An Adaptive Algorithm for Finding the Optimal Base-Stock Policy in Lost Sales Inventory Systems with Censored Demand

Woonghee Tim Huh
Department of Industrial Engineering and Operations Research, Columbia University, New York, NY 10027
email: huh@ieor.columbia.edu http://www.columbia.edu/~th2113/

Ganesh Janakiraman
IOMS-OM Group, Stern School of Business, New York University, 44 W. 4th Street, Room 8-71, New York, NY 10012-1126
email: gjanakir@stern.nyu.edu http://pages.stern.nyu.edu/~gjanakir/

John A. Muckstadt
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853
email: jack@orie.cornell.edu http://people.orie.cornell.edu/~jack/

Paat Rusmevichientong
School of Operations Research and Information Engineering, Cornell University, Ithaca, NY 14853
email: paatrus@cornell.edu http://legacy.orie.cornell.edu/~paatrus/

We consider a periodic-review single-location single-product inventory system with lost sales and positive replenishment lead times. It is well known that the optimal policy does not possess a simple structure. Motivated by recent results showing that base-stock policies perform well in these systems, we study the problem of finding the best base-stock policy in such a system. In contrast to the classical inventory literature, we assume that the manager does not know the demand distribution *a priori*, but must make the replenishment decision in each period based only on the past sales (censored demand) data. We develop a nonparametric adaptive algorithm that generates a sequence of order-up-to levels whose T-period running average of the inventory holding and lost sales penalty cost converges to the cost of the optimal base-stock policy at the rate of $O(1/T^{1/3})$. Our analysis is based on recent advances in stochastic online convex optimization and on the uniform ergodicity of Markov chains associated with bases-stock policies.

Key words: Base-Stock Policy; Censored Demand; Lost Sales Inventory System; Online Optimization

MSC2000 Subject Classification: Primary: 90B05, 62N01; Secondary: 68T05, 60J20

OR/MS subject classification: Primary:Inventory/Production:Uncertainty:Stochastic; Secondary: Analysis of Algorithms, Statistics:Nonparametric

---

1. Introduction We study the problem of managing a periodically reviewed inventory system with the following features. Inventory is replenished from a supplier with ample supply, where the replenishment lead time is deterministic and is an integer multiple of the review period. Any demand that cannot be satisfied immediately with the on-hand inventory leads to *lost sales* while any excess inventory at the end of a period is carried over to the next period. At the end of each period, either inventory holding cost of lost sales cost is incurred, and is proportional to the amount of lost sales or on-hand carry-over inventory. The manager wants to minimize the long-run average cost per period.

Assume demands in different periods are independently and identically distributed. However, contrary to the classical inventory literature, the common distribution of demand is not known to the manager *a priori*. In each period, only sales are known, but not demand. Since sales are strictly smaller than demand if demand exceeds the available supply, the demand information is *censored*.

Even when the demand distribution is known, it is well known that the optimal policy for this problem does not possess any simple structure [11], and is difficult to compute when the lead time is long. For this problem, the class of base-stock policies, though not optimal, are known to perform well, especially when the ratio of the lost sales cost parameter to the holding cost parameter is high [7]. We use as a benchmark the long-run average cost of the best base-stock policy, which could be computed if the demand distribution were known. In this paper, we provide an algorithm for computing a base-stock level in each period under the condition of the unknown demand distribution and censored demand information, and show that the average cost of using this algorithm over T periods converges to the benchmark at the rate of $1/\sqrt[3]{T}$.

1.1 Connections to the Literature We first discuss papers that study the lost sales inventory problem under the assumption that the demand distribution is known. Morton [16] and Karlin and Scarf [11] study the dynamic program and establishes that the optimal ordering quantity is a decreasing function of the on-hand and on-order inventory vector with the rate of decrease at most 1. Zipkin [22] presents a new derivation of this result and extends it to more general settings, for example, allowing capacity restrictions. While it is possible to determine the optimal replenishment policy via dynamic programming, the size of the state space increases exponentially with the lead time, making the approach intractable even for problems with reasonably short lead times. As a result, various heuristics have been proposed; however, it is unclear which algorithm, if any, performs better than the others in general. A recent paper by Zipkin [21] contains a numerical comparison of several inventory policies, such as the myopic policy of Morton [15], the base-stock policy, the dual-balancing policy of Levi et al. [13], the constant-order policy of Reiman [19], and their variants.

Recently, Huh et al. [7] show the asymptotic optimality of the base-stock policies. As the ratio of the unit penalty cost to the unit holding cost increases to infinity, they prove, under mild technical conditions, that the ratio of the cost of the best base-stock policy to the optimal cost converges to 1. Since the penalty cost is typically much larger than the holding cost (with the ratio exceeding 200 in many applications), it is reasonable to expect that the best base-stock policy performs well compared to the optimal policy. This hypothesis is confirmed by computational results by Huh et al. [7] and Zipkin [21]. In fact, when the ratio between the ratio of the lost sales penalty and the holding cost is 100, the cost of the best base-stock policy is typically within 1.5% of the optimal cost. Although base-stock policies have been shown to perform reasonably well in lost sales systems, finding the best base-stock policy, in general, cannot be accomplished analytically, and involves simulation optimization techniques.

Whereas the demand distribution is assumed to be known to the manager *a priori* in the classical lost sales inventory literature, in many applications, however, the manager does not know the underlying demand distribution, and must make the ordering decision in each period based on the historical data. Since unsatisfied demand is immediately lost, the data available to the manager often consists of historical sales data, corresponding to the smaller of the beginning on-hand inventory level and the demand realization for that period. The demand data is thus *censored*.

The first contribution of our paper is to develop an adaptive algorithm with a provable performance guarantee. It generates a sequence of order-up-to levels $\{S_t : t \geq 1\}$ such that the order-up-to level S_t in period t depends only on the sales data observed in the previous $t-1$ periods. The T-period running average expected cost under this algorithm converges to the cost of the best base-stock policy. We also establish the rate of convergence, showing that the average expected cost after T periods differs from the cost of the best base-stock policy by at most $O\left(1/T^{1/3}\right)$.

There exist a number of adaptive methods for the lost sales system with censored demand, but all of them address only the case of *zero replenishment lead-time*. Burnetas and Smith [2] propose a stochastic approximation method for estimating the newsvendor quantile. Godfrey and Powell [5] and Powell et al. [18] develop a method of iteratively approximating the convex objective function with piece-wise linear functions. Huh and Rusmevichientong [9] apply stochastic online convex optimization to this problem; in their setting, the adaptive control problem is much easier because the Markov chain is independent of the starting state, and one can obtain an unbiased derivative estimator in each period.

While the above adaptive methods are nonparametric, Nahmias [17] and Agrawal and Smith [1] consider Bayesian settings, and use censored historical data to estimate the parameters of the normal and negative binomial distributions, respectively. All of the papers mentioned here only consider the case of zero lead time. When replenishment is instantaneous, the lost sales model turns out to be analytically equivalent to the backorder system, and the best base-stock level is the newsvendor quantile of the demand distribution. When lead times are positive, however, the problem is much more difficult and there is no explicit formula that describes the optimal base-stock level. To the best of our knowledge, our result represents the first adaptive algorithm for finding the best base-stock policy in lost sales inventory systems with positive replenishment lead times.

The second contribution of our paper is the analysis of the long-run average cost under a base-stock policy. It is well known that the stochastic process that tracks the on-hand and on-order inventories under any base-stock policy forms a Markov chain. The Markov chain, however, may not be *ergodic*,

that is, it may not have a stationary distribution. We provide a sufficient condition on the base-stock level that ensures that the distribution of the on-hand inventory under the base-stock policy converges to a stationary distribution, and furthermore establish the rate of convergence. The ergodicity result simplifies the expression for the long-run average cost, leads to new insights about the structure of the cost functions under base-stock policy, and provides a foundation for our adaptive algorithm. We believe the sufficient condition for the ergodicity represents the first such results for Markov chains associated with order-up-to policies in a stochastic inventory system despite the extensive literature in this area. Our analysis is based on the uniform ergodicity of Markov chains. We expect a similar analysis to be applicable to other inventory systems.

The third contribution of the paper is to provide a framework for applying an adaptive algorithm to a stochastic system where the performance measure depends on its stationary distribution. In these systems, it is often not possible to obtain the gradient of the objective function or its unbiased estimate. The bias of the estimate often depends on how long the system has been running. As a result, an adaptive algorithm needs to balance the benefit of smaller bias by continuing to implement the current decision, and the benefit of switching quickly to a potentially better decision. We believe that the adaptive method developed in this paper can be useful in other stochastic systems provided that the convergence rate to the stationary distribution can be established uniformly for any choice of decision variables.

1.2 Organization The remainder of the paper is organized as follows. In Section 2, we formally describe the inventory control problem with lost sales and positive lead times. In Section 3, we consider the long-run average cost under any base-stock policy and establish a sufficient condition that guarantees the distribution of the on-hand inventory converges to its stationary distribution. We also establish the rate of convergence. Then, we consider the problem of estimating the long-run average cost and its derivative using censored demand samples. We establish bounds on the bias of the sample-based estimates for the objective function and its derivative. Based on the findings in Section 3, we present the main result of the paper in Section 4, where we develop an adaptive algorithm and establish a provable performance bound for the algorithm.

2. Problem Formulation and Model Description Let $t \in \{1, 2, \ldots\}$ represent the time period, which is indexed forward. The demand in period t is denoted by D_t, and we assume that the demands over time $\{D_1, D_2, \ldots\}$ are independent and identically distributed random variables. We will denote by D the generic demand random variable having the same distribution as D_t. We assume that D is nonnegative satisfying $E[D] > 0$. Let $\mu = E[D]$. Let F denote the cumulative distribution function of D. Throughout the paper, we will assume that D is a continuous random variable. Let $\tau \geq 1$ denote the replenishment lead time. Given a replenishment policy π, we denote by $Q_t(\pi)$ the quantity ordered in period t, which arrives at the beginning of period $t + \tau$. Let $Q_{-\tau+1}(\pi), Q_{-\tau+2}(\pi), \ldots, Q_0(\pi)$ be the amounts of delivery scheduled to arrive in periods $1, 2, \ldots, \tau$, respectively. Furthermore, let $I_t(\pi)$ denote the after-delivery on-hand inventory level in period t under the replenishment policy π.

For any replenishment policy π, we assume that events in period $t \geq 1$ occur in the following order. At the beginning of each period, the delivery of $Q_{t-\tau}(\pi)$ units arrives, which were ordered in period $t - \tau$. The manager observes the outstanding procurement orders $(Q_{t-1}(\pi), Q_{t-2}(\pi), \ldots, Q_{t-\tau+1}(\pi))$ and the on-hand inventory $I_t(\pi)$. Let

$$X_t(\pi) \quad = \quad (Q_{t-1}(\pi), Q_{t-2}(\pi), \ldots, Q_{t-\tau+1}(\pi), I_t(\pi))$$

be the *inventory vector* associated with policy π. Note that each $X_t(\pi)$ is a τ-dimensional vector. In particular, we call $X_1(\pi) = (Q_0(\pi), Q_{-1}(\pi), \ldots, Q_{-\tau+2}(\pi), I_1(\pi))$ is the *initial inventory vector*, which is independent of π. The manager places an order of $Q_t(\pi) \geq 0$ units. Then, demand D_t is realized. The manager does *not* observe the realized demand, but observes the sales quantity $\min\{D_t,\ I_t(\pi)\}$ only.

At the end of each period, the holding cost of $\$h$ per unit is charged on excess inventory, and the lost sales penalty cost of $\$b$ per unit is charged on excess demand. Given the on-hand inventory $I_t(\pi)$, the expected cost in period t is given by $C\left(I_t(\pi)\right)$, where

$$C\left(y\right) \quad = \quad h \cdot E\left[y - D_t\right]^+ + b \cdot E\left[D_t - y\right]^+ \ , \tag{1}$$

where the expectation is taken with respect to the demand D_t in period t under the replenishment policy π. (The manager does not observe the total lost sales penalty cost, but this cost has nonetheless been

incurred.) The on-hand inventory level in the next period, $I_{t+1}(\pi)$, is the sum of the carry-over inventory and the delivery due that period; thus, it is given by the following recursion:

$$I_{t+1}(\pi) \quad = \quad [I_t(\pi) - D_t]^+ \quad + \quad Q_{t-\tau+1}(\pi) \ .$$

We wish to find the replenishment policy that minimizes the total long-run average expected holding cost and lost sales penalty, that is,

$$\inf_{\pi} \left\{ \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[C\left(I_t(\pi)\right)\right] \right\} \ ,$$

where the expectation is taken with respect to the on-hand inventory level $I_t(\pi)$.

As indicated in the introduction, we will restrict our attention to the class of *base-stock policies*. Let $S \geq 0$. Under the order-up-to-S policy, if the inventory position (inventory on hand plus on order) in each period is less than S, we place an order to bring the inventory position to S. If the inventory position exceeds S, however, we do not place any order. Let $X_t(S)$, $I_t(S)$, and $Q_t(S)$ denote the inventory vector, the on-hand inventory, and the order quantity in period t under the order-up-to-S policy, respectively. Thus,

$$Q_t(S) \quad = \quad [S - X_t(S) \cdot \mathbf{1}^{\tau}]^+ \, ,$$

where $\mathbf{1}^{\tau} = (1, 1, \ldots, 1)$ denotes a vector of length τ.

