Data-Driven Algorithms for Stochastic Supply Chain Systems: Approximation and Online Learning

by

Huanan Zhang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in The University of Michigan
2017

Doctoral Committee:

Professor Xiuli Chao, Co-Chair
Assistant Professor Cong Shi, Co-Chair
Professor Jon Lee
Assistant Professor Viswanath Nagarajan
Assistant Professor Stefanus Jasin

Huanan Zhang
zhanghn@umich.edu
ORCID iD: 0000-0002-0672-5227

ACKNOWLEDGEMENTS

First of all, I would like to thank my dissertation co-chairs Prof. Xiuli Chao and Prof. Cong Shi for their time and effort in guiding me through this Ph.D. journey. Without their support and help, this dissertation would not have been possible. I would also like to thank my committee members Prof. Jon Lee, Prof. Viswanath Nagarajan and Prof. Stefanus Jasin, for their helpful discussions and feedback.

My gratitude also goes to professors in the IOE department from whose lectures I have learned methodologies and techniques to conduct my research. And I also owe my thanks to staff members of the department for their assistance and help.

I appreciate the friendship from Prof. Xiuli Chao and Prof. Cong Shi's research group, which includes former members Dr. Xiting Gong, Dr. Jingchen Wu, Dr. Majid Al-Gwaiz, Dr. Boxiao Chen and current members Sentao Miao, Yuchen Jiang, Weidong Chen and Hao Yuan. Our motivative discussions sparked a lot of research ideas and helped me to be more productive during my Ph.D. study. I would also like to thank all my other friends in the IOE department for their caring and sharing, especially my friend Dr. Yuhui Shi and Duyi Li.

Finally, I would like to thank my wife Weiran for her persistent support which has been the best source of energy to keep me moving forward, not only in pursuing my Ph.D., but also in life.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

CHAPTER I

Introduction

In the era of Big Data, with new and emerging technologies such as the radio-frequency identification (RFID), point-of-sale systems (POS) and face recognition systems, many firms are able to gather a variety of data with very high frequency. Indeed, 90% of the data in the world today has been created in the last two years alone.[1] However, Big Data is not just about Big Data, but is about how to extract useful and insightful information from data. Acquiring data is only the first step for the firm. The second and more important step is to effectively integrate the data through learning process (mining the data) in the decision-making process, and to utilize the information extracted from data to improve the efficiency of the firm's supply chain operation.

Traditional supply chain management often assumes that the uncertainties in the system (e.g., random demands, random capacities) are well-defined probability distributions or stochastic processes that are known to the decision maker *a priori*, and the main focus is to solve the corresponding (multi-stage) stochastic optimization model. However, in many practical scenarios, correctly specifying the distributional information on these uncertainties in the system is usually very hard and sometimes impossible. One of the major challenges encountered is that the collected data is affected by the operational decisions by the decision maker, which then affect the decision maker's understanding of the underlying system in making new operational decisions. That is, in most data-driven optimization problems in supply chains, the observed data and the operational decisions are *inextricably linked*. Running an unthoughtful operational policy may lead to the so-called *spiral-down effects*, where both the quality of data collected and operational decisions deteriorate over time. This motivates us to ponder over how to design effective data-driven policies that can balance the trade-off between exploration (learning) and exploitation (earning) for optimizing

[1]Source: IBM, `https://www-01.ibm.com/software/data/bigdata/what-is-big-data.html` accessed September 20, 2016.

supply chains. It should be noted that, due to the fact that its system dynamics are usually complex and each operational decision tends to have a long-term effect on the outcome standard online learning algorithms in general cannot be directly applied or adapted to a supply chain system.

The primary focus of this dissertation is on *multistage stochastic optimization problems* arising in the context of supply chains and inventory control, and on the design of efficient algorithms to solve the respective models. This dissertation can be categorized into two broad areas as follows.

First, in Chapter 2 and 3, we address two challenging stochastic inventory control models, where the current decision has a long-term effect on future costs. We assume that the decision maker *has no demand distribution information available a priori* and can only observe past realized sales (censored demand) information to optimize the system's performance on the fly.

In Chapter 2, we focus on the perishable inventory system. Perishable products are ubiquitous and an indispensable part of our society. Examples of perishable products include meat, fruit, vegetable, dairy products, and frozen foods in the supermarket, pharmaceuticals like drugs and vitamins, and the blood products in the blood bank. Study of stochastic inventory systems with perishable products has long been a significant yet challenging topic in the literature of operations management. Due to the inherent structural complexity, finding the optimal policies is computationally intractable even for the most basic model. Motivated by the studies in the literature showing that base-stock policies perform near-optimal in these systems, we focus on finding the best base-stock policy. For this problem, we develop the first nonparametric learning algorithm called the Cycle-Update Policy (CUP). The CUP has a square-root convergence rate compared with the best base-stock policy, which is the best possible rate in the online learning literature without further assumptions.

In Chapter 3, we consider the periodic-review inventory control problem with lost-sales and positive lead times, which is one of the most fundamental yet notoriously difficult problems in the theory of inventory management. Even with complete information about the demand distribution, it is well-known that the optimal policy does not possess a simple form. From the literature, we know that the base-stock policy is asymptotically optimal for this problem, and numerically it also performs very close to the optimal cost. For this problem, Huh et al. (2009a) developed a learning algorithm with average regret $O(1/\sqrt[3]{T})$, which is not tight since the known lower bound is $O(1/\sqrt{T})$. To close the gap between the upper bound and the lower bound of regret, we develop the Simulated Cycle-Update Policy (SCU), which closed this gap. Through extensive numerical experiments, the SCU is consistently

performing better than the learning algorithm in Huh et al. (2009a).

Second, in Chapter 4, we study perishable inventory systems. Different from traditional perishable inventory literatures, we allow demands to be arbitrarily correlated and non-stationary, which means we can capture the seasonality nature of the economy, and allow the decision makers to effectively incorporate demand forecast, such as advance demand information (ADI), martingale models of forecast evolution (MMFE), autoregressive moving average (ARMA) demand models, and Markov modulated demand (MMD) process, among others. The goal is to minimize the total expected cost, with known demand distribution. we start by considering the base model with no lead time, no setup cost and the capacities are assumed to be infinity. For this base model, even with i.i.d. demands, obtaining the optimal solutions is intractable due to the well-known "curse-of-dimensionality". For this problem, we develop two approximation algorithms with worst-case performance guarantees. Through comprehensive numerical experiments, we have shown that the numerical performances of the approximation algorithms are very close to optimal.

CHAPTER II

Nonparametric Learning Algorithms for Optimal Base-Stock Policy in Perishable Inventory Systems with Censored Demand

2.1 Introduction

Perishable products are undoubtedly an indispensable part of our lives. For example, perishable products such as meat, fruit, vegetable, dairy products, and frozen foods constitute the majority of supermarket sales. Moreover, virtually all pharmaceuticals belong to the category of perishable products. Perhaps the most frequently discussed applications of perishable inventory models are inventory control of blood products in blood banks (see, e.g., Prastacos (1984)). In Cooper (2001), the author considers an air cargo management problem in commercial airlines, and shows that it is also an interesting (and surprising) example of perishable inventory control problem.

Since as early as 1960s, the study of stochastic inventory systems with perishable products has been a significant yet challenging topic in the literature of operations management (see the survey articles Karaesmen et al. (2011), Nahmias (2011), and a series of more recent works (Li et al. (2009), Deniz et al. (2010), Chen et al. (2014b), Li and Yu (2014), Chao et al. (2015b,a), Zhang et al. (2015)). To this date, most, if not all, of the papers on stochastic perishable inventory systems assume that the stochastic future demand processes are given as an input to the models, and the inventory replenishment decisions are made with full knowledge of the future demand distribution. However, in practice, the underlying demand distribution may not be available to the firm *a priori*. This raises a natural and important research question as to how to learn the underlying demand distribution while minimizing the total expected costs on the fly.

2.1.1 Model, Motivation, and Research Question

Consider a periodic-review stochastic inventory systems with perishable products. The product lifetime m is known and fixed. The demands across periods $t = 1, 2 \ldots$ are denoted by i.i.d. random variables D_t, $t = 1, 2 \ldots$. The firm makes a replenishment decision at the beginning of each period t, and then demand is realized and is satisfied to the maximum extent from the on-hand inventory. We consider the class of First-In-First-Out (FIFO) issuing policies, i.e., the oldest inventory is consumed first when demand arrives (see Karaesmen et al. (2011)). Any demand that cannot be satisfied immediately with the on-hand inventory leads to lost sales, while non-expired excess inventory at the end of a period is carried over to the next period. At the end of each period, besides the typical inventory holding cost and lost sales penalty cost, an inventory outdating cost is incurred, and that is proportional to the amount of inventory units that reach the end of their lifetimes.

However, contrary to the classical perishable inventory literature, the underlying demand distribution D_t is not known to the firm *a priori*. Instead, the firm makes ordering decisions based on observed past sales, which is the minimum of the realized demand and the on-hand inventory. In other words, the sales data are *censored* by the available inventory level, and the firm cannot observe the lost-sales quantity. We note that joint learning and optimization problems under censored demand information for non-perishable inventory systems have received much attention in the research literature and are often challenging to analyze (see Huh and Rusmevichientong (2009), Huh et al. (2009a, 2011), Besbes and Muharremoglu (2013), Shi et al. (2015), Chen et al. (2015b) for more discussions.)

Even with complete information about the demand distribution *a priori*, it is well-known that the (clairvoyant) optimal policy for perishable inventory systems does not have any simple structure (see Nahmias (1975b) and Fries (1975)), and computing the exact optimal policy is intractable using brute-force dynamic programming. Nandakumar and Morton (1993) studied these systems and noted that "*Since base stock policies are easier to implement and widely used in practice, the interest quickly turned to analyzing such policies for this problem*". Indeed, a number of authors have investigated the performance of base-stock, or fixed critical number, policies for perishable inventory systems. For example, Cooper (2001) commented on this stream of research that *"The complexity of optimal policies, as well as the difficulties involved in computing them, has led many authors to analyze heuristic methods for controlling perishable inventories. One such method, proposed by Chazan and Gal (1977), Cohen (1976), Nahmias (1976), is the fixed-critical number order policy, in which orders are placed so that the total amount of inventory is the same at the end of each time period,*

regardless of the ages of the inventory. Computational studies by Nahmias (1976, 1977c) and Nandakumar and Morton (1993) show that under a variety of different assumptions, such as fixed-critical number policies can be quite good when compared with other methods, as well as to optimal policies. In addition, these policies provide a good baseline against when other types of policies can be compared." Further theoretical and computational evidences on the effectiveness of base-stock policies have been reported in Nahmias (1978), Deniz et al. (2010) and Cooper (2001). In particular, after conducting comprehensive numerical tests, Cooper (2001) summarized that

"*... in all cases, the performance of the critical-number policies was nearly as good as that of an optimal policy, thereby supporting the assertion that, in the absence of significant fixed-charge order costs, critical-number policies provide a simple and effective means for managing inventories of a perishable product*".

These studies motivate us to *develop learning algorithms to find the best base-stock policies* for perishable inventory systems. Since the best base-stock policy performs very close to the real optimal policy, we shall use the long-run average cost of the clairvoyant best base-stock policy (had the distribution been known *a priori*) as our benchmark. We note that this choice of benchmark is similar in spirit to Huh et al. (2009a), which finds the best base-stock policy for the lost-sales inventory control problem with positive lead times (which is another notoriously difficult problem in stochastic inventory theory).

We aim to develop a nonparametric closed-loop control policy $\pi(S_t)$ for computing a *period-dependent* base-stock level S_t in each period t with unknown demand distribution *a priori* and censored demand information. Now, had the firm known the underlying demand distribution *a priori*, there exists a clairvoyant optimal base-stock policy $\pi(S^*)$. We measure the performance of our proposed policy $\pi(S_t)$ through the notion of *regret*, the difference between the T-period average cost of running our adaptive policy $\pi(S_t)$ and the long-run average cost of the clairvoyant optimal base-stock policy $\pi(S^*)$. The main research question is to devise an effective nonparametric data-driven policy $\pi(S_t)$ that drives the average regret per period to zero with a provable (and tight) convergence rate.

2.1.2 Main Results and Contribution

The main result of this chapter is to present the first nonparametric learning algorithm for periodic-review perishable inventory systems. As seen from our literature review, this class of problems is fundamental in inventory theory that has challenged researchers for decades.

Different than the conventional perishable inventory literature, we assume that the firm

does not know the demand distribution *a priori* but makes adaptive ordering decision in each period based only on the past sales (censored demand) data. Motivated by theoretical and computational results showing that the class of base-stock policies performs near-optimal in these systems, we focus on finding the best base-stock policy. In what follows, we summarize our main results and contribution.

1) Algorithms. We develop a nonparametric adaptive inventory control policy, called the cycle-update policy (CUP for short), for the perishable inventory system with lost-sales and censored demand. Our CUP algorithm is a stochastic gradient descent type of algorithm (see Burnetas and Smith (2000), Huh and Rusmevichientong (2009), Huh et al. (2009a), Shi et al. (2015)). There are, however, several points of departure from the aforementioned literature:

(a) First, as the name suggests, our CUP algorithm updates base-stock level in each cycle, not in each period, and our updating cycles are not *a priori* fixed but are triggered sequentially by lost-sales events as demand realizes over time, which is uniquely designed for our perishable inventory system.

(b) Second, when we update base-stock level at the beginning of a cycle, computing the (sample-path) gradient for the total costs accrued during the preceding cycle is a non-trivial task. The difficulty is caused by the inter-dependence between the base-stock level and the amount of outdates during a cycle. We develop a subroutine in §2.4.2 and show that it outputs the correct gradient of total costs of a cycle with respect to the base-stock level (Proposition 2.5). The main idea underlying our subroutine is that when we perturb the current base-stock level S by an infinitesimal amount δ, we count how many of these additional orders δ eventually outdate within the cycle by using an auxiliary vector (other than the inventory vector) to keep track of the remaining lifetime of the δ additional units.

2) Performance analysis. Our main theoretical result Theorem 2.8 is that the average regret per period of CUP converges to zero at the rate of $O(1/\sqrt{T})$, and also in Theorem 2.12 that, under some additional technical condition, the rate can be improved to $O(\log T/T)$. These rates cannot be improved from theory of online learning (see Hazan (2015)).

Next we present the main features of our performance analysis.

(c) The common approach in regret analysis of online convex optimization is to compare the costs in each period between the learning algorithm and the clairvoyant optimal policy. In our setting, however, comparing costs in each period is *not* helpful, and we instead

compare cycle costs. Naively defining the cycles of the two policies in similar manner will lead to different number of cycles and different cycle lengths for the two policies. Hence, the cycles for the two systems have to be coupled in such a way that they can be compared. We define the cycles, for both systems, using the successive periods stockouts occur in the system operating under the CUP learning algorithm. Since the number of periods in each cycle is random, we have a random number of cycles, leading to some technical difficulties that do not exist for the standard online optimization problems.

(d) Since CUP updates base-stock levels in cycles, in the regret analysis we need to compare and bound the total costs within a cycle between CUP and the clairvoyant optimal policy $\pi(S^*)$. However, one difficulty arises at the start of each cycle because the two policies considered have different inventory age distributions. In particular, CUP has zero initial inventory (due to a lost-sales event in the preceding period) and therefore all the inventory units ordered in that period are brand new with remaining lifetime m. But the age distribution of the optimal base-stock policy can be very different. To tackle this difficulty, we introduce a new bridging policy, called the replacement of old inventories (ROI for short), between CUP and the optimal base-stock policy. For each sample path, similar to the optimal base-stock policy, the bridging policy ROI uses S^* as its base-stock level, but at the beginning of each cycle, ROI replaces all its inventory units (regardless of their ages) with brand new inventory units (thus all having remaining lifetime m). We establish in Proposition 2.9 that for each sample path, the total cost incurred by ROI actually provides a lower bound on the total cost incurred by the optimal base-stock policy $\pi(S^*)$. The analysis of Theorem 2.8 crucially relies on this intermediate result.

(e) In general, the extension from convex case to strongly convex case is quite straightforward (see, e.g., Huh and Rusmevichientong (2009)). However, this extension is rather non-trivial in our model. The key reason is that we only have that the expected (not sample-path) cycle cost function is strongly convex, and when we work with expected regret, the number of random cycles depends on CUP, which then correlates with its random cycle cost, and the standard argument based on Wald's Theorem does not work. To circumvent this technical issue, we "stretch" the time horizon from T period to the T-th cycle of CUP, and show that the difference between the cumulative regret over T periods and that over T-th cycle is bounded.

3) **Other contribution.** To establish our main result in this chapter, we first prove a key structural result for perishable inventory systems operating under base-stock policies, which is of independent interest. More specifically, we show in Theorem 2.1 that that the

T-period total holding, lost-sales and outdating cost is convex in the base-stock level along every sample path and for any $T \geq 1$. Our approach is linear programming (LP) based, which has been used in Janakiraman and Roundy (2004) for establishing a convexity result for non-perishable inventory systems with lost-sales and positive lead times.

2.1.3 Relevant Literature

Our work is mostly closely related to the following two research streams.

Perishable inventory systems. With complete distributional information of demand, there has been a vast body of literature devoted to the study of perishable inventory systems, since as early as Veinott (1960), Bulinskaya (1964), and Van Zyl (1964). Subsequently, the two landmark papers Nahmias (1975b) and Fries (1975) then, independently, studied the optimal policy for the general lifetime problem with independent and identically distributed (i.i.d.) demands, in a backlogging model and a lost-sales model, respectively. They showed that the optimal ordering policy depends on both the age distribution of the current inventory and the remaining time until the end of planning horizon. Since then this subfield of inventory theory has taken off and grown rapidly, which attracted much attention from both academics and practitioners. We refer interested readers to the survey articles Karaesmen et al. (2011) and Nahmias (2011) for an overview. Most recently, Chen et al. (2014b) and Li and Yu (2014) derived new structural properties of optimal policies using the concepts of $L^\sharp$-convexity and multimodularity, respectively. In parallel, Chao et al. (2015b,a) and Zhang et al. (2015) developed a series of approximation algorithms to compute provably near-optimal solutions for such complex systems.

