein Haneveld, W. K.: 2008, Dynamic inventory rationing strategies for inventory systems with two demand classes, Poisson demand and backordering, *European journal of operational research* **190**, 156–178.

Topkis, D. M.: 1968, Optimal ordering and rationing policies in a nonstationary dynamic inventory model with n demand classes, *Management Science* **15**, 160–176.

Veinott, A. F.: 1965, Optimal policy in a dynamic, single product, non-stationary inventory model with several demand classes, *Operations research* **13**, 761–778.