The adaptive algorithm that we propose in this paper is a *period-dependent* base-stock policy. It generates a sequence of order-up-to levels $\phi = \{S_t : t \geq 1\}$ such that the order-up-to level S_t in period t depends only on the sales data observed in the previous $t-1$ periods. The T-period average expected cost under the constructed policy ϕ converges to the cost of the best base-stock policy, i.e.,

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[C\left(I_t(\phi)\right)\right] \quad = \quad \inf_{S} \left\{ \limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} E\left[C\left(I_t(S)\right)\right] \right\} .$$

3. Long-Run Average Costs Under a Base-Stock Policy In this section, we study properties of the Markov chain associated with a base-stock policy, and provide a characterization of the long-run average cost. When the Markov chain is *ergodic*, the expected holding and backorder costs in period t converges the long-run average cost as the time period t increases to infinity. In Section 4, we use this fact to generate a sequence of base-stock levels whose expected time-average cost approaches the cost of the best base-stock policy.

However, the Markov chain associated with a base-stock level S may not be ergodic, if S is too small relative to the size of the demand in each period. For instance, if the demand in each period exceeds S, then a stockout occurs in every period, which causes the after-delivery on-hand inventory in each period t to be exactly the same as the amount ordered in period $t-\tau$. Thus, the on-hand inventory follows a cyclic pattern and is *not* ergodic (a more detailed example of non-ergodicty is given at the end of Section 3.2). In this section, we provide a sufficient condition that guarantees ergodicity of the Markov chain associated with an order-up-to policy, characterize the rate of convergence, and provide bounds on the expected estimation error on both the cost function and its derivative. We believe that the results in this section are of independent interest to the stochastic inventory theory literature.

Instead of working with the Markov chain $\{X_t(S) : t \geq 1\}$ associated with the inventory vectors under order-up-to-S policy, it is more convenient to *augment* the Markov chain such that the state in each period also includes the *sample derivatives* of the inventory vector with respect to S. In Section 3.1, we study the derivatives of both the on-hand inventory level $I_t(S)$ and the order quantity $Q_t(S)$ with respect to the order-up-to level S, and develop recursive formulae that define the stochastic processes $\{I'_t(S) : t \geq 1\}$ and $\{Q'_t(S) : t \geq 1\}$.

In Section 3.2, we establish a sufficient condition on the order-up-to level S that guarantees that the augmented Markov chain associated with order-up-to-S policy is ergodic. When this condition is satisfied, the augmented Markov chain converges to a stationary random vector. We also establish an upper bound on the rate of convergence (Theorem 3), and provide an example when ergodicity fails.

Based on the ergodicity of the augmented Markov chain associated with order-up-to policies, Theorem 4 in Section 3.3 characterizes the long-run average cost for any base-stock level, regardless of whether

the condition for ergodicity holds. This characterization becomes useful for developing our adaptive algorithm in Section 4. Furthermore, in Section 3.4, we establish error bounds associated with the finite sample-based cost function as well as its derivative.

We remark that, in Section 3, we study the Markov chain stochastic process associated with the lost-sales inventory system under a fixed base-stock policy. The results in these sections stand alone without any reference to the adaptive algorithm of Section 4.

3.1 Sample Derivatives of the On-Hand Inventory Under a Base-Stock Policy We provide an expression for the sample derivatives of the on-hand inventory and order quantity in each period, which will be used in the development of our adaptive algorithm. For any base-stock level $S \geq 0$ and the initial inventory vector $x_1 \in \Re_+^\tau$, let the random variable $V(S, x_1)$ denote the first time that the total inventory position is less than or equal to S, assuming the we use order-up-to-S policy, that is,

$$V(S, x_1) \quad = \quad \min\{t \geq 1 : X_t(S) \cdot \mathbf{1}^\tau \leq S,\ X_1(S) = x_1\}.$$

Recall that under order-up-to-S policy, the dynamics of the order quantities and the on-hand inventory levels are given as follows: for any $t \geq 1$,

$$Q_t(S) \quad = \quad \begin{cases} 0\ , & \text{if } t < V(S,x_1) \\ [S - X_t(S)\cdot \mathbf{1}^\tau]^+\ , & \text{if } t = V(S,x_1) \\ \min\{D_{t-1},\ I_{t-1}(S)\}\ , & \text{if } t > V(S,x_1) \end{cases} \qquad \text{and}$$

$$I_t(S) \quad = \quad [I_{t-1}(S) - D_{t-1}]^+ \ + \ Q_{t-\tau}(S)\ .$$

Let $Q'_t(S) = dQ_t(S)/dS$ and $I'_t(S) = dI_t(S)/dS$ denote the sample derivatives of the order quantities and the on-hand inventory level with respect to the order-up-to level S, respectively. The main result of this section is Theorem 1.

Let $\mathbb{I}(\cdot)$ denote the indicator function.

Theorem 1 *Let $S \geq 0$ be a base-stock level, and let $x_1 \in \Re_+^\tau$ be an initial inventory vector $x_1 \in \Re_+^\tau$. Under the order-up-to-S policy, the sample derivatives of the order quantity and of the on-hand inventory satisfy the following: for any $t \geq 1$, $I'_t(S) \in \{0,1\}$ and $Q'_t(S) \in \{0,1\}$, and*

$$Q'_t(S) \quad = \quad \begin{cases} 0\ , & \text{if } 1 \leq t < V(S,x_1) \\ 1\ , & \text{if } t = V(S,x_1) \\ I'_{t-1}\cdot \mathbb{I}[D_{t-1} \geq I_{t-1}]\ , & \text{if } t > V(S,x_1) \end{cases} \qquad \text{and}$$

$$I'_t(S) \quad = \quad I'_{t-1}(S)\cdot \mathbb{I}\left[D_{t-1} < I_{t-1}(S)\right] \ + \ Q'_{t-\tau}(S)\ ,$$

where we define $I'_0(S) = 0$ and $Q'_t(S) = 0$ for all $t \leq 0$. Moreover, for any $t \geq 1$, with probability one,

$$I'_t(S) + \sum_{\ell=t-\tau+1}^{t} Q'_\ell(S) \quad = \quad 1.$$

Proof. Using the definition of the order-up-to policy, Janakiraman and Roundy [10] prove (in their Lemma 1) that for any $t \geq V(S, x_1)$,

$$I_t(S) = S - \sum_{\ell=t-\tau}^{t-1} \min\{I_\ell(S), D_\ell\}\ .$$

Moreover, they show in Corollary 2 of their paper that $Q'_t(S) \in \{0,1\}$ and $I'_t(S) \in \{0,1\}$ for all t.

The desired formulae for the derivatives $Q'_t(S)$ and $I'_t(S)$ follow immediately from the dynamics of the order quantities and the on-hand inventory under an order-up-to-S policy. Moreover, it follows from the above equation that

$$I'_t(S) \quad = \quad 1 - \sum_{\ell=t-\tau}^{t-1} I'_\ell(S)\cdot \mathbb{I}\left[D_\ell \geq I_\ell(S)\right] \quad = \quad 1 - \sum_{\ell=t-\tau+1}^{t} Q'_\ell(S),$$

where the last equality follows from the fact that $Q'_\ell(S) = I'_{\ell-1}(S) \cdot \mathbb{I}[D_{\ell-1} \geq I_{\ell-1}(S)]$ for $\ell \geq V(S,x_1)$ and $Q'_\ell(S) = 0$ for $\ell < V(S,x_1)$. This proves the desired result for $t \geq V(S,x_1)$. For $t < V(S,x_1)$, the result follows from the fact that $I'_t(S) = 1$ and $Q'_\ell(S) = 0$ for all $\ell < V(S,x_1)$. □

Let $\mathcal{N}^\tau = \{x \in \{0,1\}^\tau : \sum_{i=1}^\tau x_i \leq 1\}$ be the set of τ-dimensional binary vectors such that at most one component is 1. Theorem 1 implies $X'_t(S) = \left(Q'_{t-1}(S), \ldots, Q'_{t-\tau+1}(S), I'_t(S)\right) \in \mathcal{N}^\tau$.

3.2 A Sufficient Condition for Ergodicity of the Markov Chain Associated with a Base-Stock Policy In this section, we identify a sufficient condition for the Markov chain associated with a base-stock policy to be ergodic. Under this condition, we establish the convergence rate for the Markov chain to its stationary distribution.

We introduce an *augmented* Markov chain $\overline{X}(S) = \{(X_t(S), X'_t(S)) : t \geq 1\}$ associated with an order-up-to-S policy and establish a sufficient condition for its ergodicity. We let $\overline{X}(S)$ keep track of the inventory vector in each period as well as the sample derivatives of the order quantities and the on-hand inventory. Define, for any $t \geq 1$,

$$(X_t(S), X'_t(S)) \quad = \quad \left(Q_{t-1}(S), \ldots, Q_{t-\tau+1}(S), I_t(S),\ Q'_{t-1}(S), \ldots, Q'_{t-\tau+1}(S), I'_t(S)\right).$$

Lemma 2 *The stochastic process $\overline{X}(S) = \{(X_t(S), X'_t(S)) : t \geq 1\}$ forms a Markov chain.*

Proof. We first note that $X_{t+1}(S) = (Q_t(S), \ldots, Q_{t-\tau+2}(S), I_{t+1}(S))$ depends only on $X_t(S) = (Q_{t-1}(S), \ldots, Q_{t-\tau+1}(S), I_t(S))$ and D_t. Moreover, it follows from Theorem 1 that

$$\begin{aligned} Q'_t(S) &= 1 - I'_t(S) - \sum_{\ell=t-\tau+1}^{t-1} Q'_\ell(S) \qquad \text{and} \\ I'_{t+1}(S) &= 1 - Q'_{t+1}(S) - Q'_t(S) - \sum_{\ell=t-\tau+2}^{t-1} Q'_\ell(S)\ , \end{aligned}$$

where $Q'_{t+1}(S) = I'_t(S) \cdot \mathbb{I}[D_t \geq I_t]$. This shows that $X'_{t+1}(S)$ depends only on $X_t(S)$, $X'_t(S)$, and D_t, giving the desired result. □

We will identify a sufficient condition for the ergodicity of the Markov chain $\overline{X}(S)$. Before we proceed, we recall the definition of ergodicity (see Chapter 13 of Meyn and Tweedie [14] for more details). The Markov chain $\overline{X}(S) = \left\{(X_t(S), X'_t(S)) \in \Re_+^\tau \times \mathcal{N}^\tau : t \geq 1\right\}$ is *ergodic* if there exists a random variable $(X_\infty(S), X'_\infty(S))$ such that for any initial state $(x_1, x'_1) \in \Re_+^\tau \times \mathcal{N}^\tau$,

$$\lim_{t\to\infty} \delta_t\left(S, x_1, x'_1\right) = 0,$$

where, for any $t \geq 1$,

$$\begin{aligned} &\delta_t\left(S, x_1, x'_1\right) \\ &\quad = \quad \sup\Big\{\Big|\mathcal{P}\left[(X_t(S), X'_t(S)) \in B \mid (X_1(S), X'_1(S)) = (x_1, x'_1)\right] - \mathcal{P}\left[(X_\infty(S), X'_\infty(S)) \in B\right]\Big| : \\ &\qquad\qquad \text{measurable set } B \subseteq \Re_+^\tau \times \mathcal{N}^\tau\Big\}. \end{aligned}$$

In such a case, we say $(X_\infty(S), X'_\infty(S))$ is the *steady-state vector* of $\overline{X}(S)$.

The main result of this section is stated in the following theorem that provides a sufficient condition for the ergodicity of the Markov chain $\overline{X}(S)$. Furthermore, it shows that the rate of convergence is exponential in t. For any $S \geq 0$, define

$$\gamma(S) \quad = \quad \mathcal{P}\left[D \leq S/(\tau+1)\right]\ .$$

Theorem 3 *Let $S \geq 0$ be a base-stock level. If $\gamma(S) > 0$, then the Markov chain $\overline{X}(S) = \{(X_t(S), X'_t(S)) : t \geq 1\}$ associated with an order-up-to-S policy is ergodic with a steady-state random variable $(X_\infty(S), X'_\infty(S))$. Furthermore, for any initial inventory vector $(x_1, x'_1) \in \Re_+^\tau \times \mathcal{N}^\tau$, and $t \geq 4\tau + 1$,*

$$\begin{aligned} &\delta_{t+1}\left(S, x_1, x'_1\right) \\ &\quad \leq \quad \begin{cases} \left(1 - \gamma(S)^{2\tau}\right)^{t/(4\tau)} + F\left(\eta\right)^{\frac{t}{2}-\tau}, & \text{if } D \text{ has an infinite support} \\ \left(1 - \gamma(S)^{2\tau}\right)^{t/(4\tau)} + \exp(4\eta/\overline{D} - 2\mu^2(\frac{t}{2}-\tau)/\overline{D}^2), & \text{if } D \leq \overline{D} \text{ with probability one,} \end{cases} \end{aligned}$$

where $F(\cdot)$ denotes the distribution function of D with $\mu = E[D]$, and $\eta = x_1 \cdot \mathbf{1}^\tau - S$ denotes the difference between the initial inventory position $x_1 \cdot \mathbf{1}^\tau$ and the order-up-to level S.