Since we primarily focus on the class of base-stock policies (or fixed critical number policies), to put this chapter into the proper context, we mainly survey relevant papers that studied the performance of base-stock policies in theory and computation. As base-stock policy is easy to compute and implement, researchers quickly turned their interest to base-stock policies. Nahmias (1975a) was perhaps the first paper to analyze the performance of base-stock policies and showed that they are extremely effective in all instances tested compared with two other simple policies. Cohen (1976) computed the stationary distribution of the total stock for a two-period lifetime problem and thereby derived the optimal critical number policy. Nahmias (1976), Chazan and Gal (1977) derived bounds on the expected outdating cost and used it in the calculation of the critical number policy. Nahmias (1977c, 1978) extended this result to incorporate random lifetimes and fixed order (setup) costs, respectively. The aforementioned papers all reported excellent computational results on the performance of base-stock policies, with resulting cost less than 2% higher than the optimal

cost in almost all instances tested. Cooper (2001) derived several bounds on the limiting distribution of the number of outdates in a period, as well as an upper bound of the variance of the stationary number of outdates, and used these bounds to search for the best possible base-stock policy. His numerical results showed that base-stock policies perform within 1% of optimality under various demand distributions. Nandakumar and Morton (1993) and Deniz et al. (2010) also conducted computational studies on the effectiveness of base-stock policies.

The results in this chapter differ from the above literature by not assuming any distributional information on demand *a priori*, and focuses a joint learning and optimization problem for finding the best base-stock policy under censored demand information.

Nonparametric algorithms for inventory models. A number of papers have been published on nonparametric algorithms for non-perishable inventory systems. Burnetas and Smith (2000) developed a gradient descent type algorithm for repetitive newsvendor problem (i.e., without inventory carryover), and they showed that the average profit converges to the optimal one but did not establish the rate of convergence. Huh and Rusmevichientong (2009) proposed a gradient descent based algorithm for lost-sales systems with censored demand. Subsequently, Huh et al. (2009a) proposed an algorithm for finding the optimal base-stock policy in lost-sales inventory systems with positive lead time. Besbes and Muharremoglu (2013) examined the discrete demand case and showed that active exploration is needed. Huh et al. (2011) applied the concept of Kaplan-Meier estimator to devise another data-driven algorithm for censored demand. Recently, Shi et al. (2015) proposed algorithm for multi-product inventory systems under a warehouse-capacity constraint with censored demand. Chen et al. (2015a,b) proposed a nonparametric data-driven algorithm for the joint pricing and inventory control problem with backorders and lost-sales, respectively.

Another nonparametric approach in the inventory literature is sample average approximation (SAA) (e.g., Kleywegt et al. (2002), Levi et al. (2007b, 2015)) which uses the empirical distribution formed by *uncensored* samples drawn from the true distribution. Concave adaptive value estimation (e.g., Godfrey and Powell (2001), Powell et al. (2004)) successively approximates the objective cost function with a sequence of piecewise linear functions. The bootstrap method (e.g., Bookbinder and Lordahl (1989)) estimates the newsvendor quantile of the demand distribution. The infinitesimal perturbation approach (IPA) is a sampling-based stochastic gradient estimation technique that has been used to solve stochastic supply chain models (see, e.g., Glasserman (1991)). Maglaras and Eren (2015) employed maximum entropy distributions to solve a stochastic capacity control problem. For parametric approaches in stochastic inventory systems, see, e.g., Lariviere and Porteus (1999) and Chen

and Plambeck (2008) on Bayesian learning, and Liyanage and Shanthikumar (2005) and Chu et al. (2008) on operational statistics.

The results in this chapter contribute to the literature by developing the first nonparametric learning algorithm for finding the optimal base-stock policy in periodic-review perishable inventory systems with censored demand. The work closest to ours is perhaps Huh et al. (2009a) on non-perishable inventory system with lost-sales and positive lead time, and the authors proposed a learning algorithm for finding the optimal base-stock policy. However, the two inventory systems are significantly different, and the algorithms developed in these papers are based on different approaches. Hence, both our algorithms and results are significantly different from theirs.

2.1.4 Structure and General Notation

The rest of this chapter is organized as follows. In §2.2, we formally describe the periodic-review perishable inventory systems with lost-sales and censored demand, and also the class of base-stock policies. In §2.3, we establish an important structural result by showing that the T-period total cost is convex in the base-stock level along every sample path for any $T \geq 1$. In §2.4, we introduce the cycle-update policy (CUP) in §2.4.1-§2.4.2 and conduct a numerical study §2.4.3. In §2.5, we carry out a performance analysis of CUP and establish a regret bound. In §2.6, we consider the strongly convex extension of our model and obtain an improved rate of convergence under some mild technical conditions. In §2.7, we conclude this chapter by summarize the results in this chapter.

Throughout this chapter, we often distinguish between a random variable and its realizations using capital and lower-case letters, respectively. For any real numbers x and y, we denote $x^+ = \max\{x, 0\}$, $x \vee y = \max\{x, y\}$, and $x \wedge y = \min\{x, y\}$. The indicator function $\mathbb{1}(A)$ takes value 1 if A is true and 0 otherwise, and " $\triangleq$ " stands for "defined as". We use LHS and RHS as abbreviations for "left-hand side" and "right-hand side", respectively. The projection function is defined as $\mathbf{P}_{[a,b]}(x) = \min[b, \max(x, a)]$ for any real numbers x, a, b.

2.2 Perishable Inventory Systems with Censored Demand

We formally describe the stochastic periodic-review perishable inventory system with censored demand. The product lifetime m is known and fixed, i.e., items perish after staying in inventory for m periods if not consumed. Let $t \in \{1, 2, \ldots\}$ represent the time period, which is indexed forward. For each period t, we denote the demand in period t by a continuous random variable D_t. We assume that D_t, $t = 1, \ldots, T$, are i.i.d. across time period t.

Contrary to the classical formulation, the firm has no prior knowledge about the true underlying demand distribution *a priori*, but can observe past sales data (i.e., censored demand data), and make adaptive inventory decisions based on the available information.

2.2.1 System Dynamics and Objectives

For perishable inventory systems, any inventory unit that stays in the system for m periods without meeting the demand expires and exits the system. Thus, we use a state vector $\mathbf{x}_t$ to keep track of the inventory age information at the beginning of any period t (before ordering), i.e.,

$$\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,m-1}),$$

where for $i = 1, \ldots, m-1$, $x_{t,i}$ is the on-hand inventory level of product whose remaining lifetime is *no more* than i periods. It is clear that $x_{t,m-1}$ is the total on-hand inventory level in period t. For notational convenience, we use $x_{t,m} = x_{t,m-1} + q_t$ to denote the total on-hand inventory level (*after* ordering q_t amount of new inventory units) in period t, which is in fact our control variable.

For any FIFO issuance policy π, the sequence of events in each period t, $t = 1, 2, \ldots$, is as follows. (Note that $q_t^\pi, \mathbf{x}_t^\pi, o_t^\pi$ all depend on π; for brevity, we shall make the dependency implicit.)

(a) At the beginning of each period t, the firm observes the starting inventory vector $\mathbf{x}_t$.

(b) The firm makes a replenishment decision $q_t \geq 0$ in period t, and the replenishment order arrives instantaneously. (Note that the zero lead time assumption is predominant in perishable inventory literature, see e.g., Nahmias (2011) and Karaesmen et al. (2011)). The total on-hand inventory level (after receiving the order q_t) is $x_{t,m} = x_{t,m-1} + q_t$.

(c) Then the random demand D_t is realized (denote its realization by d_t) and satisfied to the maximum extent by FIFO issuing policy, i.e., the oldest inventory meets demand first. Under censored demand information, the firm does not observe the realized demand d_t but observes the sales quantity $\min(d_t, x_{t,m})$ only.

(d) At the end of the period, all the outstanding inventories incur a unit holding cost h and all the unsatisfied demands are *lost* with a unit lost-sales penalty cost p. Note that the lost-sales cost is unobservable in the event of stockout, due to demand censoring. Finally, all the inventories that have stayed in the system for m periods expire with unit outdating cost θ. Following the convention by Nahmias (1975b), we assume the

inventory units that perish at the end of this period also incur a holding cost. As a result, the period t cost, C_t^π, is

$$C_t^\pi(\omega) = h(x_{t,m} - d_t)^+ + p(d_t - x_{t,m})^+ + \theta o_t, \tag{2.1}$$

where $o_t = (x_{t,1} - d_t)^+$ is the outdating inventory, and the number of lost-sales in period t is $(d_t - x_{t,m})^+$, which is unobservable due to demand censoring. We will assume, without loss of generality, that the unit purchasing cost is zero (see a detailed cost transformation in Chao et al. (2015b)).

(e) At last, the system proceeds to period $t+1$ with $\mathbf{x}_{t+1}$ given by

$$x_{t+1,j} = (x_{t,j+1} - d_t - o_t)^+ = \left(x_{t,j+1} - d_t - (x_{t,1} - d_t)^+\right)^+, \quad \text{for } 1 \le j \le m-1. \tag{2.2}$$

The objective is to find a replenishment policy that only utilizes past sales data to minimize the long-run average cost.

2.2.2 The Class of Base-Stock Policies

Even with complete information about the demand distribution *a priori*, it is well-known that the (clairvoyant) optimal policy for perishable inventory systems is extremely complicated (see Karaesmen et al. (2011), Nahmias (2011)), and computing the exact optimal policy is intractable using brute-force dynamic programming. However, it has been shown in the literature that the class of *base-stock policies* has near-optimal computational performance (see Cooper (2001) and the detailed discussion in §4.1). Hence, in this chapter, we focus our attention to find the best base-stock policy. Recall that under a base-stock policy of level S, the total inventory level at the beginning of each period is always raised to S, i.e., for any period t we have $q_t = (S - x_{t,m-1})^+$. We assume that the system is initially empty, i.e., $\mathbf{x}_1 = \mathbf{0}$.

Without any prior knowledge about the demand distribution, an admissible or feasible base-stock policy $\pi(S_t)$ is represented by a sequence of *period-dependent* order-up-to levels, $\{S_t, t \ge 1\}$ with $S_t \ge x_{t,m-1}$, where S_t depends only on the sales and decisions made prior to time t, i.e, S_t is adapted to the filtration generated by $\{S_s, \min\{S_s, D_s\} : s = 1, \ldots, t-1\}$ under censored demand. Restricting to this class of policies, we wish to develop a nonparametric adaptive inventory control policy $\pi(S_t)$ so that its average cost per period converges

to that of the (clairvoyant) optimal base-stock policy, i.e.,

$$\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}C_t^{\pi(S_t)}\right]=\inf_{S}\left\{\limsup_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}C_t^{\pi(S)}\right]\right\}, \tag{2.3}$$

where our adaptive policy $\pi(S_t)$ on the LHS of (2.3) is constructed under unknown demand distribution *a priori* and censored demand information while the optimal base-stock policy $\pi(S^*)$ on the RHS of (2.3) is constructed under known demand distribution *a priori*. We will also find the rate at which the average cost of policy $\pi(S_t)$ converges to that of the (clairvoyant) policy.

2.3 Convexity for Base-Stock Policies

We first analyze the perishable inventory system operating under a base-stock policy. A natural and important question is whether the total expected cost from period 1 to T is convex in the base-stock level S for any $T \geq 1$. The answer is affirmative and we shall provide an LP-based proof in this section. This LP-based technique has been used in Janakiraman and Roundy (2004) to prove the convexity of total cost in the base-stock level in a non-perishable inventory system with lost-sales and positive lead times.

Theorem 2.1. *For the perishable inventory systems operating under a base-stock policy $\pi(S)$, for any realization of demand $\omega = (d_1, d_2, \ldots)$, the T-period total costs incurred by $\pi(S)$ is convex in S for any $T \geq 1$.*

Proof. We shall prove a stronger result that, the total holding cost, the total lost-sales cost, and the total outdating cost are all convex in S. The total holding and lost-sales costs from period 1 to T, when running a base-stock policy $\pi(S)$, are $h\sum_{t=1}^{T}(S-d_t)^+$ and $p\sum_{t=1}^{T}(d_t-S)^+$, respectively, and they are clearly convex in S. In the following, we prove that the total outdating cost is also convex in S.

First, we observe that under the base-stock policy $\pi(S)$ and zero replenishment lead time, the amount of outdating inventory units in any period for the lost-sales model is identical to that for the backlogging model. This holds for any realization of demand process. We can then turn to the backlogging counterpart model for the remainder of this proof.

The backlogging perishable inventory system also starts with zero inventory, i.e., $x_{1,i} = 0$ for all $i = 1, \ldots, m-1$. Under the base-stock policy $\pi(S)$, the on-hand inventory level (after ordering) is $x_{t,m} = S$ for all $t = 1, \ldots, T$. For any given $d_1, \ldots, d_{T-1}$, we construct a linear

program $\mathbf{LP}(S)$ as follows:

$$\min_{\{\mathbf{q},\mathbf{x}\}} \quad \sum_{t=1}^{T} q_t \tag{2.4}$$

$$\text{subject to} \quad x_{1,i} = 0, \quad i = 1, \ldots, m-1, \tag{2.5}$$

$$x_{t,m} = S, \quad t = 1, \ldots, T, \tag{2.6}$$

$$x_{t+1,i-1} = x_{t,i} - q_{t+1}, \quad t = 1, \ldots, T-1, \;\; i = 2, \ldots, m, \tag{2.7}$$

$$q_{t+1} \geq d_t, \quad t = 1, \ldots, T-1, \tag{2.8}$$

$$q_{t+1} \geq x_{t,1}, \quad t = 1, \ldots, T-1. \tag{2.9}$$

The decision variables are $\mathbf{q} = (q_t)_{t=1,\ldots,T}$, and $\mathbf{x} = (x_{t,i})_{t=1,\ldots,T,i=1,\ldots,m}$. For notational convenience, we denote the feasible region (2.5–2.9) by $\mathbf{\Gamma}(S)$.

We claim that, the unique feasible solution to $\mathbf{\Gamma}(S)$ and the system of equations

$$q_t = \max(d_{t-1}, x_{t-1,1}), \quad t = 2, \ldots, T \tag{2.10}$$

is an optimal solution to $\mathbf{LP}(S)$. Note that $\mathbf{\Gamma}(S)$ and (2.10) completely describe the evolution of a backlogging perishable inventory system operating under an order-up-to-S policy. Since for any feasible solution $(\mathbf{q}, \mathbf{x})$ to $\mathbf{LP}(S)$, $\mathbf{x}$ is completely and uniquely determined by $\mathbf{q}$ using (2.6) and (2.7), we will focus on $\mathbf{q}$ in the remainder of the proof (while leaving $\mathbf{x}$ implicit).

To prove the claim above, we first argue that for any given $d_1, \ldots, d_{T-1}$, there is a unique solution $\hat{\mathbf{q}}$ that satisfies $\mathbf{\Gamma}(S)$ and (2.10). Combining (2.6) and (2.7) from $\mathbf{\Gamma}(S)$, we have

$$\hat{x}_{t,1} = S - \hat{q}_{t-m+2} - \hat{q}_{t-m+3} - \cdots - \hat{q}_t, \quad t = 2, \ldots, T. \tag{2.11}$$

It is clear that $\hat{q}_1 = S$ by (2.5) and (2.6). By (2.10), $\hat{q}_2 = \max(d_1, \hat{x}_{1,1}) = \max(d_1, 0) = d_1$, which is unique. For $t = 3, \ldots, T$, $\hat{q}_t = \max(d_{t-1}, \hat{x}_{t-1,1})$ which is determined using only $\hat{q}_2, \ldots, \hat{q}_{t-1}$ due to (2.11). Hence, the solution $\hat{\mathbf{q}} = (\hat{q}_1, \ldots, \hat{q}_T)$ can be sequentially and uniquely determined.

We proceed to prove that $\hat{\mathbf{q}}$ is also optimal by constructing this solution from an arbitrary optimal solution $\mathbf{q}^0$ in $T-2$ steps. (It is clear that $\mathbf{LP}(S)$ is bounded below by zero so optimal solutions must exist.) For notational convenience, we denote the solution after step k by $\mathbf{q}^k$ (while keeping its corresponding $\mathbf{x}^k$ implicit). In each step $k = 1, \ldots, T-2$, we keep $\mathbf{q}^k$ feasible without changing its objective value. We shall argue that $\mathbf{q}^{T-2} = \hat{\mathbf{q}}$, which has the desired property (2.10).

In the first step, we set $\mathbf{q}^1 = \mathbf{q}^0$ and carry out the following operation. If $q_2^0 = \max(d_1, x_{1,1}^0)$, then we simply do nothing. Otherwise, if $q_2^0 > \max(d_1, x_{1,1}^0)$, then we decrease q_2^1 such that $q_2^1 = \max(d_1, x_{1,1}^0)$ and increase q_3^1 such that $q_3^1 = q_3^0 + q_2^0 - \max(d_1, x_{1,1}^0)$. We keep all other entries of $\mathbf{q}^1$ unchanged, and also determine the corresponding $\mathbf{x}^1$. It is clear that the objective value remains unchanged. Moreover, $\mathbf{q}^1$ is also feasible, since after this operation, $x_{1,1}^1 = 0$ is unchanged, $x_{2,1}^1$ is raised as much as q_3^1 is raised (hence keeping q_3^1 feasible), and $x_{t,1}^1$ is non-increasing for all $t = 3, \ldots, T-1$.

Then, in the subsequent steps $k = 2, \ldots, T-2$, we set $\mathbf{q}^k = \mathbf{q}^{k-1}$ and apply the same operation to change the entries q_{k+1}^k and q_{k+2}^k only to obtain a new $\mathbf{q}^k$. By the identical argument, we can show that $\mathbf{q}^k$ is feasible and gives the same objective value as the previous solution $\mathbf{q}^{k-1}$.

After these $T-2$ steps, we have obtained a feasible $\mathbf{q}^{T-2}$ that satisfies (2.10) for $t = 2, \ldots, T-1$. It remains to verify that $q_T^{T-2} = \max(d_{T-1}, x_{T-1,1}^{T-2})$. This holds because if otherwise $q_T^{T-2} > \max(d_{T-1}, x_{T-1,1}^{T-2})$, then we could decrease q_T^{T-2} until the inequality becomes binding, which gives rise to a new feasible solution that has a strictly lower objective value than $\mathbf{q}^0$, contradicting to the assumption that $\mathbf{q}^0$ is an optimal solution. Hence, we have that $\mathbf{q}^{T-2}$ satisfies (2.10) and is also optimal by the above construction argument. Furthermore, $\mathbf{q}^{T-2} = \hat{\mathbf{q}}$, since $\hat{\mathbf{q}}$ is the unique feasible solution that satisfies (2.10). We have proven the claim.

Since the feasible region $\mathbf{\Gamma}(S)$ is a convex subset of the space of decision variable $\{\mathbf{q}, \mathbf{x}\}$ and the parameter S, it follows that the optimal objective value of the linear program $\mathbf{LP}(S)$ is convex in S. Because there is a unique optimal solution $\hat{\mathbf{q}}$ that satisfies $\mathbf{\Gamma}(S)$ and (2.10) (completely describing the evolution of a backlogging perishable inventory system operating under an order-up-to-S policy), this shows that the total number of inventory units ordered $\sum_{t=1}^{T} \hat{q}_t$ under this order-up-to-S policy is convex in S. In addition, it is easy to see that when running an order-up-to-S policy, we have

$$\sum_{t=1}^{T} \hat{q}_t = S + \sum_{t=1}^{T-1} d_t + \sum_{t=1}^{T-1} o_t.$$

This shows that, for any sequence of demand realizations $d_1, \ldots, d_{T-1}$, the total outdating cost $\theta \sum_{t=1}^{T-1} o_t$ is convex in S. This completes the proof of Theorem 2.1. □

As a by-product of the proof of Theorem 2.1, we obtain an interesting relationship between the optimal base-stock level S^* for the perishable inventory system and the optimal base-stock level $\tilde{S}^*$ for the counterpart of non-perishable inventory system with infinite life-

times. Since the optimal base-stock level for non-perishable periodic-review inventory system has a closed form solution (i.e., the newsvendor quantile solution), this result gives an upper bound for the optimal base-stock level of perishable inventory system. Of course, this result is useful only when the demand distribution is known *a priori*.

Corollary 2.2. *Consider a perishable inventory system and its counterpart of non-perishable inventory system with infinite lifetimes, under the same initial conditions, cost parameters and demand distributions. Denote the optimal base-stock levels for the perishable inventory system and the nonperishable inventory system by S^* and $\tilde{S}^*$, respectively. Then we have $S^* \leq \tilde{S}^*$.*

Proof. Denote the expected one-period holding and lost-sales cost by $L(S) = \mathbb{E}[h(S-d)^+ - p(d-S)^+]$. It is well-known that $\tilde{S}^*$ is the unique minimizer of $L(S)$, and $L'(\tilde{S}^*) = 0$. For the corresponding perishable inventory system, besides $L(S)$, the inventory system also incurs an outdating cost. Denote the expected long-run outdating cost by using a base-stock policy $\pi(S)$ by $A(S)$. By Theorem 1, $A(S)$ is convex in S. It is clear that $A(S)$ is increasing in S, and therefore $A'(S) \geq 0$ for any $S \geq 0$. Since $L'(S^*) + A'(S^*) = 0$, we must have $S^* \leq \tilde{S}^*$. This completes the proof. □

2.4 Nonparametric Algorithm: Cycle-Update Policy (CUP)

Not knowing the true underlying demand distribution D_t *a priori*, our objective is to find a provably good adaptive data-driven algorithm for inventory control such that its total expected system cost is close to that of the clairvoyant optimal base-stock policy. In the following, we present a novel *Cylce-Update Policy* (CUP for short) for the perishable inventory system with lost-sales and censored demand, which achieves the aforementioned objective.

We first make an assumption about the (clairvoyant) optimal base-stock level S^*.

Assumption 2.3. *There is a known finite number $\bar{S}$ such that $S^* \leq \bar{S}$, and $\mathbb{P}(D_t > \bar{S}) > 0$.*

This is a mild and reasonable assumption since typically the firm has some idea about the maximum possible base-stock level. Similar boundedness assumptions on the optimal base-stock levels have also appeared in Huh and Rusmevichientong (2009), Huh et al. (2009a, 2011), Shi et al. (2015), Chen et al. (2015b) for other inventory systems. We also remark that if the firm has some prior estimate of the demand distribution, the firm can readily compute $\tilde{S}^*$ for the counterpart (non-perishable) inventory system, which then serves an upper bound for S^* by Corollary 2.2.

We introduce the following notation to denote the total cost in periods $\{n_1, n_1+1, \ldots, n_2-1\}$ for any $1 \leq n_1 < n_2 \leq T+1$ operating under a base-stock policy S with *brand new* (after ordering) inventory level S in period n_1:

$$G(S, (n_1, n_2); \omega) = \sum_{t=n_1}^{n_2-1} C_t^{\pi(S)}(\omega), \tag{2.12}$$

where $C_t^{\pi(S)}(\omega)$ is given in (2.1). Then, by Theorem 1, $G(S, (n_1, n_2); \omega)$ is convex in S for any n_1, n_2 on every sample path ω.

2.4.1 Cycle-Update Policy (CUP)

The key idea of our cycle-update policy (CUP) algorithm is to update the base-stock level every time the system experiences a stockout, and keeps the current base-stock level unchanged otherwise. Define the stockout period as the end of a cycle. Algorithm 1 below describes the algorithm, which calls to a detailed routine for computing a cycle gradient described in §2.4.2, that is used to compute the base-stock level for a new cycle.

In contrast to the existing literature on online convex optimization, the cycles in our algorithm are not *a priori* fixed or known, and they are triggered sequentially by lost-sales events as demands realize over time. Specifically, let τ_k be the beginning of k-th cycle for which CUP implements a newly computed base-stock level S_k, $k = 1, \ldots$, the cycle ends the first time after τ_k that stockout occurs. That is, for each sample path $\omega = \{d_1, d_2, \ldots\}$, $\tau_1(\omega) = 1$, and for $k \geq 1$,

$$\tau_{k+1}(\omega) = \inf \{t \geq \tau_k(\omega) + 1 : x_{t,m-1}(\omega) = 0\}.$$

The k-th cycle cost of the CUP is $G(S_k, (\tau_k, \tau_{k+1}); \omega)$. Note that $\tau_{k+1} - \tau_k$ is geometrically distributed with parameter $P(D > S_k)$, where D is a generic single-period demand. In addition, τ_{k+1} is *not* independent of the costs $C_{\tau_k}^{\pi(S_k)}, \ldots, C_{\tau_{k+1}-1}^{\pi(S_k)}$ incurred in cycle k.

In what follows, *we let* $\nabla_1 G(S_k, (n_1, n_2); \omega))$ *denote the partial derivative of* $G(S_k, (n_1, n_2); \omega)$ *with respect to* S_k. For notational convenience, we often make the dependency on ω implicit.

Remark 2.4. One important observation of our CUP algorithm is that the starting inventory level in each period is always below the base-stock level. This is because CUP updates the base-stock level only when the system becomes empty, and keep the same base-stock level otherwise. This implies that CUP can always attain the desired base-stock level exactly in each period. It shall be noted that many papers in the demand learning literature need to deal

Algorithm 1 Cycle-Update Policy (CUP)

(Initialization.) Set $\tau_1 = 1$, and the initial base-stock level S_0 for period 1 is arbitrarily chosen from $(0, \bar{S})$. Set $x_{1,m-1} = S_0$, and set the cycle counter to $k = 1$. For each period $t \geq 2$, repeat the following procedure:

Case 1: If the starting inventory level $x_{t,m-1} > 0$ (meaning that lost-sales did not occur in period $t-1$), then keep the same base-stock level as in period $t-1$, i.e., order up to S_k in period t so that $x_{t,m} = S_k$. Go to the next period.

Case 2: If the starting inventory level $x_{t,m-1} = 0$ (meaning that lost-sales occurred in period $t-1$), then set $\tau_{k+1} = t$ as the beginning of a new cycle $k+1$, and update the base-stock level S_{k+1} by

$$S_{k+1} := \mathbf{P}_{[0,\bar{S}]}\big(S_k - \eta_k \nabla_1 G(S_k, (\tau_k, \tau_{k+1}))\big), \tag{2.13}$$

where the step-size $\eta_k = \gamma/\sqrt{k}$ for some positive constant γ, and $\nabla_1 G(S_k, (\tau_k, \tau_{k+1}))$ is the gradient of the k-th cycle cost with respect to S_k *fixing* τ_k and τ_{k+1}, which can be efficiently computed using a *subroutine* presented in §2.4.2 via (2.15). Order up to S_{k+1} for period t so that $x_{t,m} = S_{k+1}$, and set $k := k+1$. Go to the next period.

with the "overshooting" or "undershooting" issue of not being able to achieve the desired base-stock levels in some periods due to either positive inventory carry-over or capacity constraints (e.g., Huh and Rusmevichientong (2009), Huh et al. (2009a), Shi et al. (2015), Chen et al. (2015b)). We resolve this issue by the clever algorithmic design of CUP, which greatly simplifies the performance analysis.

We also remark that one cannot design cycles based on successive stockouts of any feasible policy, e.g., between successive events $\{D > \bar{S}\}$. Although this particular cycle design makes coupling of two policies much easier as it is independent of policies, the event $\{D > \bar{S}\}$ is unobservable due to censored demand. In contrast, designing cycles based on successive stockouts of CUP remains feasible under censored demand. We point out that even when the event $\{D > \bar{S}\}$ can be observed (for example, in the backlog case of our model), we prefer to not use it as the chance for $\{D > \bar{S}\}$ may be rather small. □

2.4.2 Computing the Cycle Gradient

The above CUP algorithm requires computing the (sample-path) gradient of the total cost within a cycle with respect to S_k for every sample path ω. The cycle gradient is the sum of the following two parts. It is important to note that, τ_{k+1} depends on S_k, however, we only compute the partial derivative of cycle cost for fixed τ_k and τ_{k+1}.

The cycle gradient of the holding and lost-sales cost. The computation of the gradient of the cycle holding and lost-sales cost is straightforward. For each cycle $k = 1, 2, \ldots$, the gradient (with respect to the base-stock level S_k) is simply

$$h \cdot (\tau_{k+1} - \tau_k - 1) - p. \tag{2.14}$$

This is because, by the definition of the k-th cycle, namely periods τ_k to $\tau_{k+1} - 1$, the CUP algorithm ends with strictly positive inventory during the first $(\tau_{k+1} - \tau_k - 1)$ periods, and experiences a stockout in the last period $\tau_{k+1} - 1$. Hence the result follows from the assumption that the demand is a continuous random variable.

The cycle gradient of the outdating cost. The computation of the gradient of the cycle outdating cost is more involved (but can be efficiently computed). Focus on the k-th cycle, namely $\{\tau_k, \ldots, \tau_{k+1} - 1\}$, where the base-stock level is kept at S_k, and objective is to compute the gradient of the cycle outdating cost with respect to S_k. Let $u_{t,i}$ denote the gradient of inventory with remaining lifetime i with respect to S_k for fixed τ_k and τ_{k+1}. In each period t, besides the inventory vector $\mathbf{x_t}$, where the newly received order q_t in period t is considered as inventory with remaining lifetime m, we will keep track of another m-dimensional vector

$$\mathbf{u}_t = (u_{t,1}, \ldots, u_{t,m}),$$

which represents the derivatives of inventory levels of different remaining lifetimes with respect to S_k. Since the sum of all inventory units is S_k in each period t during cycle k, the sum of all the entries of $\mathbf{u_t}$ must be 1. In fact, it can be argued that $u_{t,i} \in \{0, 1\}$ for all $i = 1, \ldots, m$. For notational convenience, we use $\mathbf{e}_i^m$ to denote an m-dimensional vector whose i-th entry is 1 and all other entries are 0. Then $\mathbf{u_t} \in \{\mathbf{e}_1^m, \ldots, \mathbf{e}_m^m\}$ for each period t.

The following subroutine specifies how $\mathbf{u_t}$ is updated during the k-th cycle, and it is used to determine the cycle gradient of outdating cost.

The main idea underlying this subroutine is the following: We perturb the base-stock level S_k by an infinitesimal amount to $S_k + \delta$, and compute the additional amount of outdating inventory due to such a change within the cycle (see Figure 3.3 as an example). Call the two systems, with base-stock levels S_k and $S_k + \delta$, the original system and the perturbed system, respectively. By base-stock policy, the total inventory of different remaining lifetimes in the two systems are always S_k and $S_k + \delta$, respectively, in each period of cycle k. By the assumption of continuous demand, with probability 1, the inventory levels of different remaining lifetimes in the two systems are identical except at one entry, at which the perturbed system has δ more units of inventory than the original system. By keeping track of

Subroutine Computing the k-th Cycle Gradient for Algorithm 1

Initialization: In period $t = \tau_k$, initialize $\mathbf{u_t} := \mathbf{e}_m^m$, and a counter $n = 0$.

Main Step: For each subsequent period $t = \tau_k + 1, \ldots, \tau_{k+1}$, suppose $\mathbf{u_{t-1}} := \mathbf{e}_i^m$ for some $i \in \{1, \ldots, m\}$.

Case 1: If no outdating occurs in period $t-1$, then let $j = \min\{\ell : x_{t,\ell} > 0\}$ denote the remaining lifetime of the oldest inventory in period t (after ordering), and set $\mathbf{u_t} := \mathbf{e}^m_{\max(i-1,j)}$.

Case 2: If outdating occurs in period $t-1$, then

(i) if $i = 1$, then set $\mathbf{u_t} := \mathbf{e}_m^m$, and set $n := n + 1$;

(ii) otherwise set $\mathbf{u_t} := \mathbf{e}_{i-1}^m$.

---

the extra inventory level δ, which is precisely the unit vector $\mathbf{u_t}$, the subroutine allows us to compute how much more inventory outdate in the perturbed system than in the original system, and it is exactly $n\delta$ if n is the output of the subroutine. The following result presents the gradient of the cycle outdating cost using the described subroutine.

Proposition 2.5. *Let n be the output of the subroutine for cycle k, then the gradient for the k-th cycle outdating cost is θn.*

Proof. The main idea underlying this subroutine is that when we perturb the current base-stock level S_k by an infinitesimal amount δ, we compute how many δ's outdate within the cycle (see Figure 3.3 as an example). To that end, we keep track of the additional δ units of inventory by using the auxiliary vector $\mathbf{u}_t$ in each period t. More precisely, if $\mathbf{u}_t = e_i^m$ for some $1 \le i \le m$, then there is an additional δ (as a result of raising the base-stock level to $S_k + \delta$) that have remaining lifetime i in the on-hand inventory in period t. Indeed, since the inventory levels of different remaining lifetimes add to $S_k + \delta$, the extra δ units have to appear somewhere (and the assumption of continuous demand ensures that they do not split to different age groups).

We prove, by induction, that for any $t = \tau_k, \ldots, \tau_{k+1} - 1$, if $\mathbf{u}_t = \mathbf{e}_i^m$ then after we raise the base-stock level S_k by an infinitesimal amount δ, the inventory with remaining lifetime i will increase by δ in period t, while inventory levels with any other remaining lifetime remain unchanged.

First, following our subroutine, $u_{\tau_k} = e_m^m$. Recall that the system is empty at the beginning of period τ_k, hence the ordering quantity $q_{\tau_k} = S_k + \delta$ and inventory levels with remaining lifetime not equal to m are all 0. So the claim is true for $t = \tau_k$. Suppose that

the claim has been shown for period $t-1$, and we want to prove that it also holds for $\mathbf{u}_t$.

For clarity we consider two systems, one with base-stock level S_k, called the original system, and the other with base-stock level $S_k+\delta$, called by the perturbed system. By induction assumption suppose $\mathbf{u}_{t-1}=\mathbf{e}_i^m$ for some i, i.e., in period $t-1$ the perturbed system has the same inventory vector as the original system except for an additional δ with remaining lifetime i. We consider two cases separately below.

Case 1: If no outdating occurs in period $t-1$, then we have the following subcases.

a) If $i=1$ or $\mathbf{u}_{t-1}=\mathbf{e}_1^m$, then all units with remaining lifetime 1 are consumed by demand in both the original and the perturbed systems, and by FIFO, there will be δ extra units in the perturbed system with the oldest inventory in period t, i.e., $\mathbf{u}_t=\mathbf{e}_j^m$ where $j=\min\{\ell : x_{t,\ell}>0\}$.

b) If $i>1$, then the δ extra units in the perturbed system are either consumed or still in system in period t. In the former case, $\mathbf{u}_t=\mathbf{e}_j^m$ where $j=\min\{\ell : x_{t,\ell}>0\}$ is the oldest inventory in period t; and in the latter case, $\mathbf{u}_t=\mathbf{e}_{i-1}^m$ since the δ units have one less period of remaining lifetime in period t.

For both subcases a) and b), we update the vector $\mathbf{u}_t=\mathbf{e}^m_{\max(i-1,j)}$, but no change is made on outdating quantities. This is consistent with Case 1 in the Subroutine.

Case 2: If outdating occurs in period $t-1$, then we have the following subcases.

c) If $i=1$ or $\mathbf{u}_{t-1}=\mathbf{e}_1^m$, then all units with remaining lifetime being 1 either satisfy demand or outdate in both systems, and because there are δ more units in the perturbed system with remaining lifetime being 1, δ more units of inventory outdate in the perturbed system, incurring extra outdating cost. In this subcase the number of outdated inventory is increased by δ in the perturbed system. At the beginning of period t, an ordering quantity of S_k and $S_k+\delta$ will be ordered, respectively, in the original and perturbed systems all having remaining lifetime m, hence $\mathbf{u}_t=\mathbf{e}_m^m$.

d) If $i>1$, then the fact that there is oudatinng in period $t-1$ implies that the extra δ units with remaining lifetime i will still be in system and its remaining lifetime will be reduced to $i-1$ in period t. Thus, $\mathbf{u}_t=\mathbf{e}_{i-1}^m$.