The proof of the above theorem appears in Appendix A, and it is based on the standard coupling argument in Markov chain theory. The main idea of the proof is the observation that, regardless of the initial state, all sample paths of the Markov chain couple after a certain pattern (or sequence) of consecutive demands occurs. If the initial inventory position is at most S, then an example of such a demand pattern is a sequence of τ consecutive periods of zero demand, which will result in a state in which the inventory on hand is S units and there is no outstanding order regardless of starting inventory levels. When initial inventory exceeds S, we can construct a similar demand pattern that will result in coupling. Thus, we can obtain an upper bound on $\delta_{t+1}(S, x_1, x_1')$ based on the probability that certain demand patterns do not occur by period t.

An Example of Non-Ergodicity We now show that the Markov chain $\overline{X}(S) = \{(X_t(S), X_t'(S)) : t = 1, 2, \ldots\}$ may not be ergodic if the condition of Theorem 3 fails, that is, if $\gamma(S) = 0$. The key idea behind this example is that if S is too small, then a stockout occurs in every period, which causes the after-delivery on-hand inventory in each period t to be exactly the amount ordered in period $t - \tau$. Thus, the inventory vector $X_t(S)$ follows a cyclic pattern.

Consider a base-stock level S such that there exists sufficiently small ε such that $0 < \varepsilon < \tau S$ and $\gamma(S + \varepsilon) = \mathcal{P}[D \le (S + \varepsilon)/(\tau + 1)] = 0$. Suppose that the initial inventory vector $X_1(S) = (Q_0(S),\ Q_{-1}(S),\ \ldots,\ Q_{-\tau+2}(S),\ I_1(S))$ is given by

$$I_1(S) = \frac{S}{\tau+1} + \frac{\varepsilon}{\tau+1} \qquad \text{and} \qquad Q_t(S) = \frac{S}{\tau+1} - \frac{\varepsilon/\tau}{\tau+1} \qquad \text{for each } t = -\tau+2,\ \ldots,\ 1,\ 0\ .$$

Then, the quantity ordered in period 1 is given by

$$\begin{aligned} Q_1(S) &= S - (Q_0(S) + Q_1(S) + \cdots + Q_{-\tau+2}(S) + I_1(S)) \\ &= S - \left((\tau-1)\cdot\left(\frac{S}{\tau+1} - \frac{\varepsilon/\tau}{\tau+1}\right) + \left(\frac{S}{\tau+1} + \frac{\varepsilon}{\tau+1}\right)\right) \\ &= S - \left(\tau\cdot\frac{S}{\tau+1} + \frac{\varepsilon/\tau}{\tau+1}\right) = \frac{S}{\tau+1} - \frac{\varepsilon/\tau}{\tau+1}\ , \end{aligned}$$

which is strictly positive since $\varepsilon < S\tau$. Since $\gamma(S + \varepsilon) = 0$, it follows that the event

$$D > \frac{S}{\tau+1} + \frac{\varepsilon}{\tau+1}$$

occurs with probability 1, that is, D is greater than $Q_1(S)$ and each component of $X_1(S)$ with probability 1. Thus, the process of inventory vectors $\{X_t(S) : t = 1, 2, \ldots\}$ follows a cyclic process where at most one of the components of each inventory vector $X_t(S)$ is $S/(\tau+1) + \varepsilon/(\tau+1)$ and all the other components are $S/(\tau+1) - (\varepsilon/\tau)/(\tau+1)$. Similarly, we can show that $\{X_t'(S) : t = 1, 2, \ldots\}$ is also cyclic.

3.3 Structure of the Cost Function and the Optimal Base-Stock Level We now provide a characterization of the long-run average holding cost and lost sales penalty under any order-up-to policy. The main result of this section is stated in Theorem 4, which expresses the long-run average cost as a function of the steady-state stationary distribution of inventory levels, and establishes the convexity of the cost function with respect to the base-stock level.

Recall $\gamma(S) = \mathcal{P}[D \le S/(\tau+1)]$. To simplify our exposition, we use the expressions $C(I_t(S))$ to denote the expected holding cost and lost sales penalty in period t under the order-up-to-S policy, that is,

$$C(I_t(S)) = h \cdot E[D_t - I_t(S)]^+ + b \cdot E[I_t(S) - D_t]^+\ ,$$

where the expectation is taken with respect to *both* the random variables D_t and $I_t(S)$. Similarly, we use $C(I_\infty(S))$ to denote the long-run average expected cost under the order-up-to-S policy.

Theorem 4 *For any $S \ge 0$, the long-run average holding cost and lost sales penalty under an order-up-to-S policy always exists, is independent of the initial starting inventory vector, and satisfies*

$$C(I_\infty(S)) := \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} C(I_t(S)) = \begin{cases} b\cdot\left(E[D] - \frac{S}{\tau+1}\right), & \text{if } \gamma(S) = 0, \\ b\cdot E[D - I_\infty(S)]^+ + h\cdot E[I_\infty(S) - D]^+, & \text{if } \gamma(S) > 0. \end{cases}$$

Moreover, the function $C(I_\infty(S))$ is convex and differentiable in S, and has a minimizer S^* satisfying $\gamma(S^*) > 0$.

PROOF. If $\gamma(S) > 0$, we know from Theorem 3 that the Markov chain $\overline{X}(S) = \{(X_t(S), X'_t(S)) : t \geq 1\}$ converges to the stationary random vector $(X_\infty(S), X'_\infty(S))$, and the stated expression for the long-run average cost follows from Markov chain theory. When $\gamma(S) = 0$, it follows that $D \geq S/(\tau+1)$ with probability one. Huh et al. [7] show (in Lemma 10 of their paper) that in this case, the long-run average cost is equal to $b(E[D] - \frac{S}{\tau+1})$, which is the desired result.

It is easy to verify that $C(I_\infty(S))$ is continuous in S. The differentiability of $C(I_\infty(\cdot))$ follows from the above formula since D is a continuous random variable. Moreover, let $\hat{S} = \sup\{x : \gamma(x) = 0\}$. Huh et al. [7] (in their Appendix E and F) show that the left and the right derivatives of $C(I_\infty(S))$ at $\hat{S}$ are $-b/(\tau+1)$, i.e.,

$$\lim_{S\uparrow\hat{S}} \frac{d}{dS} C(I_\infty(S)) \;=\; \lim_{S\downarrow\hat{S}} \frac{d}{dS} C(I_\infty(S)) \;=\; \frac{-b}{\tau+1}.$$

Thus, there exists a minimizing S^* such that $\gamma(S^*) > 0$. For $S > \hat{S}$, the convexity of $C(I_\infty(S))$ is established in Theorem 12 in [10]. Since the function is linear for $S < \hat{S}$ and the left and right derivatives at $\hat{S}$ coincide at $\hat{S}$, the convexity of $C(I_\infty(S))$ follows for all S. □

3.4 Sample-Based Estimation of the Cost and Its Derivative To estimate the cost function $C(I_\infty(S))$ and its derivative with respect to S, one can run the system for a long time, and obtain appropriate sample-based estimates. However, for any finite $t \geq 1$, the distribution of the state in period t is not in general exactly the same as its stationary distribution, resulting in a bias in the above estimates. In this section, we establish error bounds associated with sample-based estimates of the cost function $C(I_\infty(S))$ and its derivative. The main result of this section is stated in the following theorem.

Theorem 5 *Let $S \geq 0$ be a base-stock level such that $\gamma(S) = \mathcal{P}[D \leq S/(\tau+1)] > 0$. Let $(x_1, x'_1) \in \Re_+^\tau \times \mathcal{N}^\tau$. If we apply the order-up-to-S policy with the initial inventory vector $X_1(S) = x_1$ and $X'_1(S) = x'_1$, then, for any $t \geq 1$,*

$$|C(I_\infty(S)) - C(I_t(S))| \;\leq\; (b+h) \cdot \max\{S,\ x_1 \cdot \mathbf{1}^\tau\} \cdot \delta_t(S, x_1, x'_1)\ .$$

Moreover,

$$\left| \frac{d}{dS} C(I_\infty(S)) - \frac{d}{dS} C(I_t(S)) \right| \;\leq\; (b+h) \cdot \delta_t(S, x_1, x'_1)\,.$$

In the proof of Theorem 5, we express the difference in the cost functions $|C(I_\infty(S)) - C(I_t(S))|$ in terms of the truncated expectations of $I_\infty(S)$ and $I_t(S)$. We then upper bound the difference in terms of $\delta_t(S, x_1, x'_1)$ using the Markov chain ergodicity results. The details of the proof appear in Appendix B.

4. An Adaptive Algorithm Building upon the results of the previous section, we propose an adaptive algorithm that determines the base-stock level for each period, where the decision in each period depends only on the observed sales data in the past. We also establish the convergence rate of our algorithm. As a benchmark, we compare the running average holding cost and lost sales penalty of our algorithm to the cost of the optimal base-stock policy. Let S^* be the optimal base-stock level. We make the following assumption throughout Section 4.

Assumption 1 *The manager has an* a priori *knowledge of a lower bound $\underline{M} \geq 0$ and an upper bound $\overline{M} \geq 0$ on S^*, i.e., $\underline{M} \leq S^* \leq \overline{M}$, and $\gamma(\underline{M}) = \mathcal{P}[D \leq \underline{M}/(\tau+1)] > 0$.*

We note for any demand distribution with positive probability at zero, the choice of $\underline{M} = 0$ satisfies the condition of Assumption 1. Throughout the remainder of this section, we will also assume without loss of generality that the demand random variable has an infinite support. We emphasize that this assumption is taken primarily to simplify our exposition and the formula for the error bounds. When the demand is bounded almost surely, **exactly** the same argument applies. (See the error bounds given in Theorem 3.) Although our algorithm applies to both bounded and unbounded demands, we note that Assumption 1 requires us to know in advance the upper bound $\overline{M}$ on the optimal order-up-to level S^*. Extending our analysis to the case when $\overline{M}$ is unknown remains an open research question.

4.1 Description of the Algorithm Leveraging the convexity of $C\left(I_\infty(S)\right)$ as a function of the order-up-to level S, we extend an existing result from the online convex optimization literature, which requires an unbiased estimate of the derivative $dC\left(I_\infty(S)\right)/dS$ of the cost function (see, for example, [4, 12, 20]). However, in our case, we cannot obtain an unbiased sample of the cost function and its derivative because they depend on the steady-state on-hand inventory level $I_\infty(S)$; furthermore, the magnitude of the bias depends on how long the system has been running at a given base-stock level. In this section, we propose an algorithm that carefully balances (i) the benefit of reducing the bias in the estimator by fixing the base-stock level for a number of periods, and (ii) the ability to explore new base-stock levels based on our estimate of the gradient.

To address the issue of bias, we propose the notion of cycles which consists of multiple periods. We divide time into a sequence of cycles, and maintain the same base-stock level within a cycle. Base-stock levels may be adjusted from one cycle to another. Furthermore, the cycles have unequal lengths; as time elapses, we gain more confidence in our solution and increase the length of the cycle. While our algorithm is a modification of existing online algorithms such as Flaxman et al. [4], our results do not immediately follow from the existing literature because our algorithm makes use of unequal cycle lengths. The running average cost under our algorithm is a *weighted* average of cycle costs, where the weights correspond to the lengths of cycles. Since the weights are unequal across cycles, we need additional arguments and analysis. However, in the analysis the modification with unequal cycle lengths, we make extensive use of existing results (without cycles) as a "black box" (see, for example, Lemma 10).

Let S_k denote the order-up-to level for the k^{th} cycle. We will use the sample derivative of the cost function evaluated in *the last period of the cycle* as a proxy for $dC\left(I_\infty(S_k)\right)/dS$, which will be discussed subsequently. If the length of the k^{th} cycle is sufficiently long, the ergodicity of the Markov chain $\{(X_t(S_k), X'_t(S_k)) : t \geq 1\}$ should ensure that our estimate has a small bias compared to $dC\left(I_\infty(S_k)\right)/dS$.

Our adaptive algorithm, which we refer to as ADAPTIVE(α, β), is parameterized by two parameters $\alpha, \beta \in (0,1)$. The first parameter α controls the adjustment of the order-up-to level between two successive cycles while the second parameter β controls the length of each cycle. (It will be shown that the choice of $\alpha = \beta = 1/2$ minimizes the asymptotic bound of the regret; however, we keep our exposition general to show how the bounds depend on these two parameters.) We use k to index cycles and j to index periods within a given cycle. Let (k, j) denote the j^{th} period in the k^{th} cycle. We now describe the algorithm in details.

Algorithm ADAPTIVE(α, β)

INITIALIZATION: For the first cycle, set the order-up-to level S_1 to any number in $\left[\underline{M}, \overline{M}\right]$, and set the initial inventory vector $X_{(1,1)} \in \Re^\tau_+$ such that $X_{(1,1)} \cdot \mathbf{1}^\tau \leq \overline{M}$.