This shows that in period t, $\mathbf{u}_t$ also represents the position of the extra δ inventory units in the perturbed system, which completes the induction proof. The argument above also

shows that, at the end of the Subroutine procedure, it has counted the number of extra δ's that outdate during the k-th cycle, and hence the output n represents the gradient of outdating cost with respect to S_k. This completes the proof of Proposition 2.5. □

Combining Proposition 2.5 and (2.14), we obtain the gradient of the k-th cycle total cost

$$\nabla_1 G(S_k, (\tau_k, \tau_{k+1})) = \theta n + h \cdot (\tau_{k+1} - \tau_k - 1) - p. \tag{2.15}$$

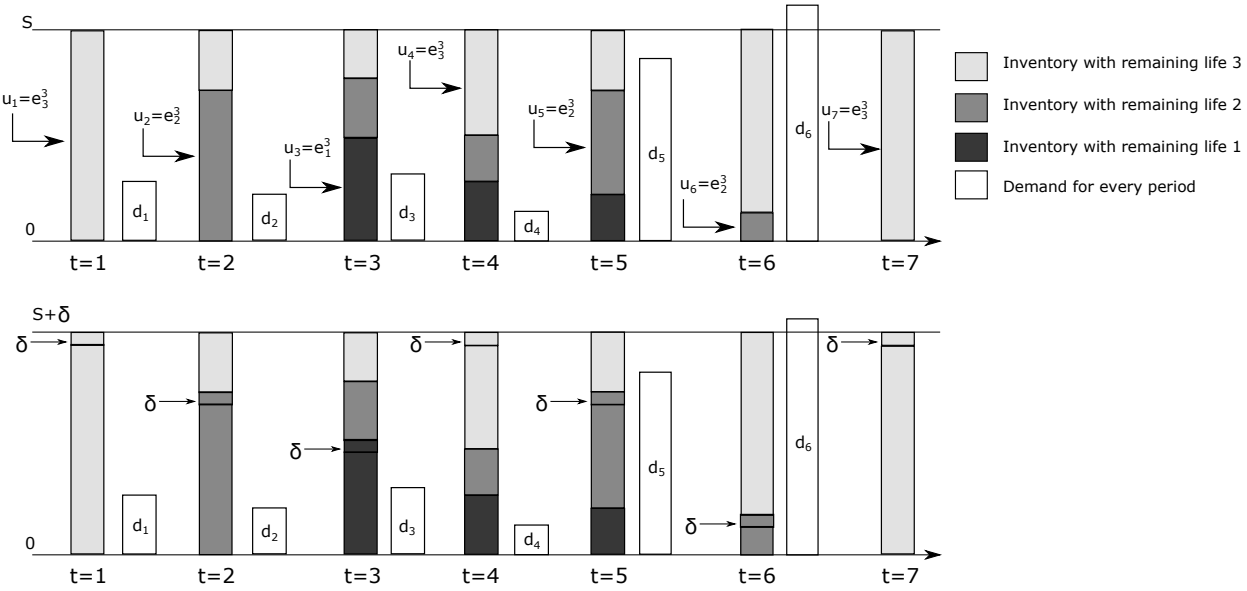

Figure 2.1: An example of how $\mathbf{u_t}$ is updated, with $m = 3$.

Example 2.6. To better understand the above subroutine, we use a concrete example with $m = 3$ to illustrate how $\mathbf{u_t}$ is updated during a cycle. This example is designed to cover all possible scenarios. In the upper portion of Figure 3.3, we keep track of $\mathbf{u}_t$ for every period t, while in the lower portion of Figure 3.3, we perturb the base-stock level by a small amount δ. We can see that after this perturbation, there is always an additional δ amount of inventory in each period t, and more importantly, the exact position (or remaining lifetime) of this δ amount of inventory is tracked using $\mathbf{u}_t$ for every period t.

In this example, there is no outdating in periods $1, 2, 5, 6$. As a result, when we update $\mathbf{u_t}$ for $t = 2, 3, 6, 7$, we identify the remaining lifetime of the oldest on-hand inventory unit in period t, and they are $2, 1, 2, 3$, respectively. We then update $\mathbf{u_t} = \mathbf{e}^m_{\max(i-1,j)}$ as shown in

Figure 3.3. Outdating happens in period 3 and 4, and we need to check the events $\{\mathbf{u_3} = \mathbf{e}_1^3\}$ and $\{\mathbf{u_4} = \mathbf{e}_1^3\}$. It turns out that the event $\{\mathbf{u_3} = \mathbf{e}_1^3\}$ is true but the event $\{\mathbf{u_4} = \mathbf{e}_1^3\}$ is false. Hence, the δ-outdating event happens only once during this cycle, and therefore the cycle gradient for the outdating cost is θ. □

Remark 2.7. One may initially think that the gradient of the total outdating cost within a cycle is simply θ times the number of outdating periods within a cycle. We remark that this naive way of computing the gradient does not give the right answer. As shown in Figure 3.3, when we raise the base-stock level by δ, the outdating amount in period 3 increases by δ, but the outdating amount in period 4 stays unchanged (which in fact equals to $d_1 - d_4$ in both cases). In this example, the naive way of computing the cycle gradient of total outdating cost would give us 2θ, but in fact it should be θ. □

2.4.3 A Numerical Study

An important question is how the proposed CUP algorithm performs numerically. We have conducted a numerical study, and the results are reported below. Our computations were done using Matlab R2014a on a desktop computer with an Intel(R) Xeon(R) CPU E31230 @ 3.20 Ghz.

We compare the performance of CUP against the (clairvoyant) optimal base-stock policies. To the best of our knowledge, there exist no benchmark learning algorithms or even heuristics reported in the literature for perishable inventory systems with demand distribution unknown *a priori*, and neither can we adapt any of the existing learning algorithms for other inventory systems to our perishable setting. The *performance* of CUP is measured by the percentage of total T-period cost increase compared with that of the (clairvoyant) optimal base-stock policies.

We test two types of demand distributions: uniform on $[0, 100]$ and truncated normal on $[0, 100]$ with mean 50 and standard deviation 25. The cost parameters are $h = 1, \theta = 5$ and $p \in \{5, 10\}$. We set the step-size $\gamma = 1$ in CUP, $\bar{S} = 95$, and the initial inventory level is 50. The (clairvoyant) optimal base-stock policy is computed through simulation. Each instance is run 5000 times to compute the average costs of CUP and the (clairvoyant) optimal base-stock policy.

Figure 2.2 reports the percentage of cost increase of CUP under different settings when T goes from 200 to 2000 periods. In every graph in Figure 2.2, the x-axis denotes the number of periods and y-axis is percentage of cost increase compared with that of the clairvoyant optimal base-stock policy. The exact numerical results in these graphs are given in Table

2.1.

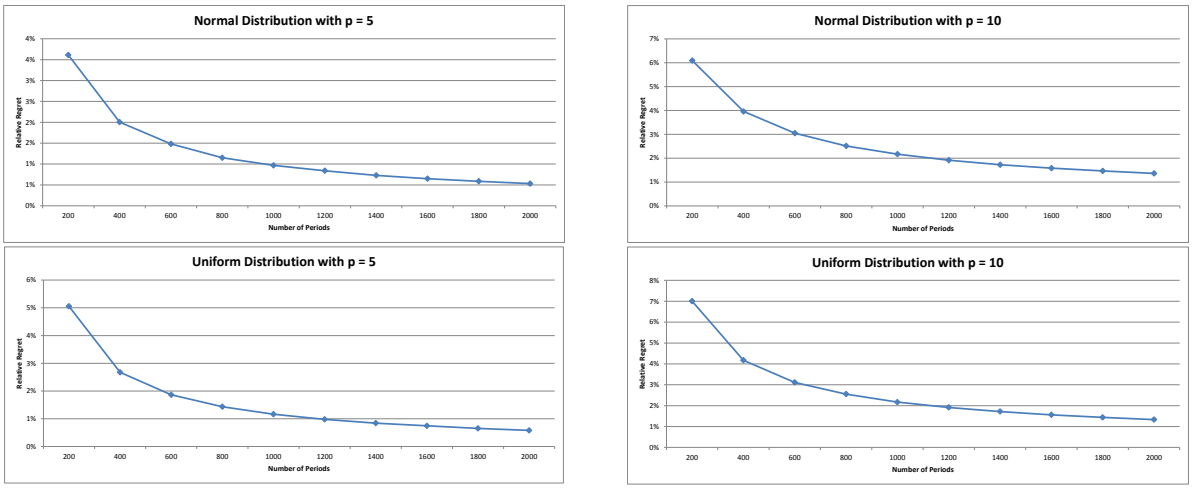

Figure 2.2: Percentage of total expected cost increase of CUP under different problem instances

Number of Periods		200	400	600	800	1000	1200	1400	1600	1800	2000
Normal	p=5	3.61%	2.00%	1.48%	1.15%	0.97%	0.84%	0.73%	0.65%	0.59%	0.53%
	p=10	6.09%	3.96%	3.05%	2.51%	2.17%	1.91%	1.72%	1.58%	1.46%	1.36%
Uniform	p=5	5.05%	2.67%	1.87%	1.44%	1.16%	0.98%	0.84%	0.75%	0.65%	0.58%
	p=10	7.00%	4.18%	3.11%	2.55%	2.17%	1.92%	1.72%	1.56%	1.44%	1.33%

Table 2.1: Percentage of total expected cost increase of CUP under different problem instances

From Figure 2.2 and Table 2.1, it is seen that CUP performs consistently well on all the tested problem instances, with a maximum average cost increment of 7.00% after 200 periods, 2.17% after 1000 periods, and 1.36% after 2000 periods.

2.5 Performance Analysis of CUP

Given a sample path $\omega = \{d_1, d_2, \ldots, \}$ of demand process, the T-period regret of our nonparametric adaptive inventory policy CUP is defined as the difference between the clairvoyant optimal cost (given the demand distribution *a priori*) and the cost incurred by CUP (which learns the demand distribution over time under censored demand information) over

T periods. More precisely,

$$\mathcal{R}_T^{\mathbf{CUP}}(\omega) = \sum_{t=1}^{T} \left(C_t^{\pi(S_t)}(\omega) - C_t^{\pi(S^*)}(\omega) \right),$$

where S_t is the base-stock level implemented by our nonparametric (closed-loop) algorithm CUP, and S^* is the (clairvoyant) optimal base-stock level defined in (2.3). The average regret of CUP is $\mathbb{E}[\mathcal{R}_T^{\mathbf{CUP}}]$, and the average regret per period is defined as $\mathbb{E}[\mathcal{R}_T^{\mathbf{CUP}}]/T$.

Theorem 2.8 below states the main result in this chapter.

Theorem 2.8. *Suppose Assumption 2.3 holds. Then, there exists some positive constant K_1, such that for each problem instance of the perishable inventory system described in §2.2, the expected regret of the cycle-update policy (CUP) satisfies*

$$\mathbb{E}\left[\mathcal{R}_T^{\mathbf{CUP}}\right] \leq K_1\sqrt{T}, \qquad \textit{for all } T \geq 1.$$

In other words, the average regret per period approaches 0 at the rate of $O(1/\sqrt{T})$.

It is known that in the general convex case (without assuming smoothness and strong convexity), this rate of $O(\sqrt{T})$ is unimprovable (see, e.g., Theorem 3.2. of Hazan (2015)).

2.5.1 A Bridging Policy – Replacement of Old Inventories (ROI)

Our general strategy towards establishing Theorem 2.8 is to compare and bound the k-th ($k = 1, 2, \ldots$) cycle cost between CUP and the clairvoyant optimal policy. A difficulty arises at the start of each cycle τ_k. More precisely, our CUP algorithm will update the base-stock level to S_k and order up to it. Due to the construction of CUP, the system is empty at the beginning of τ_k, and therefore all the S_k inventory units ordered are new with remaining lifetime m. However, the age distribution of the (clairvoyant) optimal inventory at the beginning of τ_k is unknown and impossible to determine. This creates much difficulty when we compare the cost difference between these policies going forward towards $\tau_{k+1} - 1$.

To circumvent this difficulty, we introduce a bridging policy, called the replacement of old inventories (ROI for short), between CUP and the optimal base-stock policy $\pi(S^*)$. For each sample path, similar to the optimal policy, the bridging policy ROI uses S^* as its base-stock level. However, at the beginning of τ_k ($k = 1, 2, \ldots$), ROI replaces all its inventory units (regardless of their ages) with brand new inventory units with remaining lifetime m.

We then establish that for each sample path, the total cost incurred by ROI in fact provides a lower bound on the total cost incurred by the optimal base-stock policy $\pi(S^*)$.

Proposition 2.9. *For each problem instance of the perishable inventory system described in §2.2, given any sample path $\omega = \{d_1, d_2, \ldots\}$ and any $T \geq 1$, the total cost incurred by the bridging policy ROI is less or equal to the total cost incurred by the optimal base-stock policy $\pi(S^*)$.*

Proof. It suffices to show that for a given sample path ω and a given base-stock level S, an empty system with zero initial inventory gives the lowest total cost from period 1 to any period T, among all possible configurations of initial inventory that is less or equal to S and with any age distributions.

We first make a simple yet important observation. That is, under a base-stock policy $\pi(S)$, the holding and lost-sales costs are independent of the initial inventory as well as its age distribution, and is only affected by demands and the given base-stock level S. Hence, the initial inventory and its age distribution only affect the outdating cost. To analyze the outdating cost, we consider a variant of the linear program (LP) introduced in the proof of Theorem 1.

Denote the initial inventory configuration by $\mathbf{a} = (a_1, \ldots, a_{m-1})$, where $0 \leq a_1 \leq a_2 \leq \cdots \leq a_{m-1} \leq S$. Note that a_i represents the initial inventory with remaining lifetime no more than i periods. For any given $d_1, \ldots, d_{T-1}$ and any initial inventory configuration $\mathbf{a}$, we construct a linear program $\mathbf{LP}'(S, \mathbf{a})$ as follows.

$$\min_{\{\mathbf{q},\mathbf{x}\}} \quad \sum_{t=1}^{T} q_t \tag{2.16}$$

$$\text{subject to} \quad x_{1,i} = a_i, \quad i = 1, \ldots, m-1, \tag{2.17}$$

$$x_{t,m} = S, \quad t = 1, \ldots, T, \tag{2.18}$$

$$x_{t+1,i-1} = x_{t,i} - q_{t+1}, \quad t = 1, \ldots, T-1, \;\; i = 2, \ldots, m, \tag{2.19}$$

$$q_{t+1} \geq d_t, \quad t = 1 \ldots, , T-1, \tag{2.20}$$

$$q_{t+1} \geq x_{t,1}, \quad t = 1, \ldots, T-1. \tag{2.21}$$

The decision variables are $\mathbf{q} = (q_t)_{t=1,\ldots,T}$, and $\mathbf{x} = (x_{t,i})_{t=1,\ldots,T,i=1,\ldots,m}$. For notational convenience, we denote the feasible region (2.17–2.21) by $\mathbf{\Gamma}'(S, \mathbf{a})$.

Let the unique solution satisfying (2.10) and $\mathbf{\Gamma}'(S, \mathbf{0})$ be $\mathbf{q}^0$, and the optimal solution satisfying (2.10) and $\mathbf{\Gamma}'(S, \mathbf{a})$ be $\hat{\mathbf{q}}$. Following an identical argument as that in the proof of Theorem 1, we have that $\mathbf{q}^0$ is optimal for $\mathbf{LP}'(S, \mathbf{0})$ and $\hat{\mathbf{q}}$ is optimal for $\mathbf{LP}'(S, \mathbf{a})$. Hence, to prove Proposition 2.9, it suffices to prove the following claim: The optimal objective value of $\mathbf{LP}'(S, \mathbf{0})$ is less than or equal to that of $\mathbf{LP}'(S, \mathbf{a})$ for any $\mathbf{a}$ with $0 \leq a_1 \leq a_2 \leq \cdots \leq$

$a_{m-1} \leq S$, i.e., the objective value of solution $\mathbf{q}^0$ is less than or equal to that of solution $\hat{\mathbf{q}}$.

We shall prove this claim by constructing $\hat{\mathbf{q}}$ from $\mathbf{q}^0$ in at most $T-1$ steps, where the objective value is kept non-decreasing in each step.

In the optimal solution $\mathbf{q}^0$ for $\mathbf{LP}'(S, \mathbf{0})$, we have $q^0_{t+1} = d_t$ for $t = 1, \ldots, m-1$, since $x^0_{t,1} \leq x^0_{1,t} = 0 \leq d_t$ for $t = 1, \ldots, m-1$. In fact, this means that with zero starting inventory, the system will not have any outdating units in the first $m-1$ periods. This solution $\mathbf{q}^0$ may not be feasible for $\mathbf{LP}'(S, \mathbf{a})$; however, if $\mathbf{q}^0$ turns out to be feasible, then it must be also optimal for $\mathbf{LP}'(S, \mathbf{a})$ because $\max(d_t, x_{t,1})$ remains unchanged for each $t = 1, \ldots, , T-1$.

Now consider the more involved case where $\mathbf{q}^0$ is not feasible for $\mathbf{LP}'(S, \mathbf{a})$. In this case, we have $q^0_t \leq \max(d_{t-1}, x^0_{t-1,1})$ for any $2 \leq t \leq T$. We shall construct $\hat{\mathbf{q}}$ from $\mathbf{q}^0$ in at most $T-1$ steps. Denote the solution after step k by $\mathbf{q}^k$ (while keeping its corresponding $\mathbf{x}^k$ implicit).

In the first step, we set $\mathbf{q}^1 = \mathbf{q}^0$ and carry out the following operation. If $q^0_2 = \max(d_1, x^0_{1,1})$, then we simply do nothing. Otherwise, if $q^0_2 < \max(d_1, x^0_{1,1})$, then we increase q^1_2 such that $q^1_2 = \max(d_1, x^0_{1,1})$ and decrease q^1_3 such that $q^1_3 = q^0_3 - \max(d_1, x^0_{1,1}) + q^0_2$. We keep all other entries of $\mathbf{q}^1$ unchanged, and also determine the corresponding $\mathbf{x}^1$. It is clear that the objective value remains unchanged. Then in the subsequent step $k = 2, \ldots, T-2$, we set $\mathbf{q}^k = \mathbf{q}^{k-1}$ and apply the same operation to change the entries q^k_{k+1} and q^k_{k+2} only to obtain a new $\mathbf{q}^k$ satisfying $q_{k+1} = \max(d_k, x^{k-1}_{k,1})$. The objective value remains unchanged after these operations. In the final step $k = T-1$, we set $\mathbf{q}^{T-1} = \mathbf{q}^{T-2}$, and increase q^{T-1}_T such that $q^{T-1}_T = \max(d_{T-1}, x^{T-2}_{T-1,1})$. Then, we obtain $\mathbf{q}^{T-1} = \hat{\mathbf{q}}$, with unchanged objective value in the first $T-2$ steps and a possible increase in objective value in the final step $k = T-1$. This proves the claim and the desired result then follows. □

2.5.2 Establishing the Regret Rate using ROI

With the bridging policy ROI introduced in the preceding subsection, we are ready to prove the regret rate for our CUP algorithm.

Proof of Theorem 2.8. Consider an arbitrary sample path $\omega = \{d_1, d_2, \ldots\}$ and a fixed T. We use $N = N(\omega)$ to denote the total number of cycles before period T, including possibly the last incomplete cycle. If the last cycle is not completed at T, then we truncate the cycle and also let $\tau_{N+1} - 1 = T$.

By Proposition 2.9, we know that the bridging policy ROI provides a lower bound on the (clairvoyant) optimal base-stock policy $\pi(S^*)$. We shall compare the costs between CUP

and ROI.

$$\begin{aligned}
\mathcal{R}_T^{\mathbf{CUP}}(\omega) &= \sum_{t=1}^{T} \Big(C_t^{\mathbf{CUP}}(\omega) - C_t^{\pi(S^*)}(\omega) \Big) \\
&\leq \sum_{t=1}^{T} \big(C_t^{\mathbf{CUP}}(\omega) - C_t^{\mathbf{ROI}}(\omega) \big) \\
&= \sum_{k=1}^{N} \sum_{i=\tau_k}^{\tau_{k+1}-1} \big(C_i^{\mathbf{CUP}}(\omega) - C_i^{\mathbf{ROI}}(\omega) \big) \\
&= \sum_{k=1}^{N} \big(G(S_k, (\tau_k, \tau_{k+1}); \omega) - G(S^*, (\tau_k, \tau_{k+1}); \omega) \big). \qquad (2.22)
\end{aligned}$$

Note that CUP starts with brand new inventory units in period τ_k for all $k = 1, \ldots, N$, because CUP has experienced lost-sales in the previous period $\tau_k - 1$ (by the construction of CUP). Similarly, ROI starts with brand new inventory units in period τ_k as well for all $k = 1, \ldots, N$, as we have replaced all the old inventory units in these periods with new ones. By Theorem 2.1, we have that the cycle cost function $G(S_k, (n_1, n_2); \omega)$ is convex in S_k, thus

$$G(S_k, (\tau_k, \tau_{k+1}); \omega) - G(S^*, (\tau_k, \tau_{k+1}); \omega) \leq \nabla_1 G(S_k, (\tau_k, \tau_{k+1}); \omega) \, (S_k - S^*) \, . \qquad (2.23)$$

Substituting (2.23) into (2.22) yields

$$\mathcal{R}_T^{\mathbf{CUP}}(\omega) \leq \sum_{k=1}^{N} \nabla_1 G(S_k, (\tau_k, \tau_{k+1}); \omega) \, (S_k - S^*) \, . \qquad (2.24)$$

On the other hand, by our CUP algorithm (2.13), we have that for every ω and $k = 1, \ldots, N$,

$$(S_{k+1} - S^*)^2 \leq (S_k - S^*)^2 - \frac{2\gamma}{\sqrt{k}} (S_k - S^*) \nabla_1 G(S_k, (\tau_k, \tau_{k+1}); \omega) + \frac{\gamma^2 \left(\nabla_1 G(S_k, (\tau_k, \tau_{k+1}); \omega) \right)^2}{k}. \qquad (2.25)$$

Combining (2.24) and (2.25), and taking expectation on both sides over all sample paths, we obtain

$$\mathbb{E}\left[\mathcal{R}_T^{\mathbf{CUP}}\right] \leq \quad \mathbb{E}\left[\sum_{k=1}^{N} \frac{\sqrt{k}}{2\gamma}\left((S_k - S^*)^2 - (S_{k+1} - S^*)^2\right)\right] + \mathbb{E}\left[\sum_{k=1}^{N} \frac{\gamma}{2\sqrt{k}}\left(\nabla_1 G(S_k, (\tau_k, \tau_{k+1}))\right)^2\right]. \tag{2.26}$$

We first analyze the first term on the RHS of (2.26). By some simple algebra, we have

$$\begin{aligned} & \mathbb{E}\left[\sum_{k=1}^{N} \frac{\sqrt{k}}{2\gamma}\left((S_k - S^*)^2 - (S_{k+1} - S^*)^2\right)\right] \\ \leq \; & \frac{1}{\gamma}\mathbb{E}\left[\frac{1}{2}(S_1 - S^*)^2 - \frac{\sqrt{N}}{2}((S_{N+1} - S^*)^2\right] + \frac{1}{2\gamma}\mathbb{E}\left[\sum_{k=2}^{N}\left(\sqrt{k} - \sqrt{k-1}\right)(S_k - S^*)^2\right] \\ \leq \; & \frac{1}{\gamma}\bar{S}^2\left[\frac{1}{2} + \frac{1}{2}\sum_{k=2}^{N}\left(\sqrt{k} - \sqrt{k-1}\right)\right] = \frac{\sqrt{N}}{2\gamma}\bar{S}^2 \leq \frac{\sqrt{T}}{2\gamma}\bar{S}^2. \end{aligned} \tag{2.27}$$

Then we analyze the second term on the RHS of (2.26). By (2.15), it is seen that for almost every ω, the absolute value of cycle gradient $\nabla_1 G(S_k, (\tau_k, \tau_{k+1}); \omega)$ is bounded above by $\max(h + \theta, p) \cdot (\tau_{k+1} - \tau_k)$. Noting that $\tau_{k+1} - \tau_k$ is a geometric random variable with parameter $\mathbb{P}(D > S_k)$. Letting $\mu \triangleq \mathbb{P}(D > \bar{S}) > 0$ (by Assumption 2.3), then we have

$$\mathbb{P}(D > S_k) \geq \mathbb{P}(D > \bar{S}) = \mu.$$

Denote $U \sim \text{Geo}(\mu)$ as a geometric random variable with parameter μ, then we can write

$$\begin{aligned} \mathbb{E}\left[\sum_{k=1}^{N} \frac{\gamma}{2\sqrt{k}}\left(\nabla_1 G(S_k, (\tau_k, \tau_{k+1}))\right)^2\right] \; & \leq \; \gamma\left(\max(h + \theta, p)\right)^2 \cdot \mathbb{E}[U^2] \cdot \sum_{k=1}^{T} \frac{1}{2\sqrt{k}} \\ & \leq \; \frac{\gamma(2 - \mu)\left(\max(h + \theta, p)\right)^2}{\mu^2} \cdot \sqrt{T}, \end{aligned} \tag{2.28}$$

where the second inequality follows from $\sum_{k=1}^{T} 1/\sqrt{k} \leq 2\sqrt{T}$ and

$$\mathbb{E}[U^2] = \text{VAR}(U) + (\mathbb{E}[U])^2 = \frac{2 - \mu}{\mu^2}.$$

Combining (2.27) and (2.28), we obtain

$$\mathbb{E}\left[\mathcal{R}_T^{\mathbf{CUP}}\right] \leq \frac{\sqrt{T}}{2\gamma}\bar{S}^2 + \frac{\gamma(2-\mu)\left(\max(h+\theta,p)\right)^2}{\mu^2}\cdot\sqrt{T} \leq K_1\sqrt{T} \tag{2.29}$$

for some positive constant K_1. This completes the proof of Theorem 2.8. □

Remark 2.10. If the value of μ is known *a priori*, then the theoretical bound for regret can be optimized by choosing γ to be

$$\gamma = \frac{\mu\bar{S}}{\max(h+\theta,p)\cdot\sqrt{4-2\mu}}.$$

This balances the two terms in the middle of (2.29), and it gives a minimized value of K_1 as

$$K_1 = \frac{\max(h+\theta,p)\cdot\bar{S}\cdot\sqrt{4-2\mu}}{\mu}.$$

2.6 Strongly Convex Extension

We extend our algorithm and results to the strongly convex case, and obtain an improved regret rate. A differentiable function $g(\cdot)$ defined on a convex set of $\mathbb{R}$ is called strongly convex with parameter $\lambda > 0$ (see e.g., Hazan (2015)), if for all points x, y in its domain,

$$g(y) \geq g(x) + \nabla g(x)^T(y-x) + \frac{\lambda}{2}\big(y-x\big)^2. \tag{2.30}$$

Assumption 2.11. *There exist three known finite numbers $\bar{S}$, $\underline{S}$ and λ, such that*

(i) $0 \leq \underline{S} < \bar{S}$, $\lambda > 0$,

(ii) $\underline{S} \leq S^* \leq \bar{S}$, *and* $\mathbb{P}(D_t > \bar{S}) > 0$, *and*

(iii) the probability density function $f(x)$ of single-period demand D satisfies $\inf_{x\in[\underline{S},\bar{S}]} f(x) \geq \lambda$.

Let $\bar{u} \triangleq \mathbb{P}(D > \underline{S}) > 0$ and $\underline{u} \triangleq \mathbb{P}(D > \bar{S}) > 0$. Then $1 \geq \bar{u} \geq \underline{u} > 0$. With the slightly stronger Assumption 2.11 (in place of Assumption 2.3), we can show that the expected cost of a cycle is strongly convex in the base-stock level S, and a modified version of CUP achieves a logarithmic regret rate, i.e., the average regret of CUP converges to zero at the rate of $O(\log T/T)$, which is formally stated in Theorem 2.12 below.

Theorem 2.12. *For the perishable inventory system described in §2.2, we modify CUP or Algorithm 1 as follows:*

(1) Use the projection operator $\mathbf{P}_{[\underline{S},\bar{S}]}$*, instead of using* $\mathbf{P}_{[0,\bar{S}]}$*.*

(2) Change the step-size to $\eta_k = \left(\frac{1}{\lambda(h+p)}\right)\frac{1}{k}$ *for all* k*.*

Then under Assumption 2.11, there exists some positive constant K_2*, such that for any* $T \geq 1$*, the expected regret of CUP for any problem instance satisfies*

$$\mathbb{E}\left[\mathcal{R}_T^{\mathbf{CUP}}\right] \leq K_2 \log T.$$

It is well-known that in the strongly convex case, this rate of $O(\log T)$ is unimprovable (see, e.g., Hazan (2015)).

In general, the extension from general convex case to strongly convex case is straightforward (see, e.g., Huh and Rusmevichientong (2009)). However, this extension is rather non-trivial in our model. The key reason is that we only have that the expected (not sample-path) cycle cost function is strongly convex, and the CUP algorithm involves random cycles that correlate with the random cycle costs, which leads to some technical difficulties in developing the regret bound. Indeed, when we work with expected regret, the number of random cycles depends on CUP, which then correlates with its random cycle cost, and the standard argument based on Wald's Theorem does not work. To circumvent this technical issue, we "stretch" the time horizon from period T to the T-th cycle of CUP, then show that the cumulative regret over T periods is upper bounded by the cumulative regret over T cycles plus a constant, and study the regret of the T-cycle problem.

To facilitate our analysis, we call the event $\{D > \bar{S}\}$ as β, and define N_β^i as the period in which the event β occurs the i-th time. More precisely, given a sample path $\omega = \{d_1, d_2, \ldots\}$,

$$N_\beta^{i+1}(\omega) = \inf\left\{t \geq N_\beta^i(\omega) + 1 :\ d_t > \bar{S}\right\}, \qquad N_\beta^0(\omega) = 0.$$

Recall that given a sample path ω, $C_t^{\pi(S)}(\omega)$ is the cost incurred in period t when applying a base-stock level S in every period. We first present the following auxiliary result.

Lemma 2.13. *The (clairvoyant) optimal base-stock level satisfies*

$$S^* = \underset{S}{\operatorname{argmin}}\, \mathbb{E}\left[\sum_{t=N_\beta^i+1}^{N_\beta^{i+1}} C_t^{\pi(S)}\right], \quad i = 0, 1, 2, \ldots.$$

Proof. Since the inventory system becomes empty every time event β occurs, the costs between $N_\beta^i + 1$ and N_β^{i+1} are i.i.d. random variables, $i = 0, 1, \ldots$, where we define N_β^0 as 0.

Hence, it suffices to prove

$$S^* = \operatorname*{argmin}_S \mathbb{E}\left[\sum_{t=1}^{N_\beta^1} C_t^{\pi(S)}\right].$$

Consider an arbitrary base-stock level S and an arbitrary sample path. For any $t \geq 1$, let $J(t) = \max\{k : N_\beta^k \leq t\}$ be the number of cycles completed by time t, then we have $N_\beta^{J(T)} \leq T \leq N_\beta^{J(T)+1}$. As the total cost is non-decreasing in the number of periods, we must have

$$\frac{\sum_{j=1}^{J(T)}\left[\sum_{t=N_\beta^{j-1}+1}^{N_\beta^j} C_t^{\pi(S)}\right]}{J(T)} \cdot \frac{J(T)}{T} \leq \frac{\sum_{t=1}^{T} C_t^{\pi(S)}}{T} \leq \frac{\sum_{j=1}^{J(T)+1}\left[\sum_{t=N_\beta^{j-1}+1}^{N_\beta^j} C_t^{\pi(S)}\right]}{J(T)+1} \cdot \frac{J(T)+1}{T}.$$

Since the cycles have i.i.d. length with geometric distribution of mean $1/\underline{\mu}$, it follows from renewal theory that $J(T)/T$, as well as $(J(T)+1)/T$, converge almost surely to $\underline{\mu}$. Moreover, since $\sum_{t=N_\beta^{j-1}+1}^{N_\beta^{j+1}} C_t^{\pi(S)}$ for $i = 1, 2, \ldots$ are also i.i.d., it follows from the Strong Law of Large Numbers that, with probability 1,

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} C_t^{\pi(S)} = \underline{\mu} \cdot \mathbb{E}\left[\sum_{t=1}^{N_\beta^1} C_t^{\pi(S)}\right]. \tag{2.31}$$

It can be seen that the LHS of (2.31) is almost surely bounded by $(h+\theta)\cdot\bar{S} + p \cdot \frac{1}{T}\sum_{t=1}^{T} D_t$, which is integrable, thus applying Lebesgue's Dominated Convergence Theorem we obtain

$$\lim_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\sum_{t=1}^{T} C_t^{\pi(S)}\right] = \mathbb{E}\left[\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} C_t^{\pi(S)}\right] = \underline{\mu} \cdot \mathbb{E}\left[\sum_{t=1}^{N_\beta^1} C_t^{\pi(S)}\right]. \tag{2.32}$$

Because S^* minimizes the first term of (2.32), it also minimizes the third term. □

With Lemma 2.13, we are ready to prove Theorem 2.12.

Proof of Theorem 2.12.

We define $b(t)$ as the first period after t that an event β occurs, i.e., $b(t) = \inf\left\{s \geq t : \ d_t > \bar{S}\right\}$. It is important to note that these stopping times $b(t)$'s are policy-independent. We further introduce the notation $l(T)$ to denote the end of T-th cycle of CUP, and it is clear that

$l(T) \geq T$ almost surely. With the new notation $l(T)$ and $b(t)$, we have

$$\begin{aligned}
\mathbb{E}\left[\mathcal{R}_{l(T)}^{\mathbf{CUP}} - \mathcal{R}_T^{\mathbf{CUP}}\right] &= \mathbb{E}\left[\left(\sum_{t=1}^{l(T)} C_t^{\mathbf{CUP}} - \sum_{t=1}^{l(T)} C_t^{\pi(S^*)}\right) - \left(\sum_{t=1}^{T} C_t^{\mathbf{CUP}} - \sum_{t=1}^{T} C_t^{\pi(S^*)}\right)\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\mathbf{CUP}} - \sum_{t=1}^{b(T)} C_t^{\mathbf{CUP}}\right] - \mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\pi(S^*)} - \sum_{t=1}^{b(T)} C_t^{\pi(S^*)}\right] \\
&\quad -\mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\mathbf{CUP}} - \sum_{t=1}^{l(T)} C_t^{\mathbf{CUP}}\right] + \mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\pi(S^*)} - \sum_{t=1}^{l(T)} C_t^{\pi(S^*)}\right] \\
&\quad +\mathbb{E}\left[\sum_{t=1}^{b(T)} C_t^{\mathbf{CUP}} - \sum_{t=1}^{T} C_t^{\mathbf{CUP}}\right] - \mathbb{E}\left[\sum_{t=1}^{b(T)} C_t^{\pi(S^*)} - \sum_{t=1}^{T} C_t^{\pi(S^*)}\right] \quad (2.33) \\
&\geq -\mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\mathbf{CUP}} - \sum_{t=1}^{l(T)} C_t^{\mathbf{CUP}}\right] + \mathbb{E}\left[\sum_{t=1}^{b(l(T))} C_t^{\pi(S^*)} - \sum_{t=1}^{l(T)} C_t^{\pi(S^*)}\right] \\
&\quad +\mathbb{E}\left[\sum_{t=1}^{b(T)} C_t^{\mathbf{CUP}} - \sum_{t=1}^{T} C_t^{\mathbf{CUP}}\right] - \mathbb{E}\left[\sum_{t=1}^{b(T)} C_t^{\pi(S^*)} - \sum_{t=1}^{T} C_t^{\pi(S^*)}\right], \quad (2.34)
\end{aligned}$$

where the inequality follows from that the sum of the first two terms of (2.33) is non-negative since S^* minimizes the total cost between events β by Lemma 2.13.

Since the expected one-period cost difference between any two feasible policies is bounded above by $(\bar{S} - \underline{S})\max(h + \theta, p)$, and the expected number of periods between $b(l(T))$ and $l(T)$ is $1/\underline{\mu}$ which is also the same as that between $b(T)$ and T. By (2.34), we have

$$\mathbb{E}\left[\mathcal{R}_{l(T)}^{\mathbf{CUP}} - \mathcal{R}_T^{\mathbf{CUP}}\right] \geq -\frac{2}{\underline{\mu}}(\bar{S} - \underline{S})\max(h + \theta, p). \quad (2.35)$$

Thus, in what follows we shall focus on the evaluation of $\mathbb{E}[\mathcal{R}_{l(T)}^{\mathbf{CUP}}]$, the expected regret of a T-cycle problem. Since under our CUP algorithm, τ_k is a stopping time determined by demand process and previous base-stock levels and in particular, S_{k-1}. To emphasize its dependency on S_{k-1}, in the following we shall also write it as $\tau_k(S_{k-1})$, $k = 1, 2, \ldots$.