ALGORITHM DEFINITION: For each cycle $k = 1, 2, \ldots,$

- The length of cycle k, denoted by T_k, is defined by $T_k := \lceil k^\beta \rceil$, and cycle k begins at period $\sum_{k'=1}^{k-1} T_{k'} + 1$ and ends at $\sum_{k'=1}^{k} T_{k'}$ (inclusive).
- Let S_k denote the base-stock level for this cycle. The initial inventory vector in cycle k is given by $X_{(k,1)}$. We will use the order-up-to-S_k policy for *every* period in cycle k. Let $X_{(k,j)}$ and $I_{(k,j)}(S_k; X_{(k,1)})$ denote the inventory vector and the on-hand inventory level, respectively, in the j^{th} period of the k^{th} cycle.
- For each period $1 \leq j \leq T_k$ in cycle k, compute an estimate of the sample-path derivative of the on-hand inventory $I'_{(k,j)}\left(S_k; X_{(k,1)}\right)$ using the following recursion from Theorem 1:

$$I'_{(k,j)}\left(S_k; X_{(k,1)}\right) = 1 - \sum_{\ell=j-\tau}^{j-1} I'_{(k,\ell)}\left(S_k; X_{(k,1)}\right) \cdot \mathbb{I}\left[I_{(k,\ell)}\left(S_k; X_{(k,1)}\right) \leq D_{(k,\ell)}\right],$$

  where $D_{(k,\ell)}$ is the realized demand in the ℓ^{th} period of the k^{th} cycle. Note that we define $I'_{(k,j)}\left(S_k; X_{(k,1)}\right) = 0$ if $j \leq 0$. Thus, to compute the sample-based derivative in each period, we only need to keep the derivative values from at most the τ previous periods. Moreover, note that the event $\mathbb{I}\left[I_{(k,\ell)}\left(S_k; X_{(k,1)}\right) \leq D_{(k,\ell)}\right]$ can be computed based on the *sales* data in the ℓ^{th} period of the j^{th} cycle. We simply need to check whether or not we have a stockout.
- At the end of the k^{th} cycle (period T_k of the k^{th} cycle), update the base-stock level as follows.

Let

$$\epsilon_k = \frac{(\overline{M} - \underline{M})}{\max\{b,h\} \cdot k^{\alpha}},$$

and let $H_k(S_k)$ be defined by

$$H_k(S_k) \quad = \quad \begin{cases} h, & \text{if } I'_{(k,T_k)}(S_k; X_{(k,1)}) = 1 \text{ and } I_{(k,T_k)} > D_{(k,T_k)}, \\ -b, & \text{if } I'_{(k,T_k)}(S_k; X_{(k,1)}) = 1 \text{ and } I_{(k,T_k)} \leq D_{(k,T_k)}, \\ 0, & \text{if } I'_{(k,T_k)}(S_k; X_{(k,1)}) = 0. \end{cases}$$

The base-stock level for the cycle $k+1$ is then given by

$$S_{k+1} \quad = \quad P_{[\underline{M},\overline{M}]}(S_k - \epsilon_k \cdot H_k(S_k)) \ ,$$

where $P_{[\underline{M},\overline{M}]}(z) = \max\{\underline{M},\ \min\{z, \overline{M}\}\}$ is the projection operator.

- The initial inventory vector $X_{(k+1,1)}$ for cycle $k+1$ will correspond to the inventory vector *after ordering* at the beginning of the first period of cycle $k+1$.

For any $L \geq 1$, let $N(L) = \sum_{k=1}^{L} T_k$ denote the total number of time periods in the first L cycles. We define the *L-cycle regret* $\Lambda(L)$ as follows:

$$\Lambda(L) \quad = \quad E\left[\sum_{k=1}^{L}\sum_{j=1}^{T_k} C(I_{(k,j)}(S_k; X_{(k,1)}))\right] - E\left[C(I_\infty(S^*))\right] \cdot N(L) \ ,$$

where S^* is the optimal base-stock level. The main result of this section is that the *L-cycle per-period average regret*, the expression $\Lambda(L)$ divided by $N(L)$, converges to zero at the rate of $O\left(N(L)^{-1/3}\right)$ if the α and β parameters are chosen carefully. This result is stated in Theorem 6, whose proof is given in Section 4.3.

Theorem 6 *Under Assumption 1, let $\nu = \max\left\{1 - \gamma(\underline{M})^{2\tau}, F(\overline{M}), 1/e\right\}$. Then, for any $\alpha, \beta \in (0,1)$, the L-cycle per-period average regret under the algorithm* Adaptive(α,β) *satisfies*

$$\frac{\Lambda(L)}{N(L)} \quad \leq \quad (b+h) \cdot \left(\overline{M} - \underline{M}\right) \cdot \left\{ \frac{C_1(\alpha,\beta)}{N(L)^{\frac{1-\alpha}{1+\beta}}} + \frac{C_2(\alpha,\beta)}{N(L)^{\frac{\alpha}{1+\beta}}} + \frac{C_3(\alpha,\beta)}{N(L)^{\frac{1}{1+\beta}}} + \frac{C_4(\alpha,\beta)}{N(L)^{\frac{\beta}{1+\beta}}} \right\},$$

where the constants are given by:

$$\begin{aligned} C_1(\alpha,\beta) &= 4, \\ C_2(\alpha,\beta) &= \frac{4}{1-\alpha}, \\ C_3(\alpha,\beta) &= \frac{12\,(4\tau)^{1/\beta}\,\Gamma(1/\beta)\,(1/\beta)}{(\ln(1/\nu))^{1/\beta}}, \\ C_4(\alpha,\beta) &= \frac{24\tau}{\ln(1/\nu)}, \end{aligned}$$

and $\Gamma(\cdot)$ denotes the Gamma function. If we set $\alpha = \beta = 1/2$, then $\Lambda(L)/N(L) = O\left(N(L)^{-1/3}\right)$.

We note that the choice of $\alpha = \beta = 1/2$ is the one that minimizes the asymptotic per-period regret. To see this, note that

$$\max\left\{ \frac{1}{N(L)^{\frac{1-\alpha}{1+\beta}}}, \frac{1}{N(L)^{\frac{\alpha}{1+\beta}}}, \frac{1}{N(L)^{\frac{1}{1+\beta}}}, \frac{1}{N(L)^{\frac{\beta}{1+\beta}}} \right\} = \max\left\{ \frac{1}{N(L)^{\frac{1-\alpha}{1+\beta}}}, \frac{1}{N(L)^{\frac{\alpha}{1+\beta}}}, \frac{1}{N(L)^{\frac{\beta}{1+\beta}}} \right\} \ ,$$

where the equality follows from the fact that $0 \leq \beta \leq 1$. The expression in the above right hand side achieves the minimum when $\alpha = \beta = 1/2$.

4.2 Preliminary Results Online convex optimization is the minimization of a convex function, for which little is known *a priori* except the convexity of the objective function. At each iteration, we choose a point in the feasible region and incur the cost associated with this point; however, we obtain some information about the function at this point, such as the gradient or its stochastic estimator. The objective is to minimize the average cost over time.

The following theorem, which appears in [8], is an extension of the result of Zinkevich [20], Flaxman et al. [4], and Kleinberg [12] to allow for a *biased* stochastic gradient estimate in each iteration and for a general step size of the form $\epsilon_t = O(1/t^\alpha)$ with $0 < \alpha < 1$. These extensions are necessary for the analysis of the regret under our proposed ADAPTIVE(α, β) policy. Note that for any compact set $\mathcal{S}$, $P_{\mathcal{S}}(\cdot)$ denotes the projection operator on $\mathcal{S}$.

Theorem 7 *Let $\Phi : \mathcal{S} \to \mathbb{R}$ be a convex function and let $z^* = \arg\min_{z\in\mathcal{S}} \Phi(z)$ be its minimizer. For any $z \in \mathcal{S}$, let $H_t(z)$ be an n-dimensional random vector defined on $\mathcal{S}$, and suppose that there exists $\overline{B} > 0$ such that $E\left[\|H_t(z)\|^2\right] \le \overline{B}^2$ holds for all $z \in \mathcal{S}$. Let the sequence $(Z_t : t \ge 1)$ be defined by*

$$Z_{t+1} \;=\; P_{\mathcal{S}}\left(Z_t - \epsilon_t \cdot H_t(Z_t)\right), \qquad \textit{where} \qquad \epsilon_t \;=\; \frac{\zeta\, diam(\mathcal{S})}{\overline{B}} \cdot \frac{1}{t^\alpha}$$

for some $\zeta > 0$ and $\alpha \in (0,1)$, where Z_1 is any point in $\mathcal{S}$. Let $\eta^A(z) \;=\; E\left[H_t(z) \mid z\right] - \bigtriangledown\Phi(z)$. Then, for all $T \ge 1$,

$$\sum_{t=1}^{T} E\left[\Phi(Z_t) - \Phi(z^*)\right] \;\le\; diam(\mathcal{S}) \left\{ \overline{B} \cdot \left[\frac{T^\alpha}{2\zeta} + \frac{\zeta\; T^{1-\alpha}}{2(1-\alpha)}\right] + \sum_{t=1}^{T} E\left|\eta^A(Z_t)\right| \right\}.$$

The next two lemmas are used in the analysis of Section 4.3 and their proofs appear in Appendix C. Let $\Gamma(\cdot)$ represent the gamma function, which is defined by $\Gamma(z) = \int_0^\infty w^{z-1}e^{-w}dw$ for any real number $z > 0$. The proofs of these results are based on algebraic manipulation and the expression for the cumulative density function of a gamma distribution.

Lemma 8 *For any $\rho \in (0,1)$, $\beta \in (0,1)$, and $L \ge 1$,*

$$\sum_{k=1}^{L} \rho^{\lceil k^\beta \rceil} \;\le\; \sum_{k=1}^{L} \rho^{k^\beta} \;\le\; \frac{\Gamma\left(1/\beta\right)\left(1/\beta\right)}{\left(\ln\left(1/\rho\right)\right)^{1/\beta}}.$$

From the description of the algorithm described in Section 4.1, $T_k = \lceil k^\beta \rceil$ is the length of the k^{th} cycle, and $N(k) = \sum_{k'=1}^{k} T_{k'}$ denotes the total length of the first k cycles. The following lemma establishes the relationship among k, $\lceil k^\beta \rceil$ and $N(k)$.

Lemma 9 *Let $\beta \in (0,1)$. For $k \ge 1$, let $N(k) = \sum_{k'=1}^{k} \lceil k'^\beta \rceil$. Then,*

(i) $k \;\le\; \left[(\beta+1)\cdot N(k)\right]^{1/(\beta+1)}$.

(ii) $k \cdot \lceil k^\beta \rceil \;\le\; 2(\beta+1)\cdot N(k)$.

(iii) $\lceil k^\beta \rceil \;\le\; 2\left[(\beta+1)\cdot N(k)\right]^{\beta/(\beta+1)}$.

(iv) $\lceil k^\beta \rceil^\alpha \;\le\; 2\left[(\beta+1)\cdot N(k)\right]^{\alpha\beta/(\beta+1)}$ *for any $\alpha \in (0,1)$.*

4.3 Proof of Theorem 6 We express the L-cycle total regret $\Lambda(L)$ as a sum of the following two expressions:

$$\Lambda_1(L) \;=\; \sum_{k=1}^{L} T_k \cdot \left\{E\left[C(I_\infty(S_k))\right] - E\left[C(I_\infty(S^*))\right]\right\}, \qquad \text{and}$$

$$\Lambda_2(L) \;=\; E\left[\sum_{k=1}^{L}\sum_{j=1}^{T_k} \left\{C(I_{(k,j)}(S_k; X_{(k,1)})) - C(I_\infty(S_k))\right\}\right].$$

The first expression $\Lambda_1(L)$ corresponds to the regret due to the deviation of S_k from S^* – note that $\Lambda_1(L)$ is the weighted sum of the loss of optimality due to S_k, where weights correspond to the cycle length T_k. The second expression $\Lambda_2(L)$ reflects how much the on-hand inventory levels $\{I_{(k,j)} \mid j = 1, 2, \ldots T_k\}$ differ from the stationary on-hand inventory level of that cycle. We provide an upper bound for each term in Lemmas 10 and 11.

Let $\nu = \max\left\{1 - \gamma(\underline{M})^{2\tau}, F(\overline{M}), 1/e\right\}$. The proof of the following lemma is obtain from our earlier treatment of Markov chain ergodicity (Section 3), and also from addressing the issue of unequal cycle lengths by applying, as a black box, an existing online algorithm result (Theorem 7), which does not allow the notion of cycles.

Lemma 10 *Suppose Assumption 1 holds and D has an infinite support. Then, for any $\alpha, \beta \in (0,1)$, the algorithm* ADAPTIVE(α, β) *satisfies*

$$\Lambda_1(L) \quad \leq \quad \left(\overline{M} - \underline{M}\right) \cdot (b+h) \cdot T_L \cdot \left\{ \frac{L^\alpha}{2} + \frac{L^{1-\alpha}}{2(1-\alpha)} + \frac{3\,(4\tau)^{1/\beta}\,\Gamma\,(1/\beta)\,(1/\beta)}{(\ln\,(1/\nu))^{1/\beta}} \right\}.$$

PROOF. From the definition of $\Lambda_1(L)$ and the fact that $T_1 \leq \cdots \leq T_L$, we have

$$\frac{\Lambda_1(L)}{T_L} \quad \leq \quad \sum_{k=1}^{L} \left\{ E\left[C(I_\infty(S_k))\right] - E\left[C(I_\infty(S^*))\right] \right\} \ .$$

From Theorem 4, $E\left[C(I_\infty(S))\right]$ is a convex function of the base-stock level S. Moreover, the dynamics of S_k defined in the algorithm ADAPTIVE(α, β) are exactly the same as the gradient descent method defined in Theorem 7, with $\mathcal{S} = \left[\underline{M}, \overline{M}\right]$, $\zeta = 1$, and $\overline{B} = \max\{b, h\}$. Thus, we obtain

$$\begin{aligned} &\sum_{k=1}^{L} \left\{ E\left[C(I_\infty(S_k))\right] - E\left[C(I_\infty(S^*))\right] \right\} \\ &\quad \leq \quad \left(\overline{M} - \underline{M}\right) \cdot \left\{ \max\{b,h\} \cdot \left[\frac{L^\alpha}{2} + \frac{L^{1-\alpha}}{2(1-\alpha)} \right] + \sum_{k=1}^{L} E\left|\eta^A(S_k)\right| \right\} \ , \end{aligned}$$

where

$$\eta^A(S) \quad = \quad \frac{d}{dS} E\left[C(I_{(k,T_k)}(S; X_{(k,1)}))\right] - \frac{d}{dS} E\left[C(I_\infty(S))\right] \ . \tag{2}$$

Note $\max\{b,h\} \leq b + h$.