To derive the regret for the strongly convex case, similar as in §2.5, for an arbitrary S, we define $G(S, \tau_k(S_{k-1}), \tau_{k+1}(S_k))$ as the total cost of base-stock policy S during a fixed cycle between periods $\tau_k(S_{k-1})$ and $\tau_{k+1}(S_k) - 1$, with brand new (after ordering) inventory level S in period $\tau_k(S_{k-1})$. It is important to note that not only the cost in each period is random, the number of periods in the cycle is also random. Then, the conditional expected cost of the k-th cycle is $\mathbb{E}[G(S_k, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k]$, where $F_k \triangleq ((D_1, \ldots, D_{\tau_k - 1}); (S_1, \ldots, S_k); \tau_k)$. We shall compare this conditional expected cost with that of the bridging problem ROI,

i.e., $\mathbb{E}[G(S^*, (\tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k]$, that starts in period $\tau_k(S_{k-1})$ with all brand new inventories. Our idea is to evaluate, for a fixed S_k, the cost difference between policies S and S^*, i.e.,

$$\mathbb{E}[G(S, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k] - \mathbb{E}[G(S^*, (\tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k], \tag{2.36}$$

where the expectation is taken with respect to future random demand $(D_t; t \geq \tau_k)$.

We first show that fixing S_k, $\mathbb{E}[G(S, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k]$ is strongly convex in S. By Wald's Theorem, the expected total holding and shortage cost during the cycle is

$$\frac{h\mathbb{E}[(S-D)^+] + b\mathbb{E}[(D-S)^+]}{\mathbb{P}(D > S_k)}.$$

By Theorem 1, the expected outdating cost during a cycle is also convex in S, hence

$$\big(\mathbb{E}[G(S, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k]\big)''_S \geq \frac{(h+b)f(S)}{\mathbb{P}(D > S_k)} \geq \frac{(h+b)f(S)}{\mathbb{P}(D > \underline{S})} \geq \frac{(h+b)\lambda}{\bar{\mu}} \geq (h+b)\lambda > 0. \tag{2.37}$$

This shows that $\mathbb{E}[G(S, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k]$ is strongly convex in S with parameter $(h+p)\lambda$. Therefore, applying Taylor's expansion in (2.36) on S then set $S = S_k$ we obtain

$$\begin{aligned} & \mathbb{E}[G(S_k, \tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k] - \mathbb{E}[G(S^*, (\tau_k(S_{k-1}), \tau_{k+1}(S_k)) \mid F_k] \\ \leq \; & \nabla_1 \mathbb{E}[G(S_k, (\tau_k(S_{k-1}), \tau_{k+1}(S_k))) \mid F_k](S_k - S^*) - \frac{1}{2}(h+p)\lambda(S_k - S^*)^2, \end{aligned} \tag{2.38}$$

where $\nabla_1 \mathbb{E}[G(S, (\tau_k(S_{k-1}), \tau_{k+1}(S_k))) \mid F_k]$ is the partial derivative of $\mathbb{E}[G(S, (\tau_k(S_{k-1}), \tau_{k+1}(S_k))) \mid F_k]$ with respect to the first argument S. Using (2.38), we have

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{R}^{\mathbf{CUP}}_{l(T)}\right] \leq \mathbb{E}\left[\sum_{k=1}^{T}\Big(G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) - G(S^*,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\Big)\right] \\
= \;& \mathbb{E}\left[\sum_{k=1}^{T}\Big(\mathbb{E}[G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) \mid F_k] - \mathbb{E}[G(S^*,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) \mid F_k]\Big)\right] \\
\leq \;& \mathbb{E}\left[\sum_{k=1}^{T}\nabla_1\mathbb{E}[G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) \mid F_k](S_k - S^*)\right] \\
& -\mathbb{E}\left[\sum_{k=1}^{T}\frac{1}{2}(h+p)\lambda(S_k-S^*)^2\right]. \qquad (2.39)
\end{aligned}
$$

Following the same argument used in (2.25), we have, for $k = 1, \ldots, T$,

$$
\begin{aligned}
& (S_{k+1}-S^*)^2 \\
\leq \;& (S_k-S^*)^2 - 2\eta_k(S_k-S^*)\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) + \eta_k^2\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2.
\end{aligned}
$$

Conditioning on F_k and taking expectation with respect to future demand, yield

$$
\begin{aligned}
& \nabla_1\mathbb{E}[G_k(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k))) \mid F_k](S_k-S^*) \qquad (2.40) \\
\leq \;& \frac{1}{2\eta_k}\Big(\mathbb{E}[(S_k-S^*)^2 \mid F_k] - \mathbb{E}[(S_{k+1}-S^*)^2 \mid F_k]\Big) + \frac{\eta_k}{2}\mathbb{E}\left[\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2 \mid F_k\right].
\end{aligned}
$$

Combining (2.39) and (2.40), we have

$$\begin{aligned}
\mathbb{E}\big[\mathcal{R}^{\mathbf{CUP}}_{l(T)}\big] &\leq \mathbb{E}\Big[\sum_{k=1}^{T}\frac{1}{2\eta_k}\Big(\mathbb{E}[(S_k-S^*)^2 \mid F_k]-\mathbb{E}[(S_{k+1}-S^*)^2 \mid F_k]\Big) \\
&\qquad +\frac{\eta_k}{2}\mathbb{E}\big[\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2 \mid F_k\big]-\frac{(h+p)\lambda}{2}\sum_{k=1}^{T}(S_k-S^*)^2\Big] \\
&= \mathbb{E}\Big[\sum_{k=1}^{T}\Big(\frac{1}{2\eta_k}[(S_k-S^*)^2-(S_{k+1}-S^*)^2\Big) \\
&\qquad +\frac{\eta_k}{2}\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2-\frac{(h+p)\lambda}{2}\sum_{k=1}^{T}(S_k-S^*)^2\Big] \\
&\leq \mathbb{E}\Big[\sum_{k=1}^{T}\frac{1}{2\lambda(h+p)k}\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2-T(S_{T+1}-S^*)^2\Big] \\
&\leq \mathbb{E}\Big[\sum_{k=1}^{T}\frac{1}{2\lambda(h+p)k}\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2\Big] \\
&\leq \big(\max(h+\theta,p)\big)^2\cdot\frac{2-\underline{\mu}}{2(h+p)\lambda\underline{\mu}^2}\cdot\sum_{k=1}^{T}\frac{1}{k}, \qquad (2.41)
\end{aligned}$$

where the second inequality follows from plugging in the step-size $\eta_k=\frac{1}{(h+p)\lambda k}$, and the last inequality holds because, using the identical argument used in (2.28), we have that for each $k=1,\ldots,T$,

$$\mathbb{E}\left[\big(\nabla_1 G(S_k,(\tau_k(S_{k-1}),\tau_{k+1}(S_k)))\big)^2\right]\leq\big(\max(h+\theta,p)\big)^2\cdot\frac{2-\underline{\mu}}{\underline{\mu}^2}.$$

Consequently, combining (2.41) and (2.35), we have

$$\mathbb{E}\left[\mathcal{R}^{\mathbf{CUP}}_{T}\right] \leq \big(\max(h+\theta,p)\big)^2\cdot\frac{2-\underline{\mu}}{2(h+p)\lambda\underline{\mu}^2}\cdot\sum_{k=1}^{T}\frac{1}{k}+\frac{2}{\underline{\mu}}(\bar{S}-\underline{S})\max(h+\theta,p)\leq K_2\cdot\log T,$$

for some positive constant K_2 when T is large enough. This completes the proof of Theorem 2.12. □

Remark 2.14. The algorithm, as well as the regret analysis, for the strongly convex case assumes that $\bar{\mu}$ is *not* known to the firm *a priori*. If $\bar{\mu}$ is known *a priori*, then as seen from (2.37), the strong convexity coefficient can be improved to $(h+b)\lambda/\bar{\mu}$. In this case, the step-size of the algorithm is modified to $\eta_n=\left(\frac{\bar{\mu}}{\lambda(h+b)}\right)\frac{1}{k}$, and the corresponding regret is reduced by the factor $\bar{\mu}$. □

2.7 Conclusions

We developed the first nonparametric learning algorithm for periodic-review perishable inventory systems. Our CUP algorithm converges to the long-run average cost of the best base-stock policy at the theoretical best rate of convergence. To design and analyze CUP, we first established an important structural result that the total holding, lost-sales and outdating cost is convex in the base-stock level along every sample path. We devised a novel (stochastic) cycle updating scheme for adjusting the base-stock levels, and designed a key subroutine to compute the (sample-path) gradient of total cost over a finite number of periods. Finally, we introduced a clever bridging policy (called the replacement of old inventories) that plays an important role in comparing the total cost of CUP and that of the (clairvoyant) optimal base-stock policy. Our numerical results demonstrated the effectiveness of the proposed algorithms. In this chapter we focused on lost-sales models, and it should be noted that, because of zero ordering lead time, our results and analysis extend almost immediately to the backlogging model.

CHAPTER III

Nonparametric Learning Algorithms for Optimal Base-Stock Policy in Lost-sales Inventory Systems with Positive Lead Times and Censored Demand

3.1 Introduction

The periodic-review inventory control problem with lost-sales and positive lead times is one of the most fundamental yet notoriously difficult problems in the theory of inventory management (see Zipkin (2000)). The model assumes that unmet demand at the end of each period is *lost*, rather than being backlogged and carried over to the next period. For example, in many retail applications demand can be met by competing suppliers, making lost-sales a more appropriate modeling assumption (cf. Bijvank and Vis (2011)). There is a constant delivery lead time measured by the delay between placing an order and receiving it, which leads to an enlarged state-space in which the pipeline orders need to be tracked (cf. Zipkin (2008b)). In this chapter, contrary to the classical inventory setting, we assume that the firm does not know the demand distribution *a priori* but can only collect past sales data over time. Because the sales in a period are the minimum of the actual demand and the on-hand inventory level, the demand information is *censored* (cf. Huh et al. (2009a)). The firm wishes to minimize the long-run average holding and lost-sales penalty cost per period.

Even with complete information about the demand distribution, it is well-known that the optimal policy does not possess a simple form (see Karlin and Scarf (1958), Morton (1969), Janakiraman and Roundy (2004), Janakiraman et al. (2007)). To analyze the structure of the optimal policies, Zipkin (2008b) used a partial sum of inventory to represent the state and showed that the minimum cost function is $L^\sharp$-convex, and as a result, the optimal order quantities exhibit monotonicity and bounded sensitivity (more sensitive to newer orders). Although analyzing the dynamic program with large state space yields such nice structural

properties, the computation of optimal policies remains intractable due to the well-known *curse of dimensionality*. As a result, a considerable amount of efforts has been devoted to designing various effective heuristic policies (Reiman (2004), Levi et al. (2008a), Zipkin (2008a), Lu et al. (2015), Goldberg et al. (2016), Xin and Goldberg (2016)). In particular, Huh et al. (2009b) showed that the best base-stock policy is an effective heuristic. Levi et al. (2008a) proposed a dual-balancing policy for this problem so that the expected cost of their policy is always within two times the expected optimal cost, and Chen et al. (2014a) applied the $L^\sharp$-convexity results to devise a pseudo-polynomial time approximation scheme that solves this problem within an arbitrary prespecified additive error. More recently, Xin and Goldberg (2016) showed that the best constant-order policy converges to optimality exponentially fast as lead time grows large.

As we have witnessed the recent progress for this fundamental class of problems, the incomplete information counterpart problem (under censored demand) remains relatively under-explored. In many practical scenarios (e.g., furniture retailing), the firm does not know the underlying demand distribution *a priori* and is forced to make replenishment decisions based on historical sales data. However, the sales data, as we discussed earlier, are in fact censored demand information. The joint learning and optimization problem in the underlying lost-sales system is therefore practically relevant and theoretically challenging. The only paper (and the closest to ours) in the literature is Huh et al. (2009a) who studied the exact same model and proposed an online learning algorithm whose regret against the full-information *optimal base-stock policy* is $O(T^{2/3})$ over a T-period problem. The motivations and justifications for using the optimal base-stock policy as a valid benchmark for this incomplete information problem are two-fold. First, the class of base-stock policies is easily implemented and widely used (see e.g., Janakiraman and Roundy (2004)). Second, Huh et al. (2009b) showed that, with complete information, as the unit penalty cost increases, with other parameters unchanged, the ratio of the cost of the best base-stock policy to the optimal cost converges to one. Their numerical results suggest "*when the ratio between the lost-sales penalty and the holding cost is 100, the cost of the best base-stock policy typically is within 1.5% of the optimal cost*". In many applications, this ratio "typically exceeds 200" (see Huh et al. (2009a)). We also refer interested readers to Bijvank et al. (2014) for a robustness result on the asymptotic optimality of base-stock policy in lost-sales inventory systems. Therefore, the class of base-stock policies is expected to perform very well.

An important open question left by Huh et al. (2009a) is that whether there exists a nonparametric learning algorithm whose regret matches the theoretical lower bound $O(\sqrt{T})$.

3.1.1 Main Results and Contributions

This chapter provides an affirmative answer to the open question left by Huh et al. (2009a). More specifically, for the periodic-review inventory control problem with lost-sales and a positive lead time $L \geq 1$ under censored demand information, we present a new non-parametric learning algorithm, termed the *simulated cycle-update algorithm* (SCU for short), and show that the expected regret, defined as the difference in cost between the SCU and the optimal base-stock policy, is in the order of $O(\sqrt{T})$ for a T-period problem, which matches the theoretical lower bound (see Theorem 3.3 and Proposition 3.4). Our numerical results also show that the SCU algorithm performs better than the learning algorithm proposed in Huh et al. (2009a).

The SCU algorithm belongs to the broad family of *online gradient decent* (OGD) type of algorithms developed for various other inventory systems (cf. Burnetas and Smith (2000), Huh and Rusmevichientong (2009), Shi et al. (2015), Zhang et al. (2016)). Most studies on lost-sales inventory systems, with the exception of Huh et al. (2009a), considered models with zero lead times, that are significantly easier to analyze. One major challenge is that with positive lead times, each order placed has a prolonged impact (for at least a lead time of L periods) on the state of the system as well as the cost. Conventional online learning algorithms in the literature cannot be readily adapted to such a stochastic system, due to this lasting impact on decision-making and the complex system dynamics.

To tackle the aforementioned challenge, at a high-level, we develop a random cycle-updating rule (on the base-stock levels) based on another simulated system running in parallel, so that the (prolonged) cost impact of revising a target base-stock level can be readily quantified and compared between two feasible policies. Next, we highlight the main novelties of our approach below.

(a) First, our SCU algorithm cyclically updates base-stock level in a subset of periods termed the *triggering periods*. More specifically, the triggering periods are sequentially determined whenever another parallel auxiliary simulated system (operating under a lower base-stock level) experiences no lost-sales for L consecutive periods, then it triggers the beginning of a new cycle. The intuition is as follows. Consider two systems operating under different base-stock levels, if both systems experience no lost-sales for L consecutive periods, then the difference in state between these two systems would be only in the on-hand inventory, as both systems would share the same pipeline inventories. As a result, we can effectively compare the costs of any two feasible policies within a cycle (between two consecutive triggering periods). It can also be shown that the cost of

any feasible base-stock policy within each cycle is convex with respect to the base-stock level. Note that this needed convexity result does not hold for pre-determined fixed cycles, where the initial state at the start of each cycle remains unknown.

(b) The second idea is that we introduce a new concept termed *withheld on-hand inventory* in which we iteratively temporarily mark off some inventory units (according to a well-defined rule). The purpose of introducing this concept is to trigger ordering decisions that allow us effectively learn about demand. Note that we are not throwing these withheld inventory units away, but rather we pretend them to be nonexistent when ordering decisions are made. We use the withheld on-hand inventory to serve demand only when all the other on-hand inventory has been consumed. The rationale is as follows. By temporarily marking off these inventory units, we will order minimum extra inventory to maintain no less on-hand inventory than the simulated system, thereby allowing us to gather sufficient demand information to keep the simulated system running properly. If it happens that the on-hand inventory is less than the simulated system, then when our system experiences a lost-sales, we will be unable to determine if the simulated system also experiences a lost-sale or not. While having the withheld on-hand inventory is necessary to run the simulated system, we show that the additional average regret introduced by the withheld on-hand inventory is bounded by $O(\sqrt{T})$, so that it does not affect the overall regret bound.

(c) The third idea is that we use a double-phase approach to obtain a biased but good cost gradient estimator. The reason for introducing this new approach is that, with positive lead times, whenever we revise the base-stock level, it is not possible to immediately adjust the on-hand inventory level and therefore we may not have enough demand information to extract the cost gradient within a cycle. The cost gradient obtained by our double-phase approach is subject to estimation bias. However, we show that this estimation bias vanishes by establishing some convergence results for Markov chains with continuous state space (or Harris chains).

There are key differences between the results in this chapter and Huh et al. (2009a). The first difference is the cycles constructed: The cycles in Huh et al. (2009a) are pre-determined and increasing in length (with cycle k containing $\lceil\sqrt{k}\rceil$ periods), while the cycles in the SCU algorithm have random lengths. The second difference is in the estimation of gradient: The estimate of gradient in SCU is based on data from the second phase of each cycle, while the estimate in Huh et al. (2009a) only uses demand information from *one period* of each cycle (considering the fact that their cycle lengths are increasing). Result-wise, the main

improvement is that the regret upper bound of the SCU algorithm matches the theoretical lower bound of any learning algorithms, which closes the gap.

3.1.2 Outline and General Notation

The rest of this chapter is organized as follows. In §3.2, we formally describe the periodic-review inventory systems with lost-sales and positive lead times under censored demand information. In §3.3, we introduce the simulated cycle-update (SCU) algorithm and offer a detailed discussion on the main ideas underlying its algorithmic design. In §3.4, we analyze the performance of SCU and discuss how to change SCU to achieve a better numerical performance when the demand is uncensored (see §3.4.4). In §3.5, we test the empirical performance of SCU against the algorithm proposed in Huh et al. (2009a) as well as the uncensored counterpart algorithm of SCU proposed in §3.4.4. Finally, we conclude the chapter in §3.6.

For any real numbers x and y, we denote $x^+ = \max\{x, 0\}$. The indicator function $\mathbb{1}(A)$ takes value 1 if A is true and 0 otherwise. The projection function is defined as $\mathbf{P}_{[a,b]}(x) = \min[b, \max(x, a)]$ for any real numbers x, a, b.