We will now establish an upper bound for $\sum_{k=1}^{L} E\left|\eta^A(S_k)\right|$. There are two cases to consider: $\lceil k^\beta \rceil \leq 4\tau$ and $\lceil k^\beta \rceil \geq 4\tau + 1$. Suppose that $\lceil k^\beta \rceil \leq 4\tau$. The definition of $C(\cdot)$ implies $C'(\cdot) \in [-b, h]$. From $I'_{(k,T_k)}(S; X_{(k,1)}) \in \{0,1\}$ for all k, it follows that $\left|\eta^A(S_k)\right| \leq b + h$. Note that the condition $\lceil k^\beta \rceil \leq 4\tau$ is equivalent to $k \leq (4\tau)^{1/\beta}$, which implies that

$$\sum_{k=1}^{L} \left|\eta^A(S_k)\right| \cdot \mathbb{I}[\lceil k^\beta \rceil \leq 4\tau] \quad \leq \quad (b+h) \cdot (4\tau)^{1/\beta} \ .$$

Suppose that $\lceil k^\beta \rceil \geq 4\tau + 1$. Let $\eta_k = X_{(k,1)} \cdot \mathbf{1}^\tau - S_k$. Theorem 3, Theorem 5 and Assumption 1 imply that

$$\begin{aligned} \left|\eta^A(S_k)\right| \quad &\leq \quad (b+h) \cdot \left[\left(1 - \gamma(S_k)^{2\tau}\right)^{\lceil k^\beta \rceil/(4\tau)} \ + \ F(\eta_k)^{\lceil k^\beta \rceil/2-\tau} \right] \\ &\leq \quad (b+h) \cdot \left[\left(1 - \gamma(\underline{M})^{2\tau}\right)^{\lceil k^\beta \rceil/(4\tau)} \ + \ F(\overline{M})^{\lceil k^\beta \rceil/2-\tau} \right] \\ &\leq \quad 2(b+h) \cdot \max\left\{ \left(1 - \gamma(\underline{M})^{2\tau}\right), F(\overline{M}) \right\}^{\lceil k^\beta \rceil/(4\tau)} \\ &\leq \quad 2(b+h) \cdot \nu^{\lceil k^\beta \rceil/(4\tau)}, \end{aligned}$$

where the second inequality follows from the fact that $\gamma(\cdot)$ and $F(\cdot)$ are nondecreasing functions. The third inequality follows from the fact that $\left\lceil k^\beta \right\rceil \geq 4\tau + 1$ and $\tau \geq 1$, which implies that $\lceil k^\beta \rceil/(4\tau) \leq \lceil k^\beta \rceil/2 - \tau$.

It thus follows from Lemma 8 that

$$\begin{aligned}\sum_{k=1}^{L}\left|\eta^A(S_k)\right|\cdot\mathbb{I}[\lceil k^\beta\rceil\geq 4\tau+1] &\leq 2(b+h)\cdot\frac{\Gamma\left(1/\beta\right)\left(1/\beta\right)}{\left(\ln\left(1/\nu^{1/(4\tau)}\right)\right)^{1/\beta}}\\ &= 2(b+h)\cdot(4\tau)^{1/\beta}\cdot\frac{\Gamma\left(1/\beta\right)\left(1/\beta\right)}{\left(\ln\left(1/\nu\right)\right)^{1/\beta}}\ .\end{aligned}$$

Combining the two cases, we see that

$$\begin{aligned}\sum_{k=1}^{L}\left|\eta^A(S_k)\right| &\leq (4\tau)^{1/\beta}\cdot\left((b+h)+\frac{2(b+h)\cdot\Gamma\left(1/\beta\right)\left(1/\beta\right)}{\left(\ln\left(1/\nu\right)\right)^{1/\beta}}\right)\\ &\leq 3(b+h)\cdot\frac{(4\tau)^{1/\beta}\,\Gamma\left(1/\beta\right)\left(1/\beta\right)}{\left(\ln\left(1/\nu\right)\right)^{1/\beta}},\end{aligned}$$

where the last inequality follows from the fact that $0\leq\ln(1/\nu)\leq 1$ and $1\leq\Gamma(1/\beta)(1/\beta)$, and we obtain the required result. □

Lemma 11 *Suppose Assumption 1 holds and D has an infinite support. Then, for any $\alpha,\beta\in(0,1)$, the algorithm* Adaptive(α,β) *satisfies*

$$\Lambda_2(L) \leq (b+h)\cdot(\overline{M}-\underline{M})\cdot L\cdot\frac{12\tau}{\ln\left(1/\nu\right)}.$$

Proof. Recall that

$$\Lambda_2(L)=E\left[\sum_{k=1}^{L}\sum_{j=1}^{T_k}\left\{C(I_{(k,j)}(S_k;X_{(k,1)}))-C(I_\infty(S_k))\right\}\right].$$

Consider the summand $C(I_{(k,j)}(S_k;X_{(k,1)}))-C(I_\infty(S_k))$ for $1\leq j\leq T_k$. There are two cases to consider: $j\leq 4\tau$ and $j\geq 4\tau+1$. Suppose $j\leq 4\tau$. By the convexity of the cost function $C(\cdot)$, $\left|E\left[C(I_{(k,j)}(S_k;X_{(k,1)}))\right]-E\left[C(I_\infty(S_k))\right]\right|$ is bounded above by $(\overline{M}-\underline{M})\cdot\max\{b,h\}$. Therefore,

$$\sum_{j=1}^{T_k}\left|E\left[C(I_{(k,j)}(S_k;X_{(k,1)}))\right]-E\left[C(I_\infty(S_k))\right]\right|\cdot\mathbb{I}[j\leq 4\tau] \leq 4\cdot\tau\cdot(\overline{M}-\underline{M})\cdot\max\{b,h\}\ .$$

Now, suppose $j\geq 4\tau+1$. By Theorem 5,

$$\begin{aligned}&\left|E\left[C(I_{(k,j)}(S_k;X_{(k,1)}))\right]-E\left[C(I_\infty(S_k))\right]\right|\\ &\quad\leq (b+h)\cdot(\overline{M}-\underline{M})\cdot\left[\left(1-\gamma(\underline{M})^{2\tau}\right)^{j/(4\tau)}+F(\overline{M})^{j/2-\tau}\right]\ .\end{aligned}$$

Therefore,

$$\begin{aligned}&\sum_{j=1}^{T_k}\left|E\left[C(I_{(k,j)}(S_k;X_{(k,1)}))\right]-E\left[C(I_\infty(S_k))\right]\right|\cdot\mathbb{I}[j\geq 4\tau+1]\\ &\quad\leq (b+h)\cdot(\overline{M}-\underline{M})\cdot\sum_{j=1}^{T_k}\left[\left(1-\gamma(\underline{M})^{2\tau}\right)^{j/(4\tau)}\ +\ F(\overline{M})^{j/2-\tau}\right]\\ &\quad\leq 2(b+h)\cdot(\overline{M}-\underline{M})\cdot\sum_{j=1}^{T_k}\max\left\{1-\gamma(\underline{M})^{2\tau},F(\overline{M})\right\}^{j/(4\tau)}\\ &\quad\leq 2(b+h)\cdot(\overline{M}-\underline{M})\cdot\int_0^\infty\nu^{z/(4\tau)}dz\\ &\quad= 2(b+h)\cdot(\overline{M}-\underline{M})\cdot\frac{4\tau}{\ln\left(1/\nu\right)},\end{aligned}$$

where the second inequality follows from the fact that $j\geq 4\tau+1$, which implies that $j/(4\tau)\leq j/2-\tau$. The last inequality follows from $\max\left\{1-\gamma(\underline{M})^{2\tau},F(\overline{M})\right\}\leq\nu<1$.

Combining the two cases, it follows that

$$\begin{aligned}\sum_{j=1}^{T_k} \left|E\left[C(I_{(k,j)}(S_k; X_{(k,1)}))\right] - E\left[C(I_\infty(S_k))\right]\right| \\ \leq \quad 4\cdot\tau\cdot(\overline{M}-\underline{M})\cdot\left(\max\{b,h\}+\frac{2(b+h)}{\ln(1/\nu)}\right) \\ \leq \quad (b+h)\cdot(\overline{M}-\underline{M})\cdot\frac{12\tau}{\ln(1/\nu)},\end{aligned}$$

where we use the fact that $0 \leq \ln(1/\nu) \leq 1$ for the second inequality. Summing the above inequality over all possible values of $k = 1, \ldots, L$ gives the required result. □

We will now prove Theorem 6.

Proof. From Lemma 10, we have

$$\begin{aligned}\Lambda_1(L) &\leq \left(\overline{M}-\underline{M}\right)\cdot(b+h)\cdot T_L\cdot\left\{\frac{L^\alpha}{2}+\frac{L^{1-\alpha}}{2(1-\alpha)}+\frac{3\,(4\tau)^{1/\beta}\,\Gamma\,(1/\beta)\,(1/\beta)}{(\ln(1/\nu))^{1/\beta}}\right\} \\ \Lambda_2(L) &\leq \left(\overline{M}-\underline{M}\right)\cdot(b+h)\cdot L\cdot\frac{12\tau}{\ln(1/\nu)}.\end{aligned}$$

It follows from Lemma 9 and $T_L = \lceil L^\beta \rceil$ that

$$\begin{aligned}T_L\cdot L^\alpha &= \left(L\cdot\lceil L^\beta\rceil\right)^\alpha\cdot\lceil L^\beta\rceil^{1-\alpha} \\ &\leq (2\cdot(\beta+1)\cdot N(L))^\alpha\cdot 2\cdot((\beta+1)\cdot N(L))^{(1-\alpha)\beta/(\beta+1)} \\ &= 2^{1+\alpha}(\beta+1)^{\alpha+(1-\alpha)\beta/(\beta+1)}\cdot N(L)^{\alpha+(1-\alpha)\beta/(\beta+1)} \\ &\leq 8\cdot N(L)^{(\alpha+\beta)/(1+\beta)}.\end{aligned}$$

A similar argument shows that $T_L \cdot L^{1-\alpha} \leq 8\cdot N(L)^{(1-\alpha+\beta)/(1+\beta)}$. Also, by Lemma 9,

$$T_L \;=\; \lceil L^\beta\rceil \;\leq\; 2\left((\beta+1)\cdot N(L)\right)^{\beta/(\beta+1)} \;\leq\; 4N(L)^{\beta/(1+\beta)}\;.$$

Thus, we obtain

$$\frac{\Lambda_1(L)}{N(L)} \quad\leq\quad \left(\overline{M}-\underline{M}\right)\cdot(b+h)\cdot\left\{\frac{4}{N(L)^{(1-\alpha)/(1+\beta)}}+\frac{4/(1-\alpha)}{N(L)^{\alpha/(1+\beta)}}+\frac{12\,(4\tau)^{1/\beta}\,\Gamma\,(1/\beta)\,(1/\beta)}{(\ln(1/\nu))^{1/\beta}\cdot N(L)^{1/(1+\beta)}}\right\}.$$

Since $L \leq ((\beta+1)N(L))^{1/(1+\beta)} \leq 2\cdot N(L)^{1/(1+\beta)}$ by Lemma 9,

$$\frac{\Lambda_2(L)}{N(L)} \quad\leq\quad \left(\overline{M}-\underline{M}\right)\cdot(b+h)\cdot\frac{24\tau}{\ln(1/\nu)\cdot N(L)^{\beta/(1+\beta)}}.$$

Combining the above two inequalities gives the desired result. □

4.4 Remarks Theorem 6 shows that the T-period expected running-average regret is $O\left(T^{-1/3}\right)$. The proof of Theorem 6 can easily be modified for other stochastic systems where the gradient depends on the steady-state distribution. We require, as in most papers in the online convex optimization literature, that the objective is convex with respect to the decision vector, the feasible set is a convex compact set, and the gradient of the objective function is bounded. Furthermore, the Markov chain obtained by fixing the decision vector displays the property that both the sample costs and the sample derivatives converge to their steady-state distributions, and that their convergence rates are exponential and independent of the decision vector (analogous to Theorem 5). Then the arguments in the proof of Theorem 6 also become applicable.

We explain the above generalization in more detail. Suppose S is a control parameter that we want to optimize. Let $X_t(S)$ denote the state vector of the system in period t, and let $X'_t(S)$ denote the sample derivative of $X_t(S)$ with respect to S. Let $\overline{X}(S) = \{(X_t(S), X'_t(S) : t \geq 1)\}$. Suppose that the following conditions are satisfied. (i) The feasible set $\mathcal{S}$ of S is convex and compact. (ii) For any $S \in \mathcal{S}$, $\overline{X}(S)$ is a Markov chain, and its state space belongs to a bounded set $\mathcal{M}$ independent of S. (iii) For any $S \in \mathcal{S}$, $\overline{X}(S)$ is ergodic, and the rate of convergence δ_t can be uniformly bounded by an exponentially decreasing

function, regardless of S and the initial state (analogous to Lemma A.1 and A.2). (iv) The average-cost criterion, denoted by $C(X_\infty(S))$, is convex with respect to S. (v) In period t, the manager can obtain the estimates for both the cost and the derivatives having biases whose magnitudes are no more than a multiple of δ_t (analogous to Theorem 5).

Conditions (i) and (iv) above ensure that the problem is a convex minimization problem over a compact set, a requirement for applying Theorem 7. This theorem, together with the bound on the bias of the derivative estimators in Condition (v), implies a result similar to Lemma 10. Meanwhile, Conditions (ii) and (iii) ensure that the stationary process exists for any choice of the control, and the convergence rate (mixing time) can be uniformly bounded for each cycle. These conditions, along with the bound on the bias of the cost estimators in Condition (v), imply a result similar to Lemma 11. Therefore, we establish a result analogous to Theorem 6, and show that the algorithm ADAPTIVE(α, β) can be easily adapted to result in the time-average regret of $O(T^{-1/3})$.

Finally, we remark that the objective in our paper is *regret* minimization where the demand distribution is fixed, albeit unknown, whereas the analysis of Flaxman et al. [4] is in a stronger bandit setting where the demand realizations are chosen by the adversary. We have chosen to model demand as a distribution for two reasons: (i) such a model would be closer to the demand models used in the inventory literature, and (ii) the stationarity of demand is crucial in establishing the Markov chain ergodicity which is a key component of our proof.

5. Conclusion In this paper, we have considered an adaptive control of replenishment quantities in a periodic-review inventory system with lost sales and a positive lead time. Contrary to the classical inventory literature, the manager does not know the demand distribution *a priori*, and only observes the *sales* data in each period. Under the long-run average-cost criterion, we have proposed an adaptive method such that its T-period average cost converges to the cost of the optimal base-stock policy, and we have shown the convergence rate of $O(1/T^{1/3})$. We achieve this by characterizing the ergodicity and the mixing time of the inventory system under a fixed base-stock policy. We believe that our adaptive method is applicable to other settings where the objective function is convex with respect to the control variable, and depends on the steady-state distribution of the system under consideration.

Acknowledgement We would like to thank the associate editor and the referees for helpful comments and suggestions that greatly improve the quality and the presentation of the paper. The research of the first and last authors was supported in part by the National Science Foundation through grants DMS-0732169 and DMS-0732196, respectively.

Appendix A. Proof of Theorem 3 In this section, we prove Theorem 3. We first show this result for the case where the starting inventory position at the beginning of period 1 is at most S (Lemma A.1 in Case I), and then extend the result to a general setting (Case II).

Case I: Initial Inventory Position is at Most S The main idea used in this section is that all the sample paths couple after a certain pattern or sequence of demands occurs. An example of such a demand pattern is the τ consecutive periods of zero demands, which results in the on-hand inventory of S units with no outstanding order regardless of the inventory vector before the pattern occurs. Yet, this particular example of zero demands may never occur depending on the distribution of demand. Another example of demand pattern, as we shall see, is as follows: the 2τ consecutive periods in each of which demand is at most $S/(\tau+1)$. This pattern of demands is used in the proof of Lemma A.1.

Lemma A.1 *If $\gamma(S) = \mathcal{P}[D \le S/(\tau+1)] > 0$, then the Markov chain $\overline{X}(S) = \{(X_t(S), X'_t(S)) : t \ge 1\}$ associated with the order-up-to-S policy is ergodic with a steady-state random vector $(X_\infty(S), X'_\infty(S))$. Moreover, for any $t \ge 2\tau+1$, any initial inventory vector $x_1 \in \Re_+^\tau$ satisfying $x_1 \cdot \mathbf{1}^\tau \le S$, and any $x'_1 \in \mathcal{N}^\tau$,*

$$\delta_{t+1}(S, x_1, x'_1) \quad \le \quad \left(1 - \gamma(S)^{2\tau}\right)^{t/(2\tau)}.$$

PROOF. We make use of the following result on the uniform ergodicity of Markov chains (see Meyn and Tweedie [14] for more details). We say a measurable set $\overline{\mathbf{U}} \subseteq \Re_+^\tau \times \mathcal{N}^\tau$ is a *small set* with respect to a nontrivial measure ν provided that there exists $t^* > 0$ such that for any $(x_1, x'_1) \in \overline{\mathbf{U}}$ and any measurable

set $B \times N \subseteq \Re_+^\tau \times \mathcal{N}^\tau$,

$$\mathcal{P}\left[(X_{t^*}(S), X'_{t^*}(S)) \in B \times N \;\middle|\; (X_1(S), X'_1(S)) = (x_1, x'_1)\right] \;\geq\; \nu\,(B \times N)\;.$$

The following result appears in Theorem 16.0.2 in Meyn and Tweedie [14]. If $\overline{\mathbf{U}}$ is a small set with respect to ν, then there exists stationary random variable $(X_\infty(S), X'_\infty(S))$ such that for any $(x_1, x'_1) \in \overline{\mathbf{U}}$ and $t \geq t^*$,

$$\delta_{t+1}\,(S, x_1, x'_1) \;\leq\; \left(1 - \nu\left(\Re_+^\tau \times \mathcal{N}^\tau\right)\right)^{t/(t^*-1)}.$$

To apply the above result, we let $\overline{\mathbf{U}} = \{x \in \Re_+^\tau \mid x \cdot \mathbf{1}^\tau \leq S\} \times \mathcal{N}^\tau$. We define a nontrivial measure ν such that $\overline{\mathbf{U}}$ is a small set with respect to this measure ν with $t^* = 2\tau + 1$, and $\nu\left(\Re_+^\tau \times \mathcal{N}^\tau\right) \geq \gamma(S)^{2\tau}$. Let the measure ν be defined on $\Re_+^\tau \times \mathcal{N}^\tau$ as follows. For any $0 \leq \ell \leq \tau - 1$, let $B_\ell \subseteq \Re_+$ be any measurable set and let

$$B \;=\; \left\{(q_{-1}, q_{-2}, \ldots, q_{-\tau+1}, i_0) \in \Re_+^\tau \;\middle|\; q_{-\ell} \in B_\ell \text{ for } 1 \leq \ell \leq \tau - 1, \text{ and } S - i_0 - \sum_{\ell=1}^{\tau-1} q_{-\ell} \in B_0\right\}.$$

(Note that $S - i_0 - \sum_{\ell=1}^{\tau-1} q_{-\ell}$ represents the order quantity associated with the state.) For any subset $N \subseteq \mathcal{N}^\tau$, define $\nu\,(B \times N)$ by

$$\nu(B \times N) \;=\; \gamma(S)^\tau \cdot \prod_{i=0}^{\tau-1} \mathcal{P}\left[D \in B_i \cap \left[0, \frac{S}{\tau+1}\right]\right] \cdot \mathbb{I}\left[(0, 0, \ldots, 0, 1) \in N\right].$$

From the above definition of ν, it is straightforward to verify that $\nu\left(\Re_+^\tau \times \mathcal{N}^\tau\right) = \gamma(S)^{2\tau} > 0$. Thus, to complete the proof, it remains to show that $\overline{\mathbf{U}}$ is a small set with respect to ν where $t^* = 2\tau + 1$. For any $1 \leq i \leq \tau - 1$, let $\widehat{B}_i = B_i \cap [0, S/(\tau+1)]$, and let $\widehat{B}$ be defined similarly to B, except that B_i's are replaced by $\widehat{B}_i$'s. It follows that

$$\begin{aligned}
&\mathcal{P}\left[\left(X_{2\tau+1}(S), X'_{2\tau+1}(S)\right) \in B \times N \;\middle|\; (X_1(S), X'_1(S)) = (x_1, x'_1)\right] \\
&\quad\geq\; \mathcal{P}\left[\left(X_{2\tau+1}(S), X'_{2\tau+1}(S)\right) \in \widehat{B} \times N \;\middle|\; (X_1(S), X'_1(S)) = (x_1, x'_1)\right]
\end{aligned}$$

From the definition of ν, it follows that $\nu\,(B \times N) = \nu\left(\widehat{B} \times N\right)$. Thus, it suffices to show that

$$\mathcal{P}\left[\left(X_{2\tau+1}(S), X'_{2\tau+1}(S)\right) \in \widehat{B} \times N \;\middle|\; (X_1(S), X'_1(S)) = (x_1, x'_1)\right] \geq \nu\left(\widehat{B} \times N\right).$$

To prove the above inequality, we can assume without loss of generality that $(0, \ldots, 0, 1) \in N$; otherwise, the definition of ν implies $\nu\left(\widehat{B} \times N\right) = 0$, and the result is trivially true. We consider the following demand pattern of length 2τ, where the demand in each of the first τ periods is at most $S/(\tau+1)$ and the demands in the next τ periods satisfy $D_{2\tau-\ell} \in \widehat{B}_i$ for each $\ell = 0, 1, \ldots, \tau - 1$. It is straightforward to verify that the probability of this event occurring is $\gamma(S)^\tau \cdot \prod_{i=0}^{\tau-1} \mathcal{P}\left[D \in \widehat{B}_i\right]$.

Claim A.1 *The above demand pattern implies that*

$$\begin{aligned}
X_{2\tau+1}(S) \;&=\; (Q_{2\tau}(S),\; \ldots,\; Q_{\tau+2}(S),\; I_{2\tau+1}(S)) \;=\; \left(D_{2\tau-1},\; \ldots,\; D_{\tau+1},\; S - \sum_{\ell=\tau+1}^{2\tau} D_\ell\right), \\
X'_{2\tau+1}(S) \;&=\; \left(Q'_{2\tau}(S),\; \ldots,\; Q'_{\tau+2}(S),\; I'_{2\tau+1}(S)\right) \;=\; (0, 0, \ldots, 0, 1)\,.
\end{aligned}$$

To prove this claim, note that since the initial inventory position $x_1 \cdot \mathbf{1}^\tau$ is less than or equal to S, we have that $Q_1(S) = S - X_1(S) \cdot \mathbf{1}^\tau$ for the first period, and $Q_{t+1}(S) = \min\{D_t,\; I_t(S)\}$ for all $t \geq 1$. This implies that for $1 \leq t \leq 2\tau$, $Q_{t+1}(S) \leq D_t \leq S/(\tau+1)$. We will now show that $Q_{t+1}(S) = D_t$ for $\tau + 1 \leq t \leq 2\tau$. Note that

$$D_{t-1} \;\geq\; Q_t(S) \;=\; S - X_t(S) \cdot \mathbf{1}^\tau \;=\; S - (Q_{t-\tau+1}(S) + Q_{t-\tau+2}(S) + \cdots + Q_{t-1}(S) + I_t(S))\;.$$

By rearranging the above inequality and using the fact that $Q_t(S) \leq S/(\tau+1)$ for all t, we obtain

$$\begin{aligned} I_t(S) &\geq S - (Q_{t-\tau+1}(S) + Q_{t-\tau+2}(S) + \cdots + Q_{t-1}(S)) - D_{t-1} \\ &\geq S - \frac{S(\tau-1)}{\tau+1} - \frac{S}{\tau+1} = \frac{S}{\tau+1}, \end{aligned}$$

and therefore $D_t \leq S/(\tau+1) \leq I_t(S)$. This implies that $Q_{t+1}(S) = \min\{D_t,\ I_t(S)\} = D_t$ for $\tau+1 \leq t \leq 2\tau$. Thus, in particular, $(Q_{2\tau}(S),\ Q_{2\tau-1}(S), \ldots, Q_{\tau+2}(S)) = (D_{2\tau-1},\ D_{2\tau-2}, \ldots, D_{\tau+1})$. Note that

$$\begin{aligned} I_{2\tau+1}(S) &= X_{2\tau+1}(S)\cdot\mathbf{1}^\tau - \sum_{\ell=\tau+2}^{2\tau} Q_\ell(S) = X_{2\tau+1}(S)\cdot\mathbf{1}^\tau - \sum_{\ell=\tau+1}^{2\tau-1} D_\ell \\ &= S - Q_{2\tau+1}(S) - \sum_{\ell=\tau+1}^{2\tau-1} D_\ell = S - D_{2\tau} - \sum_{\ell=\tau+1}^{2\tau-1} D_\ell, \end{aligned}$$

where the third equality follows from the fact that $Q_{2\tau+1}(S) = S - X_{2\tau+1}(S)\cdot\mathbf{1}^\tau$. The final equality follows from the fact that $Q_{2\tau+1}(S) = \min\{D_{2\tau},\ I_{2\tau}(S)\} = D_{2\tau}$. Moreover, since $D_t \leq I_t$ for $\tau+1 \leq t \leq 2\tau$, it follows from Theorem 1 that

$$Q'_{\tau+2}(S) = Q'_{\tau+3}(S) = \cdots = Q'_{2\tau+1}(S) = 0,$$

which implies (by Theorem 1) that $I'_{2\tau+1}(S) = 1 - \sum_{\ell=\tau+2}^{2\tau+1} Q'_\ell(S) = 1$, completing the proof of the claim.