3.2 Model Description

Consider a periodic-review inventory system with lost-sales, positive ordering lead times and censored demand. The demands over periods $\{D_1, D_2, \ldots, D_t, \ldots\}$ are i.i.d. continuous random variables. Let t denote the period, $t = 1, 2, \ldots$, and let D denote a generic one-period demand, which is non-negative with $\mathbf{E}[D] > 0$. The ordering lead time is a fixed integer $L \geq 1$. Contrary to the classical formulation, the firm has no access to the true demand distribution *a priori*. The firm can only observe the past censored demand data and adjust the ordering decisions on the fly.

For the lost-sales inventory system under consideration, any new order will stay in the pipeline for L periods before arrival. Hence, together with the on-hand inventory, we need to use an $(L+1)$-dimensional vector to keep track of the inventory information. For every period t, the starting inventory, or state of the system, is denoted by

$$\mathbf{x}_t = [q_{t-1}, \ldots, q_{t-L+1}, I_t],$$

where I_t is the on-hand inventory at the beginning of period t, and q_k is the order placed in period k. Let $\mathbf{y}_t = [q_t, q_{t-1}, \ldots, q_{t-L+1}, I_t]$ denote the inventory after ordering in period t.

Clearly, all the entries of $\mathbf{x}_t$ and $\mathbf{y}_t$ are non-negative. For simplicity, let $q_k = 0$ for all $k \leq 0$.

For any feasible policy π, the sequence of events in each period t, $t = 1, 2, \ldots$, is as follows. (Note that all the states and decisions depend on π, but in general we shall make the dependency implicit for notational simplicity. However, whenever necessary, we use $\mathbf{x}_t^\pi$ and q_t^π to represent the state and ordering decision of policy π in period t.)

(i) At the beginning of each period t, the firm observes the starting inventory vector $\mathbf{x}_t = [q_{t-1}, \ldots, q_{t-L+1}, I_t]$, and makes a replenishment decision $q_t \geq 0$.

(ii) Then, the demand D_t is realized, and we denote its realization by d_t. The demand is satisfied to the maximum extent by on-hand inventory I_t. Since demand is censored, the firm only observes sales quantity $\min(d_t, I_t)$. Thus, if $d_t \geq I_t$, then the firm does not know the exact demand.

(iii) At the end of the period, each remaining on-hand inventory unit incurs a per-unit holding cost h, and each unsatisfied demand unit incurs a per-unit lost-sales penalty cost p. As a result, the cost in period t, denoted by C_t^π, is

$$C_t^\pi = h(I_t - d_t)^+ + p(d_t - I_t)^+.$$

Note that the lost-sales quantity and its penalty cost (as an opportunity cost) are unobservable to the firm due to demand censoring.

(iv) At last, the system proceeds to period $t+1$ with system state $\mathbf{x}_{t+1}$ given by

$$\mathbf{x}_{t+1} = \left[q_t, \ldots, q_{t-L+2}, I_{t+1} = q_{t-L+1} + (I_t - d_t)^+\right]. \tag{3.1}$$

The objective is to find an ordering policy, based on historical sales information, that minimizes the expected average cost of the lost-sales inventory system with positive lead times.

As seen from §4.1, even when the demand distribution is known, the computation of an optimal policy is intractable due to the curse of dimensionality. When the demand distribution is not known *a prori*, it becomes even harder if we use the optimal policy as the benchmark. Hence in this chapter, we follow Huh et al. (2009a) to use the best base-stock policy as the benchmark. The class of base-stock policies is parametrized by a single parameter, the base-stock level $S \geq 0$. Under a base-stock policy with a base-stock level S, the ordering quantity in period t is $q_t = \left(S - I_t - \sum_{i=t-L+1}^{t-1} q_i\right)^+$. Note that $I_t + \sum_{i=t-L+1}^{t-1} q_i$ is the inventory position at the beginning of period t. Thus, essentially, the base-stock policy

orders to raise the inventory position to S if the starting inventory position is less than S, and orders nothing otherwise. We refer to Huh et al. (2009b) and Huh et al. (2009a) for the asymptotic optimality and the effectiveness of base-stock policies.

In this chapter, we will design an adaptive learning inventory policy that only uses the past sales data, and show that the expected average cost of the policy converges to that of the optimal base-stock policy at rate $O(1/\sqrt{T})$, which matches the theoretical lower bound.

3.3 Nonparametric Algorithm - Simulated Cycle-Update Policy (SCU)

We present a learning algorithm which we refer to as *simulated cycle-update policy* (SCU for short). Before introducing the SCU policy, we make the following assumption on the optimal (full information) base-stock level S^*.

Assumption 3.1. *There exist two known finite numbers $\underline{D}$ and $\bar{D}$, such that (i) $\mathbf{P}\left(D \leq \underline{D}\right) = c_1 > 0$, (ii) $\mathbf{P}\left(D \geq \bar{D}\right) = c_2 > 0$, and (iii) $S^* \in [(L+1)\cdot \underline{D},\ (L+1)\cdot \bar{D}]$.*

Assumption 3.1(i)(ii) essentially means that the decision maker has some prior knowledge about the tail demand distribution, and Assumption 3.1(iii) gives an upper and a lower bound on the optimal base-stock level S^*, which is a predominant assumption in the nonparametric learning literature in inventory management (see, e.g., Huh and Rusmevichientong (2009), Huh et al. (2009a), Shi et al. (2015), Zhang et al. (2016)).

It is worthwhile noting that when c_1 and c_2 in Assumption 3.1(i)(ii) are sufficiently small, then Assumption 3.1(iii) is automatically satisfied. The reasoning is as follows. Under a base-stock policy, the firm orders $\min(d_t, I_t)$ in each period. If demand is higher than $\bar{D}$ with very low probability, then the ordering quantity is higher than $\bar{D}$ with very low probability. As a result, the total pipeline inventories should be lower than $L\cdot\bar{D}$ with very high probability. Also, the firm would like to keep the on-hand inventory lower than $\bar{D}$ to avoid a high holding cost. Hence, the optimal base-stock level is not likely to be higher than $(L+1)\cdot\bar{D}$. Similarly, if the demand is lower than $\underline{D}$ with very low probability, then the optimal base-stock level is not likely to be lower than $(L+1)\cdot\underline{D}$. For notational convenience, we let $\underline{S} \triangleq (L+1)\underline{D}$ and $\bar{S} \triangleq (L+1)\bar{D}$.

3.3.1 Random Cycles, the Simulated System, and the Function G

One of the main challenges in designing our algorithm is that the total cost of the system cannot be readily written in a form that is amenable for online optimization. To overcome

this, our first step is to divide the time periods into appropriately designed learning cycles, and then update the inventory target levels from cycle to cycle (instead of from period to period). That is, we use the (censored) demand information collected from one particular cycle to update the base-stock level for its subsequent cycle.

Random cycles based on the simulated system. As its name suggests, the SCU algorithm is designed based on a concurrently simulated inventory system. This system is run in the background and it implements a base-stock policy $\underline{S}$ where $\underline{S} = (L+1)\underline{D}$. For convenience, we shall refer to this simulated system as the simulated $\underline{S}$-system. In what follows, we shall define a sequence of cycles using the simulated $\underline{S}$-system; and the SCU algorithm updates the base-stock level at the beginning of each cycle using data collected from the SCU-system in the previous cycle.

Specifically, we define a "triggering event" as the event that simulated $\underline{S}$-system experiences no stockouts for L consecutive periods. We call the period after a triggering event a triggering period. Let t_k denote the k-th triggering period, and for convenience, we let period 1 be the first triggering period. Mathematically, the triggering periods are defined by

$$t_1 = 1, \quad t_{k+1} = \min\left\{ n \;\middle|\; n \geq t_k + L, I_i^{\underline{S}} \geq D_i, \text{ for all } n - L \leq i < n \right\}.$$

Note that in this definition, once a triggering period is found, it resets the counter for the consecutive number of stockout periods to zero. Huh et al. (2009a) have shown that, the on-hand inventory of a lost-sales inventory system with positive lead times is non-decreasing in its base-stock levels. This implies that, if t is a triggering period, then it is also a triggering period for the inventory system operating under any base-stock policy $S \geq \underline{S}$, and therefore the pipeline inventory under any base-stock policy $S \geq \underline{S}$ is $(d_{t-1}, d_{t-2}, \ldots, d_{t-L})$.

The cycles are defined as follows: Let τ_k be the first period of cycle k, $k = 1, 2, \ldots$, then $\tau_1 = t_1 = 1$, and for $k > 1$, $\tau_k = t_{2(k-1)}$. That is, the first cycle starts in period 1, and starting from the second cycle, each cycle contains two phases, and each phase begins with a triggering period. Let τ'_k denote the first period of the second phase of cycle k, then $\tau'_k = t_{2k-1}$, $k = 2, 3, \ldots$, as depicted in Figure 3.1. Note that the cycle length is *a priori* random, it is *independent* of the learning algorithm.

The function G. Next, we define an important function, $G(S, a, b)$, which denotes the total cost from period a to period b (both included), by using a base-stock level $S \geq \underline{S}$, and its starting state is specified as follows: If $a \geq L+1$ then we assume that the starting state in period a is $\left[d_{a-1}, d_{a-2}, \ldots, d_{a-L}, S - \sum_{i=a-L}^{a-1} d_i\right]$, otherwise the starting state is $[S, 0, \ldots, 0]$ in period a. Note that the function $G(S, a, b)$ also clearly depends on $(d_{a-1}, d_{a-2}, \ldots, d_{a-L})$, but

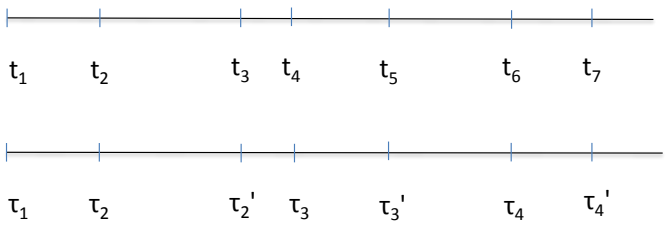

Figure 3.1: For $k \geq 1$, t_k is a triggering period, and τ_k and τ_k' are first periods of the two phases of cycle $k \geq 2$.

we will make this dependency implicit for notational simplicity. We shall only consider the vector $(d_{a-1}, d_{a-2}, \ldots, d_{a-L})$ satisfying $\sum_{t=1}^{L} d_{a-t} \leq \underline{S}$. By Theorem 8 in Janakiraman and Roundy (2004), we know that $G(S, a, b)$ is convex in the base-stock level S. Let $\nabla G(S, a, b)$ denote the *partial* gradient of $G(S, a, b)$ with respect to S. The computation of $\nabla G(S, a, b)$ is discussed in §3.3.2.

3.3.2 The Simulated Cycle-Update (SCU) Policy

With the necessary definitions in place, we present the detailed SCU algorithm. For convenience, we let $\tau_1' = \tau_1 = 1$. For any vector $\mathbf{x}$, we use $\sum \mathbf{x}$ to denote the sum of all its entries, i.e., $\sum \mathbf{x} = \sum_{i=1}^{L} x_i$ if $\mathbf{x} = (x_1, \ldots, x_L)$. Let the step-size $\eta_k = \gamma/\sqrt{k}$ for all $k = 1, 2, \ldots$, for some positive constant γ.

The algorithm below involves a new concept called the *withheld inventory* – in each period t, the algorithm divides the total on-hand inventory I_t^{SCU} into two parts, namely, the withheld on-hand inventory denoted by $\hat{I}_t^{SCU}$, and the regular (or non-withheld) on-hand inventory denoted by $\tilde{I}_t^{SCU}$. The detailed evolution of the withheld inventory is given explicitly in the description of the algorithm. We shall defer discussing the main ideas behind this concept in §3.3.3.

Algorithm 1: Simulated Cycle-Update Algorithm (SCU)

Step 0 (Initialization):

- Start with an arbitrary target base-stock level $S_1 \in [\underline{S}, \bar{S}]$.
- Initialize the withheld on-hand inventory $\hat{I}_1^{SCU} = 0$.
- Set the initial inventory of both the SCU- and the simulated $\underline{S}$- systems to $\mathbf{x}_1^{SCU} = \mathbf{x}_1^{\underline{S}} = \mathbf{0}$.

- Set the counter for consecutive no lost-sales events for the simulated system $\psi = 0$. (Recall that cycles are defined using the simulated $\underline{S}$-system. In our SCU algorithm, we use ψ to record the number of consecutive no stockout periods in the simulated $\underline{S}$-system. When ψ reaches L, it signals a triggering period and ψ is reset to 0.)

Step 1: For the first cycle $k = 1$ starting with period $t = 1$, do the following.

1(a). Order q_t^{SCU} for the SCU system and $q_t^{\underline{S}}$ for the simulated $\underline{S}$ system as follows:

$$q_t^{SCU} = \left(S_k - \sum \mathbf{x}_t^{SCU} + \hat{I}_t^{SCU}\right)^+, \tag{3.2}$$

$$q_t^{\underline{S}} = \left(\underline{S} - \sum \mathbf{x}_t^{\underline{S}}\right)^+. \tag{3.3}$$

The ordering decision in the SCU-system is given as follows: It implements the modified base-stock policy S_k based on the regular inventory only (by temporarily ignoring the withheld inventory $\hat{I}_t^{SCU}$). More precisely, since the regular (or non-withheld) inventory position is $\sum \mathbf{x}_t^{SCU} - \hat{I}_t^{SCU}$, we order q_t^{SCU} of (3.2) to raise the regular inventory to S_k.

1(b). Observe the sales quantity $\min(d_t, I_t^{SCU})$, and update the withheld on-hand inventory by

$$\hat{I}_{t+1}^{SCU} := \left[\hat{I}_t^{SCU} - \left(\min(d_t, I_t^{SCU}) - \tilde{I}_t^{SCU}\right)^+\right]^+. \tag{3.4}$$

The demand fulfillment rule for the SCU-system is given as follows: It first uses the regular (non-withheld) on-hand inventory to satisfy demand, and then uses the withheld on-hand inventory to satisfy demand (only after the regular on-hand inventory has been fully consumed). Thus, we update the withheld on-hand inventory following (3.4).

1(c). Update the states of both the SCU system and the simulated $\underline{S}$ system following the system dynamics (3.1), with the demand in period t for the simulated $\underline{S}$-system being replaced by $\min(d_t, I_t^{SCU})$.

1(d). If there is no lost-sales in the simulated $\underline{S}$ system, set the counter for consecutive no lost-sales events for the simulated system $\psi := \psi + 1$. Otherwise reset $\psi := 0$.

1(e). If $\psi = L$, then label period $t + 1$ as a triggering period and reset $\psi = 0$. The sales data is used to compute $\nabla G(S_1, 1, t)$ following a well-defined subroutine presented in §3.3.4, and update the base-stock level S_2 for the second cycle as

$$S_2 = \mathbf{P}_{[\underline{S}, \bar{S}]}\left(S_1 - \eta_1 \nabla G(S_1, 1, t)\right).$$

Set $\tau_2 := t+1$, and update the withheld on-hand inventory by

$$\hat{I}^{SCU}_{\tau_2} := \left(\hat{I}^{SCU}_{\tau_2} - (S_2 - S_1)\right)^+,$$

and proceed to Step 2 with $k = 2$. On the other hand, if $\psi < L$, then repeat procedures 1(a) to 1(e) with $t := t+1$ if $t < T$, and stop otherwise.

Step 2: For cycles $k \geq 2$, each cycle contains two phases.

Phase 1: Start from period $t = \tau_k$.

2(a) Conduct procedures 1(a)–1(d) in Step 1.

2(b) If $\psi = L$, then set $\tau'_k = t+1$ and $\psi = 0$, and proceed to Phase 2. Otherwise, repeat 2(a) with $t := t+1$ if $t < T$, and stop otherwise.

Phase 2: Start from period $t = \tau'_k$.

2(a') Conduct procedures 1(a)–1(d) in Step 1.

2(b') If $\psi = L$, then set $\tau_{k+1} := t+1$. Update the target base-stock level for the next cycle as

$$S_{k+1} = \mathbf{P}_{[\underline{S},\bar{S}]}\left(S_k - 2\eta_k \nabla G\left(S_k, \tau'_k, t\right)\right),$$

Note that here we double the gradient of the second phase to estimate the gradient of the whole cycle. We then update the withheld on-hand inventory by

$$\hat{I}^{SCU}_t := \left(\hat{I}^{SCU}_t - (S_{k+1} - S_k)\right)^+.$$

Set $\psi := 0$, $k := k+1$, and repeat Step 2. If $\psi < L$, then repeat 2(a') with $t := t+1$ if $t < T$, and stop otherwise.

This concludes the description of the SCU algorithm.

Example 3.2. In Figure 3.2, we use a simple example to illustrate how the dynamics of the SCU-system evolves and how it differs from the G-system. In this example, the lead time $L = 2$. We simulate the $\underline{S}$-system to determine the cycle length. Consider two cycles and suppose $S_2 > S_1$. In this case, the SCU policy will order more in the first period of cycle 2 to increase the inventory position from S_1 to S_2. Compared with the G-system, the inventory vectors of the two systems differ by 1 unit until τ'_2. In the fourth cycle, we have $S_4 < S_3$. In this case, the SCU policy marks 2 units of on-hand inventory as the withheld inventory. We

can see that the withheld inventory amount keeps dropping, and apart from the withheld inventory, the two systems are the same. Note that in the second period of this cycle, by ignoring the withheld on-hand inventory, the SCU policy orders 4 (instead of 5), which is the same as what the G-system orders.