Now, it follows from the claim that for $1 \leq \ell \leq \tau-1$, $Q_{2\tau+1-\ell}(S) = D_{2\tau-\ell} \in \widehat{B}_\ell$, and $S - X_{2\tau+1}(S)\cdot\mathbf{1}^\tau = D_{2\tau} \in \widehat{B}_0$. Thus, $X_{2\tau+1}(S) \in \widehat{B}$ and $X'_{2\tau+1}(S) \in N$. Since the particular demand pattern used in our proof has the probability of occurring of at least $\gamma(S)^\tau \cdot \prod_{i=0}^{\tau-1} \mathcal{P}\left[D \in \widehat{B}_i\right]$, it follows that

$$\begin{aligned} &\mathcal{P}\left[\left(X_{2\tau+1}(S), X'_{2\tau+1}(S)\right) \in \widehat{B}\times N \;\middle|\; (X_1(S), X'_1(S)) = (x_1, x'_1)\right] \\ &\geq \gamma(S)^\tau \cdot \prod_{i=0}^{\tau-1} \mathcal{P}\left[D \in \widehat{B}_i\right] = \nu\left(\widehat{B}\times N\right), \end{aligned}$$

which is the desired result. □

The pattern of demand used in the proof of Lemma A.1 has been carefully selected. In the first τ periods, demands are small such that sufficiently large quantities of inventory become available on-hand (as opposed to on-order) during the the second τ periods. The demands from period $\tau+1$ to 2τ are small enough that they do not cause any stock out, in order to ensure that the vector of outstanding orders in period $2\tau+1$ are defined in terms of these demands without censoring. The proof of Lemma A.1 is based on recognizing a set of demand patterns such that if such a pattern occurs, then all the sample paths will meet regardless of the state of the inventory vector before the demand pattern occurs. Such a demand pattern is called the "coalescing pattern", and has been used by Cooper and Tweedie [3] in the context of simulating an inventory system with age-dependent perishability.

We examine the bound given in the statement of Lemma A.1. If S is so small such that $\gamma(S) = \mathcal{P}\left[D \in [0,\ S/(\tau+1)]\right] = 0$, then this bound is equal to 1, and it is not meaningful. Otherwise, it converges to 0 exponentially with respect to t, and the convergence rate improves as $\gamma(S)$ increases, i.e., the base-stock S increases.

Case II: Initial Inventory Position Exceeds S We now extend the convergence result to the case where the initial inventory position may exceed S. We need the following lemma. Recall that $F(\cdot)$ denote the distribution function of D and $\mu = E[D]$.

Lemma A.2 *For any $\eta \in \Re$ and $t \geq 1$,*

$$\mathcal{P}\left[\sum_{\ell=1}^{t} D_\ell \leq \eta\right] \leq \begin{cases} F(\eta)^t, & \text{if } D \text{ has an infinite support} \\ e^{4\eta/\overline{D}} \cdot e^{-2t\mu^2/\overline{D}^2}, & \text{if } D \leq \overline{D} \text{ with probability one.} \end{cases}$$

Proof. It suffices to consider $\eta \geq 0$. If the demand has an infinite support, then

$$\mathcal{P}\left[\sum_{\ell=1}^{t} D_\ell \leq \eta\right] \leq \mathcal{P}\left[D_\ell \leq \eta \quad \text{for each } 1 \leq \ell \leq t\right] \leq F(\eta)^t.$$

If the demand is bounded above by $\overline{D}$, then it follows from Chernoff-Hoeffding's Inequality [6] that

$$\mathcal{P}\left[\sum_{\ell=1}^{t} D_\ell \le \eta\right] \;=\; \mathcal{P}\left[\sum_{\ell=1}^{t} (D_\ell - \mu) \le \eta - t\mu\right] \;\le\; \exp\left\{\frac{-2\,(\eta - t\mu)^2}{t\overline{D}^2}\right\}.$$

Since $\exp(\cdot)$ is an increasing function, and

$$\frac{-2\,(\eta - t\mu)^2}{t\overline{D}^2} \;=\; \frac{-2\eta^2 + 4\eta t\mu - 2t^2\mu^2}{t\overline{D}^2} \;\le\; 4\eta\mu/\overline{D}^2 - 2t\mu^2/\overline{D}^2 \;\le\; 4\eta/\overline{D} - 2t\mu^2/\overline{D}^2,$$

we obtain the required result. □

If the starting inventory position exceeds S, then under the order-up-to-S policy, the manager does not place any order until the inventory position falls below S, and the Markov chain states are transient. This result is shown in the following lemma. Recall $\mu = E[D]$.

Lemma A.3 *Consider an order-up-to-S policy, where $S \ge 0$. For any starting inventory vector $x_1 \in \Re_+^\tau$ and $t \ge \tau$,*

$$\mathcal{P}\left[X_t(S)\cdot\mathbf{1}^\tau > S \mid X_1(S) = x_1\right] \le \begin{cases} F\left(x_1\cdot\mathbf{1}^\tau - S\right)^{t-\tau}, & \text{if } D \text{ has an infinite support} \\ e^{4(x_1\cdot\mathbf{1}^\tau - S)/\overline{D}}\cdot e^{-2\mu^2(t-\tau)/\overline{D}^2}, & \text{if } D \le \overline{D} \text{ with probability one.} \end{cases}$$

Proof. Note that if the starting inventory position $x_1\cdot\mathbf{1}^\tau$ is at most S, then $X_t(S)\cdot\mathbf{1}^\tau \le S$ with probability one for all $t \ge 1$. Thus, the required result holds. We proceed by assuming otherwise, that is, $x_1\cdot\mathbf{1}^\tau > S$. By the description of the base-stock policy, $\max\{X_t(S)\cdot\mathbf{1}^\tau, S\} \le \max\{X_1(S)\cdot\mathbf{1}^\tau, S\}$ holds for any $t \ge 1$. Thus,

$$\begin{aligned} &\mathcal{P}[D_1 + D_2 + \cdots + D_{t-\tau} \;<\; X_1(S)\cdot\mathbf{1}^\tau - S] \\ &\quad= \;\mathcal{P}[D_\tau + D_{\tau+1} + \cdots + D_{t-1} \;<\; X_1(S)\cdot\mathbf{1}^\tau - S] \\ &\quad\ge \;\mathcal{P}[D_\tau + D_{\tau+1} + \cdots + D_{t-1} \;<\; X_\tau(S)\cdot\mathbf{1}^\tau - S], \end{aligned}$$

where the equality follows since demand distributions are independent and identically distributed. Also, for $t \ge \tau$, observe that

$$X_t(S)\cdot\mathbf{1}^\tau \;>\; S \quad\text{if and only if}\quad X_\tau(S)\cdot\mathbf{1}^\tau - (D_\tau + D_{\tau+1} + \cdots + D_{t-1}) \;>\; S.$$

Therefore, combining the above results,

$$\mathcal{P}\left[X_t(S)\cdot\mathbf{1}^\tau > S \;\middle|\; X_1(S) = x_1\right] \;\le\; \mathcal{P}[D_1 + D_2 + \cdots + D_{t-\tau} < x_1\cdot\mathbf{1}^\tau - S].$$

The desired result then follows immediately from Lemma A.2. □

We are now ready to prove Theorem 3. The proof of Theorem 3 combines Lemmas A.1 and A.3.

Proof. [Proof of Theorem 3] We will prove the result when the demand D has an infinite support. An analogous argument is applicable when D is bounded. If $x_1\cdot\mathbf{1}^\tau \le S$, the result follows directly from Lemma A.1. Thus, we proceed by assuming that $x_1\cdot\mathbf{1}^\tau > S$.

To facilitate our exposition, we fix the initial state (x_1, x_1'), and denote by $E_{(x_1,x_1')}[\cdot]$ and $\mathcal{P}_{(x_1,x_1')}[\cdot]$ expectation and probability that are conditioned on the event that $(X_1(S), X_1'(S)) = (x_1, x_1')$. By conditioning on the value of $X_{\lceil t/2\rceil}(S)$ and $X'_{\lceil t/2\rceil}(S)$ and applying the Markov property, it follows that, for any measurable set $B \subseteq \Re_+^\tau \times \mathcal{N}^\tau$,

$$\begin{aligned} &\mathcal{P}_{(x_1,x_1')}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B\right] \\ &\quad= \;E_{(x_1,x_1')}\left[\mathcal{P}_{(x_1,x_1')}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B \;\middle|\; X_{\lceil t/2\rceil}(S),\; X'_{\lceil t/2\rceil}(S)\right]\right] \\ &\quad= \;E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot\mathbf{1}^\tau \le S\right]\cdot\mathcal{P}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B \;\middle|\; X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right]\right] \\ &\qquad+E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot\mathbf{1}^\tau > S\right]\cdot\mathcal{P}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B \;\middle|\; X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right]\right]. \end{aligned}$$

Therefore, for any measurable set $B \subseteq \Re_+^\tau \times \mathcal{N}^\tau$, we have

$$\begin{aligned}
&\mathcal{P}_{(x_1,x_1')}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B\right] - \mathcal{P}\left[(X_\infty(S), X'_\infty(S)) \in B\right] \\
&\quad = \quad E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau \le S\right]\cdot \Delta(B)\right] + E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau > S\right]\cdot \Delta(B)\right],
\end{aligned}$$

where

$$\Delta(B) = \mathcal{P}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B \;\middle|\; X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right] - \mathcal{P}\left[(X_\infty(S), X'_\infty(S)) \in B\right].$$

The random variable $|\Delta(B)|$, however, is bounded above almost surely by $\delta_{t-\lceil t/2\rceil+2}\left(S, X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right)$ by the definition of $\delta_{t-\lceil t/2\rceil+2}(\cdot)$, and is also bounded above by 1. Therefore, we obtain

$$\begin{aligned}
&\left|\mathcal{P}_{(x_1,x_1')}\left[\left(X_{t+1}(S), X'_{t+1}(S)\right) \in B\right] - \mathcal{P}\left[(X_\infty(S), X'_\infty(S)) \in B\right]\right| \\
&\quad \le \quad E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau \le S\right]\cdot \delta_{t-\lceil t/2\rceil+2}\left(S, X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right)\right] \\
&\qquad + \mathcal{P}_{(x_1,x_1')}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau > S\right] .
\end{aligned}$$

We provide an upper bound on each term of the right-hand side of the above inequality. By Lemma A.1, the first term satisfies

$$\begin{aligned}
&E_{(x_1,x_1')}\left[\mathbb{I}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau \le S\right]\cdot \delta_{t-\lceil t/2\rceil+2}\left(S, X_{\lceil t/2\rceil}(S), X'_{\lceil t/2\rceil}(S)\right)\right] \\
&\quad \le \quad \mathcal{P}_{(x_1,x_1')}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau \le S\right]\cdot \left(1-\gamma(S)^{2\tau}\right)^{(t-\lceil t/2\rceil+1)/(2\tau)} \\
&\quad \le \quad \left(1-\gamma(S)^{2\tau}\right)^{(t-\lceil t/2\rceil+1)/(2\tau)} \\
&\quad \le \quad \left(1-\gamma(S)^{2\tau}\right)^{t/(4\tau)},
\end{aligned}$$

where the last inequality follows from the fact that $t/(4\tau) \le (t-\lceil t/2\rceil+1)/(2\tau)$. Furthermore, by Lemma A.3, the second term satisfies

$$\mathcal{P}_{(x_1,x_1')}\left[X_{\lceil t/2\rceil}(S)\cdot \mathbf{1}^\tau > S\right] \;\le\; F\left(x_1\cdot \mathbf{1}^\tau - S\right)^{\lceil t/2\rceil-\tau} \;\le\; F\left(x_1\cdot \mathbf{1}^\tau - S\right)^{\frac{t}{2}-\tau}.$$

Therefore, we obtain the required result from the definition of $\delta_{t+1}(S, x_1, x_1')$. □

Appendix B. Proof of Theorem 5 We first establish the following lemma before we prove Theorem 5.

Lemma B.1 *Under the conditions of Theorem 5,*

(i) $\left|E\left[I_t(S)-d\right]^+ - E\left[I_\infty(S)-d\right]^+\right| \;\le\; \max\left\{S, x_1\cdot \mathbf{1}^\tau\right\}\cdot \delta_t\left(S, x_1, x_1'\right)$, *for any* d.

(i) $\left|d - E\left[I_t(S)\right]^+ - E\left[d - I_\infty(S)\right]^+\right| \;\le\; \max\left\{S, x_1\cdot \mathbf{1}^\tau\right\}\cdot \delta_t\left(S, x_1, x_1'\right)$, *for any* d.

(iii) $\left|E\left[\mathbb{I}\left[D < I_t(S)\right]\cdot I'_t(S)\right] - E\left[\mathbb{I}\left[D < I_\infty(S)\right]\cdot I'_\infty(S)\right]\right| \;\le\; \delta_t\left(S, x_1, x_1'\right)$.

(iv) $\left|E\left[\mathbb{I}\left[D \ge I_t(S)\right]\cdot I'_t(S)\right] - E\left[\mathbb{I}\left[D \ge I_\infty(S)\right]\cdot I'_\infty(S)\right]\right| \;\le\; \delta_t\left(S, x_1, x_1'\right)$.

Proof. To prove part (i), note that by definition of order-up-to-S policy, both random variables $I_t(S)$ and $I_\infty(S)$ are bounded above by $\max\left\{S, x_1\cdot \mathbf{1}^\tau\right\}$ with probability one. Since they are nonnegative random variables, it follows that, for any $d \ge 0$,

$$\begin{aligned}
\left|E\left[I_t(S)-d\right]^+ - E\left[I_\infty(S)-d\right]^+\right| \quad &= \quad \left|\int_0^{\max\{S, x_1\cdot \mathbf{1}^\tau\}} \left\{\mathcal{P}\left[I_t(S)-d > z\right] - \mathcal{P}\left[I_\infty(S)-d > z\right]\right\} dz\right| \\
&\le \quad \int_0^{\max\{S, x_1\cdot \mathbf{1}^\tau\}} \left|\mathcal{P}\left[I_t(S) > z+d\right] - \mathcal{P}\left[I_\infty(S) > z+d\right]\right| dz \\
&\le \quad \int_0^{\max\{S, x_1\cdot \mathbf{1}^\tau\}} \delta_t\left(S, x_1, x_1'\right) dz \\
&= \quad \max\left\{S, x_1\cdot \mathbf{1}^\tau\right\}\cdot \delta_t\left(S, x_1, x_1'\right),
\end{aligned}$$

where the second inequality follows from the definition of $\delta(S, x_1, x_1')$, establishing (i). Similarly, (ii) holds.

Now, we prove (iii). By Theorem 1, $I_t'(S)$ is binary. It follows that $I_\infty'(S)$ is also binary with probability one. (To see this, suppose there exists a measurable set $\hat{B} \subseteq \Re_+ \setminus \{0,1\}$ such that $\mathcal{P}[I_\infty'(S) \in \hat{B}] > 0$. Then, let $B = \left\{ (q_{-1}, \ldots, q_{-\tau+1}, i_0,\ q_{-1}', \ldots, q_{-\tau+1}', i_0') \in \Re_+^\tau \times \mathcal{N}^\tau \ \middle|\ i_0' \in \hat{B} \right\}$. Thus, $\mathcal{P}\left[(X_\infty(S), X_\infty'(S)) \in B\right] > 0$, but $\mathcal{P}\left[(X_t(S), X_t'(S)) \in B\right] = 0$ for each $t \geq 1$. Therefore, $\delta_t\left(S, x_1, x_1'\right)$ does not converge to 0 as $t \to \infty$, contradicting Theorem 3.)

For any value of $D = d$, let

$$B(d) \ = \ \left\{ (q_{-1}, \ldots, q_{-\tau+1}, i_0,\ q_{-1}', \ldots, q_{-\tau+1}', i_0') \in \Re_+^\tau \times \mathcal{N}^\tau \ \middle|\ i_0 > d,\ i_0' = 1 \right\} .$$

Thus, for any fixed value of $D = d$,

$$\begin{aligned} & E\left[\mathbb{I}\left[d < I_t(S)\right] \cdot I_t'(S)\right] - E\left[\mathbb{I}\left[d < I_\infty(S)\right] \cdot I_\infty'(S)\right] \\ &= \ E\left[\mathbb{I}\left[d < I_t(S),\ I_t'(S) = 1\right]\right] - E\left[\mathbb{I}\left[d < I_\infty(S),\ I_\infty'(S) = 1\right]\right] \\ &= \ \mathcal{P}\left[d < I_t(S),\ I_t'(S) = 1\right] - \mathcal{P}\left[d < I_\infty(S),\ I_\infty'(S) = 1\right] \\ &= \ \mathcal{P}\left[(X_t(S), X_t'(S)) \in B(d)\right] - \mathcal{P}\left[(X_\infty(S), X_\infty'(S)) \in B(d)\right] . \end{aligned}$$

The absolute value of the above expression is bounded above by $\delta_t\left(S, x_1, x_1'\right)$. By taking the expectation with respect to D, we establish (iii). Similarly, (iv) holds. □

Let us now prove Theorem 5.

Proof.

Note that by definition of $C(\cdot)$,

$$\begin{aligned} C\left(I_t(S)\right) &= h \cdot E\left[\left(I_t(S) - D\right)^+\right] + b \cdot E\left[\left(D - I_t(S)\right)^+\right] \\ C\left(I_\infty(S)\right) &= h \cdot E\left[\left(I_\infty(S) - D\right)^+\right] + b \cdot E\left[\left(D - I_\infty(S)\right)^+\right] . \end{aligned}$$

It follows from Lemma B.1 (i) and (ii) that

$$\begin{aligned} & \left|C\left(I_t(S)\right) - C\left(I_\infty(S)\right)\right| \\ &\leq \ h \cdot \left| E\left[I_t(S) - D\right]^+ - E\left[I_\infty(S) - D\right]^+ \right| + b \cdot \left| E\left[\left(D - I_t(S)\right)^+\right] - E\left[\left(D - I_\infty(S)\right)^+\right] \right| \\ &\leq \ (h + b) \cdot \max\left\{S, x_1 \cdot \mathbf{1}^\tau\right\} \cdot \delta_t\left(S, x_1, x_1'\right) \end{aligned}$$

which proves the first inequality.

Now, from Section 3.3, recall

$$\begin{aligned} \frac{d}{dS} C\left(I_t(S)\right) &= h \cdot E\left[\mathbb{I}\left[D < I_t(S)\right] \cdot I_t'(S)\right] - b \cdot E\left[\mathbb{I}\left[D \geq I_t(S)\right] \cdot I_t'(S)\right], \qquad \text{and} \\ \frac{d}{dS} C\left(I_\infty(S)\right) &= h \cdot E\left[\mathbb{I}\left[D < I_\infty(S)\right] \cdot I_\infty'(S)\right] - b \cdot E\left[\mathbb{I}\left[D \geq I_\infty(S)\right] \cdot I_\infty'(S)\right]. \end{aligned}$$

Thus,

$$\begin{aligned} & \left| \frac{d}{dS} C\left(I_t(S)\right) - \frac{d}{dS} C\left(I_\infty(S)\right) \right| \\ &\leq \ h \cdot \left| E\left[\mathbb{I}\left[D < I_t(S)\right] \cdot I_t'(S)\right] - E\left[\mathbb{I}\left[D < I_\infty(S)\right] \cdot I_\infty'(S)\right] \right| \\ &\quad + \ b \cdot \left| E\left[\mathbb{I}\left[D \geq I_t(S)\right] \cdot I_t'(S)\right] - E\left[\mathbb{I}\left[D \geq I_\infty(S)\right] \cdot I_\infty'(S)\right] \right| \\ &\leq \ (h + b) \cdot \delta_t\left(S, x_1, x_1'\right) . \end{aligned}$$

where the first inequality above follows from Lemma B.1 (iii) and (iv). □

Appendix C. Proof of Lemma 8 and Lemma 9 We first prove Lemma 8.

Proof. The first inequality follows easily from $\rho \in (0,1)$ and $\lceil k^\beta \rceil \geq k^\beta$. For the second inequality, observe $\sum_{k=1}^{L} \rho^{k^\beta} \leq \int_{u=0}^{L} \rho^{u^\beta} du$. Using the substitution $y = u^\beta$, obtain

$$\int_{u=0}^{L} \rho^{u^\beta} du \ = \ \frac{1}{\beta} \int_{y=0}^{L^\beta} y^{-(\beta-1)/\beta} \rho^y dy$$

$$= \frac{1}{\beta}\int_{y=0}^{L^\beta} y^{1/\beta-1}\exp\left(\frac{-y}{-1/\ln\rho}\right)dy$$
$$= \frac{\Gamma(1/\beta)\cdot(-1/\ln\rho)^{1/\beta}}{\beta}\cdot\int_{y=0}^{L^\beta}\frac{(-1/\ln\rho)^{-1/\beta}}{\Gamma(1/\beta)}\cdot y^{1/\beta-1}\exp\left(\frac{-y}{-1/\ln\rho}\right)dy\ .$$

Here, the integral corresponds to the cumulative distribution of a gamma distribution at L^β, where the gamma distribution has the shape parameter $1/\beta$ and the scale parameter $-1/\ln\rho$. Since the cumulative density is at most 1, it follows that the above expression is at most $\beta^{-1}\cdot\Gamma(1/\beta)\cdot(-1/\ln\rho)^{1/\beta}$. □

We now prove Lemma 9.

Proof. Observe that

$$N(k) = \sum_{k'=1}^{k}\lceil k'^\beta\rceil \geq \sum_{k'=1}^{k}k'^\beta \geq \int_{u=0}^{k}u^\beta du = \left.\frac{u^{\beta+1}}{\beta+1}\right|_0^k = \frac{k^{\beta+1}}{\beta+1}\ .$$

Therefore, $N(k)\geq k^{\beta+1}/(\beta+1)$, which implyies part (i). From above, we also obtain

$$N(k) \geq \frac{k^{\beta+1}}{\beta+1} = \frac{k\cdot k^\beta}{\beta+1} \geq \frac{k\cdot(\lceil k^\beta\rceil-1)}{\beta+1}\ ,$$

which implies that $k\cdot\lceil k^\beta\rceil \leq (\beta+1)N(k)+k$. Thus, from part (i), we obtain $k\cdot\lceil k^\beta\rceil \leq (\beta+1)\cdot N(k)+[(\beta+1)\cdot N(k)]^{1/(\beta+1)}$, which in turn implies part (ii).

For (iii), observe that $\lceil k^\beta\rceil\leq k^\beta+1$. From part (i), it follows

$$\lceil k^\beta\rceil \leq 1+k^\beta \leq 1+[(\beta+1)\cdot N(k)]^{\beta/(\beta+1)}\ . \tag{3}$$

Since $N(k)\geq 1$, we obtain part (iii). To prove part (iv), consider $f(u)=u^\alpha$ where $\alpha\in(0,1)$. Since f is a concave function,

$$(1+u)^\alpha = f(1+u) \leq f(u)+1\cdot f'(u) = u^\alpha+\alpha\cdot u^{\alpha-1} \leq u^\alpha+1$$

for $u\geq 1$. Apply the above inequality to $u=[(\beta+1)\cdot N(k)]^{\beta/(\beta+1)}\geq 1$, to obtain

$$(1+[(\beta+1)\cdot N(k)]^{\beta/(\beta+1)})^\alpha \leq [(\beta+1)\cdot N(k)]^{\alpha\beta/(\beta+1)}+1\ .$$

Then, (3) implies $\lceil k^\beta\rceil^\alpha \leq 1+[(\beta+1)\cdot N(k)]^{\alpha\beta/(\beta+1)}$, from which we obtain part (iv). □

References

[1] N. Agrawal and S. A. Smith. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics*, 43:839–861, 1996.

[2] A. N. Burnetas and C. E. Smith. Adaptive ordering and pricing for perishable products. *Operations Research*, 48(3):436–443, 2000.

[3] W. L. Cooper and R. L. Tweedie. Perfect simulation of an inventory model for perishable products. *Stochastic Models*, 18, 2002.

[4] A. D. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. *Working Paper*, 2004.

[5] G. A. Godfrey and W. B. Powell. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science*, 47:1101–1112, 2001.

[6] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[7] W. T. Huh, G. Janakiraman, J. Muckstadt, and P. Rusmevichientong. Asymptotic optimality of order-up-to policies in lost sales inventory systems. *Forthcoming in Management Science*, 2006.

[8] W. T. Huh and P. Rusmevichientong. Adaptive capacity allocation with censored demand data: Application of concave umbrella functions. *Working Paper*, 2006.

[9] W. T. Huh and P. Rusmevichientong. A non-parametric asymptotic analysis of inventory planning with censored demand. *Forthcoming in Mathematics of Operations Research*, 2006.

[10] G. Janakiraman and R.O. Roundy. Lost-sales problems with stochastic lead times: Convexity results for base-stock policies. *Operations Research*, 52:795–803, 2004.

[11] S. Karlin and H. Scarf. Inventory models of the arrow-harris-marschak type with time leg. In K. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the Mathematical Theorey of Inventory and Production*. 1958.

[12] Robert Kleinberg. Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 2004.

[13] R. Levi, G. Janakiraman, and M. Nagarajan. Provably near-optimal balancing policies for stochastic inventory control models with lost sales. *Mathematics of Operations Research*, 33(2):351–374, 2008.

[14] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.

[15] T. Morton. The near-myopic nature of the lagged-proportional-cost inventory problem with lost sales. *Operations Research*, 19, 1971.

[16] T. E. Morton. Bounds on the solution of the lagged optimal inventory equation with no demand backlogging and proportional costs. *SIAM Review*, 11(4):572–596, 1969.

[17] S. Nahmias. Demand estimation in lost sales inventory systems. *Naval Research Logistics*, 41:739–757, 1994.

[18] W. Powell, A. Ruszczynski, and H. Topaloglu. Learning algorithms for separable approximations of discrete stochastic optimization problems. *Mathematics of Operations Research*, 29(4):814–836, 2004.

[19] M. Reiman. A new and simple policy for the continuous review lost sales inventory model. *Working Paper*, 2004.

[20] Martin Zinkevich. Online convex programming and generalizaed infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, DC, 2003.

[21] P. Zipkin. Old and new methods for lost-sales inventory systems. *Forthcoming in Operations Research*, 2006.

[22] P. Zipkin. On the structure of lost-sales inventory models. *Operations Research*, 56(4):937–944, 2008.