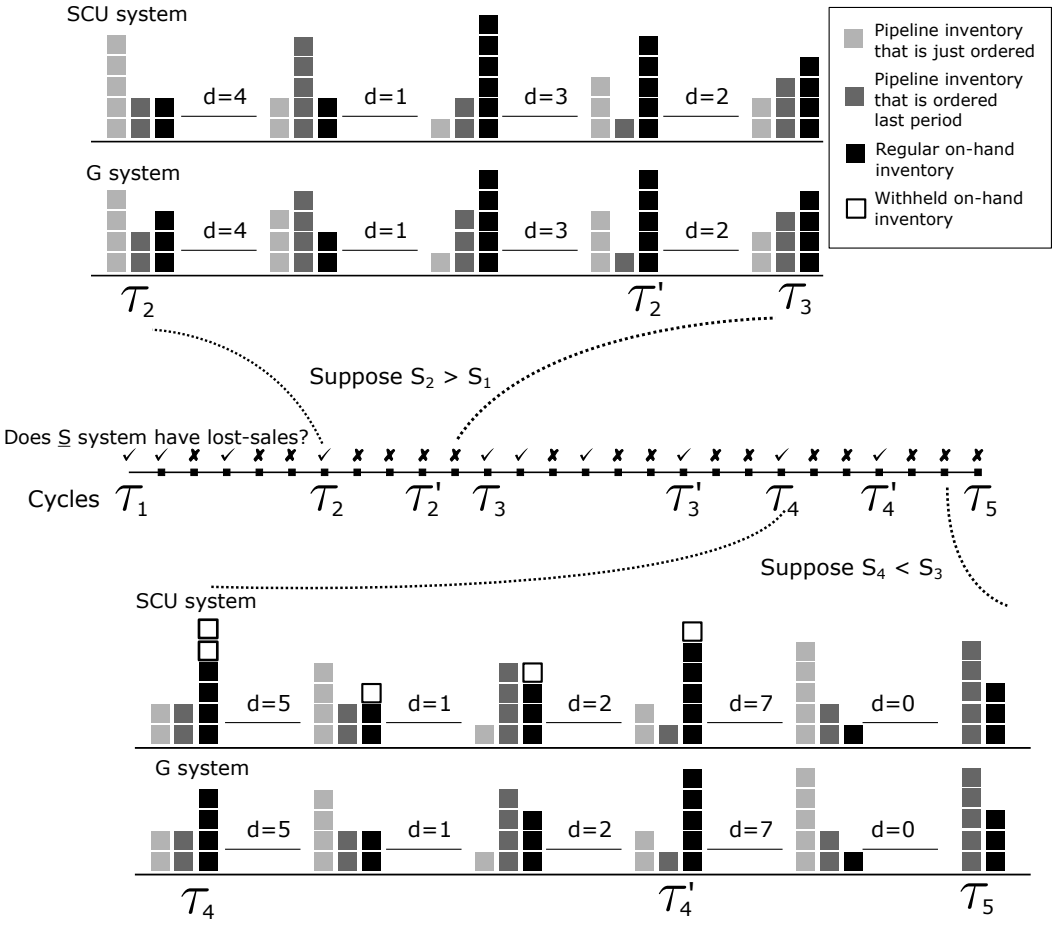

Figure 3.2: An graphical illustration of SCU policy with $L = 2$. For illustration purpose, all the numbers are integers

3.3.3 Main Ideas of the SCU Algorithm

The SCU algorithm involves several main ideas, and we have discussed one of them, which is the construction of random cycles based the simulated system in §3.3.1. In the following, we will discuss the rest of the challenges in the algorithmic design and how we resolve them.

Simulation of the $\underline{S}$-system. We have described the simulated $\underline{S}$-system in §3.3.1, and the main purpose of this simulated system is to help decide triggering periods and form

cycles.

An immediate important question is whether the simulated $\underline{S}$-system can be correctly simulated. Since the SCU algorithm is implemented under sales data (or censored demand), we do not know the exact demand in a period whenever a stockout occurs. For example, if the on-hand inventory level in our SCU-system in a period is zero, we do not know the true demand for this period since the sales is always zero regardless of demand. In this case, we cannot simulate the $\underline{S}$-system in question (as the system gives us insufficient demand information). This shows that we must design the learning algorithm in such a way that it yields the necessary demand information for simulating the $\underline{S}$-system correctly.

A sufficient condition for achieving the correct simulation of the $\underline{S}$-system is to ensure that our SCU-system always has no lower on-hand inventory than the simulated $\underline{S}$-system. To see that, suppose the states of our system and the simulated $\underline{S}$-system at the beginning of period t are $\left(q^a_{t-1}, q^a_{t-1}, \ldots, q^a_{t-L+1}, I^a_t\right)$, $a = SCU,\ \underline{S}$, respectively. Then, the on-hand inventory level at the beginning of period $t+1$ will be

$$I^a_{t+1} = q^a_{t-L+1} + \left(I^a_t - d_t\right)^+, \qquad a = SCU,\ \underline{S}.$$

In general, we may not be able to simulate the $\underline{S}$ system using only the sales quantity $\min(I^{SCU}_t,\ d_t)$. However, if $I^{SCU}_t \geq I^{\underline{S}}_t$, then the $\underline{S}$-system can be correctly simulated because

$$I^{\underline{S}}_{t+1} = q^{\underline{S}}_{t-L+1} + \left(I^{\underline{S}}_t - d_t\right)^+ = q^{\underline{S}}_{t-L+1} + \left[I^{\underline{S}}_t - \min\left(d_t, I^{SCU}_t\right)\right]^+.$$

This shows that, under the condition $I^{SCU}_t \geq I^{\underline{S}}_t$ for all t, the $\underline{S}$-system can be correctly simulated by pretending that the demand in period t is equal to the sales quantity in the SCU-system. This sufficient condition is will be carefully embedded in our algorithmic design, which will be formally established in Lemma 3.6 in §3.4.2.

The bridging G-system, and its connection with the SCU-system. We introduce an auxiliary (non-implementable) bridging system, which we refer to as the G-system. The G-system is defined as follows: i) For cycle $k = 1, 2, \ldots$, it implements base-stock policy S_k as prescribed by the algorithm (which starts in period τ_k and ends in period $\tau_{k+1} - 1$); and ii) its state at the beginning of period 1 is set at $(S_1, 0, \ldots, 0)$, and its state at the start of cycle $k \geq 2$ (i.e., in period τ_k) is *artificially set* as

$$\left(d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L}, S_k - \sum_{t=\tau_k - L}^{\tau_k - 1} d_t\right).$$

Note that the main feature in the G-system is that, its inventory state at the beginning of each cycle is artificially set (hence not implementable). This change of state essentially removes the end-of-cycle effect (from the previous cycle) when implementing a different base-stock policy for the new cycle. Thus the total cost of the G-system, with the total number of cycles denoted by N, can be written as

$$\sum_{k=1}^{N} G(S_k, \tau_k, \tau_{k+1} - 1). \tag{3.5}$$

It is well-known that dynamic optimization problem with cost function (3.5) is amenable for online algorithm design (see, e.g., Hazan (2015)). Our algorithm will be based on the stochastic gradient descent method for minimizing objective function (3.5) of the G-system, which requires the gradient evaluation of G with respect to S_k.

However, there are still several significant challenges in evaluating the G-system based on the (censored) demand data collected from the SCU-system, due to the difference in their starting states. Our learning algorithm modifies, using historical (censored) demand information, the base-stock level from cycle to cycle. Clearly, the prescribed new base-stock level for the next cycle can be either higher or lower than the previous base-stock level, each creating critical issues. This is because, due to positive lead times, when a new base-stock level is suggested by the SCU algorithm for the following cycle, there is a random transition time before this new base-stock policy can be fully implemented (with the desired starting state).

Now, suppose that the base-stock level for period $\tau_k - 1$ is S_{k-1}, and that the SCU algorithm recommends a new base-stock policy S_k in period τ_k for the next cycle. In the following, we discuss the main issues encountered for the two cases, $S_k < S_{k-1}$ and $S_k \geq S_{k-1}$. To tackle the difficulties arising in the first case $S_k < S_{k-1}$, we shall introduce a new concept called the withheld inventory. To tackle the difficulties arising in the second case $S_k \geq S_{k-1}$, we adopt a double-phase gradient estimation approach.

The concept of withheld inventory. In the first case where $S_k < S_{k-1}$, the inventory position of the SCU-system in the first few periods of cycle k may be higher than S_k even if no order is placed. In this case, if we blindly and naively implement the base-stock policy S_k, we may suffer from a severe consequence that the $\underline{S}$-system may not be simulated correctly. Indeed, under this case, the desired order quantity at the beginning of cycle k may be 0 (if the inventory position after satisfying demand in period $\tau_k - 1$ is still no lower than S_k). However, ordering 0 in period τ_k will affect the on-hand inventory level of the SCU-system at the start of period $\tau_k + L$. If, for instance, the on-hand inventory level of the SCU-system

at the beginning of period $\tau_k + L$ is 0, then it will reveal no demand information for period $\tau_k + L$. As a consequence, as we discussed earlier in §3.3.3, the $\underline{S}$-system cannot be simulated correctly in period $\tau_k + L$. This shows that we cannot naively follow the exact base-stock policy S_k, but need to revise the policy in such a way that the sufficient demand information for simulating the $\underline{S}$-system can be yielded.

Our approach to resolve this issue is to order the minimum but sufficient quantity to guarantee the correct simulation of the $\underline{S}$-system, but mark any excess on-hand inventory as what-we-define *withheld inventory*. (Note that the detailed formulae for the withheld inventory and its evolution are given in the description of the SCU algorithm.) At a high-level, in each period t, we shall divide the total on-hand inventory I_t^{SCU} into two parts, namely, the withheld on-hand inventory denoted by $\hat{I}_t^{SCU}$, and the regular (or non-withheld) on-hand inventory denoted by $\tilde{I}_t^{SCU}$. When making replenishment decisions in cycle k, we operate the base-stock policy S_k based on the regular inventory position only. More precisely, the order quantity is given in (3.4), the difference between S_k and the regular inventory position (rather than the total inventory position). Also, when satisfying demands, the withheld inventory is used only when the regular on-hand inventory has been exhausted.

The proposed (modified) base-stock policy based only on regular (or non-withheld) inventory position enables the system to *gradually* adjust its base-stock level from S_{k-1} down to S_k. This modification is essential because it ensures that the SCU system orders enough (more than what the $\underline{S}$-system orders) in each period, in order to gather sufficient demand information that guarantees the correction simulation of the $\underline{S}$-system. Note that when all the withheld on-hand inventory is consumed by demand, the SCU-system will coincide with the G-system. The exact connection between the SCU-system with the withheld inventory and the G-system will be formally established in Lemma 3.7 in §3.4.2, which plays an essential role in comparing costs between our SCU algorithm and the optimal base-stock policy.

The double-phase gradient estimation. In the second case where $S_k \geq S_{k-1}$, because the G-system artificially sets its on-hand inventory level to $S_k - \sum_{t=\tau_k-L}^{\tau_k-1} d_t$ at the beginning of period τ_k, this particular on-hand inventory level could be higher than the on-hand inventory level of the SCU-system at the beginning of cycle k. At the beginning of period τ_k, the inventory vector of the SCU-system, having just experienced no stockouts for L consecutive periods, is

$$\Big(d_{\tau_k-1}, d_{\tau_k-2}, \ldots, I_{\tau_k}^{SCU} = S_{k-1} - \sum_{t=\tau_k-L}^{\tau_k-1} d_t + \hat{I}_t\Big).$$

This is different from the starting state of the G-system, which according to our definition is

$$\left(d_{\tau_k-1}, d_{\tau_k-2}, \ldots, S_k - \sum_{t=\tau_k-L}^{\tau_k-1} d_t \right).$$

Since the on-hand inventory level of the G-system could be higher than that of the SCU-system in period τ_k, it leads to the following critical issue: Due to demand censoring and the same reasoning as in the first case, we may not able to obtain sufficient demand information from the SCU-system to compute the total cost, nor its gradient, of the G-system during periods $\tau_k, \tau_k + 1, \ldots, \tau'_k - 1$ with respect to the base-stock level S_k. This is precisely the reason why we need to use two phases for each cycle $k \geq 2$ in our algorithm design: In the triggering period τ'_k, having just experienced no stockouts for L consecutive periods, the G-system and our SCU-system become identical during the second phase if all the withheld inventory in the SCU-system is ignored.

Regardless of $S_k \leq S_{k-1}$ or $S_k \geq S_{k-1}$, we will show in Lemma 3.7 in §3.4.2 that during the second phase of each cycle, the SCU-system always has no less on-hand inventory than that of the G-system . This enables us to compute (and simulate) the total cost of the G-system during the second phase of cycle k as well as its gradient with respect to S_k. Thus, we can construct an estimate of the gradient of the entire cycle cost based on the demand data collected from the second phase. This clearly gives a *biased* estimation of the gradient. Nevertheless, we will show in §3.4.3 that the error of this estimation is very small in expectation and it vanishes at k grows.

3.3.4 Computation of Gradient $\nabla G(S, a, b)$

Let $I_t(S)$ and $q_t(S)$ denote the on-hand inventory and the ordering quantity in period t under the base-stock policy S, respectively. Also let $I'_t(S)$ and $q'_t(S)$ denote their respective gradients with respect to the base-stock level S. Since

$$\nabla G(S, a, b) = \sum_{t=a}^{b} \Big[h \cdot \mathbb{1}\left(I'_t(S) = 1, D_t < I_t(S)\right) - p \cdot \mathbb{1}\left(I'_t(S) = 1, D_t > I_t(S)\right) \Big],$$

we only need to keep track of $I_t(S)$ and $I'_t(S)$ from period a to b. The inventory level $I_t(S)$ is easy to compute. For $I'_t(S)$, it follows from Theorem 1 in Huh et al. (2009a) that

$$q'_t(S) = I'_{t-1}(S) \cdot \mathbb{1}\left[D_{t-1} > I_{t-1}(S)\right], \tag{3.6}$$

$$I'_t(S) = 1 - \textstyle\sum_{i=t-L+1}^{t} q'_i(S), \tag{3.7}$$

Thus, $I'_t(S)$ and $q'_t(S)$ can be computed recursively if we can evaluate $\mathbb{1}\left[D_{t-1} > I_{t-1}(S)\right]$ and have the necessary boundary conditions.

In the SCU algorithm, we need to compute the gradient $\nabla G(S_k, \tau'_k, \tau_{k+1} - 1)$ of the G-system (*not the SCU-system*). Note that $\nabla G(S_k, \tau'_k, \tau_{k+1}-1)$ represents the partial derivative with respect to S_k assuming τ'_k and τ_{k+1} are fixed. The boundary conditions for the G-system are $(q^G_t)'(S) = 0$ for $t < \tau'_k$ and $(q^G_{\tau'_k})'(S) = 1$. To evaluate $\mathbb{1}\left[D_{t-1} > I^G_{t-1}(S_k)\right]$ for $\tau'_k \le t < \tau_{k+1} - 1$, we need the demand information D_{t-1} in relation to $I^G_{t-1}(S_k)$. Since the only available demand data is from the SCU-system, the comparison between D_{t-1} and $I^G_{t-1}(S_k)$ is possible only when $I^{SCU}_{t-1} \ge I^G_{t-1}$. This is true, according to our algorithm design and Lemma 3.7, for the second phase of each cycle. Thus, the gradient $\nabla G(S_k, \tau'_k, \tau_{k+1} - 1)$ can be readily computed.

3.4 Performance Analysis and Discussions

We first formally define regret. Given a sample path $\omega = \{d_1, d_2, \dots, \}$ of the demand process, the T-period regret of the SCU algorithm is defined as the difference between the clairvoyant optimal cost (under full information) and the cost incurred by SCU over T periods. More specifically,

$$\mathcal{R}^{\mathbf{SCU}}_T(\omega) = \sum_{t=1}^{T} \left(C^{SCU}_t(\omega) - C^{S^*}_t(\omega)\right),$$

where $C^{SCU}_t(\omega)$ is the cost incurred in period t by our nonparametric (closed-loop) SCU algorithm, and $C^{S^*}_t(\omega)$ is the cost incurred in period t by the system operated under the (clairvoyant) optimal base-stock level S^*. The average regret of SCU algorithm is $\mathbb{E}\left[\mathcal{R}^{\mathbf{SCU}}_T\right]$, and the average regret per period is defined as $\mathbb{E}\left[\mathcal{R}^{\mathbf{SCU}}_T\right]/T$.

3.4.1 Main Result

We formally state the main theoretical result of this chapter below.

Theorem 3.3. *Suppose Assumption 3.1 holds. For each problem instance of the lost-sales inventory system with positive lead times and censored demand, the expected regret of the SCU algorithm is upper bounded by $O(\sqrt{T})$. That is, there exists some positive constant K, such that the expected regret of SCU algorithm satisfies*

$$\mathbb{E}\left[\mathcal{R}^{\mathbf{SCU}}_T\right] \le K\sqrt{T}, \qquad \textit{for all} \quad T \ge 1.$$

In other words, the average regret per period approaches 0 at the rate of $O(1/\sqrt{T})$.

We also show that the regret is tight, which is formally stated below.

Proposition 3.4. *Suppose $T > 2$. Even with uncensored demand, there exist problem instances such that the expected regret for any learning algorithm is lower bounded by $\Omega(\sqrt{T})$.*

The problem instance with continuous demand constructed for Proposition 3.4 is very similar to the discrete demand example constructed by Besbes and Muharremoglu (2013). Following their arguments, we provide the proof of Proposition 3.4.

Proof. We provide an example with continuous demand and show that its regret under any learning policy is lower bounded by $\Omega(\sqrt{T})$. This example is a slight modification of the discrete demand example provided in Besbes and Muharremoglu (2013), and the lower bound proof also follows their argument.

Example 3.5. Consider an inventory control problem with lost sales and $h = p = 1$, $L = 0$ and $T > 2$. The demand follows one of two potential distributions, with the cdf F_a and F_b given by

$$F_a(x) = \begin{cases} (50 - \frac{100}{\sqrt{T}})x & \text{for } 0 \le x < 0.01, \\ \frac{1}{2} - \frac{1}{\sqrt{T}} & \text{for } 0.01 \le x < 1, \\ (50 + \frac{100}{\sqrt{T}})(x-1) + \frac{1}{2} - \frac{1}{\sqrt{T}} & \text{for } 1 \le x < 1.01, \\ 1 & \text{for } x \ge 1.01, \end{cases}$$

and

$$F_b(x) = \begin{cases} (50 + \frac{100}{\sqrt{T}})x & \text{for } 0 \le x < 0.01, \\ \frac{1}{2} + \frac{1}{\sqrt{T}} & \text{for } 0.01 \le x < 1, \\ (50 - \frac{100}{\sqrt{T}})(x-1) + \frac{1}{2} + \frac{1}{\sqrt{T}} & \text{for } 1 \le x < 1.01, \\ 1 & \text{for } x \ge 1.01 \end{cases}$$

Then, the optimal base-stock level for F_a, denoted by S_a^*, is within $(0, 0.01)$, and the optimal base-stock level for F_b , denoted by S_b^*, is within $(1, 1.01)$. We prove that, even with observable demand, no policy can achieve a worst-case expected regret better than $\Omega(\sqrt{T})$.

Let π be an arbitrary policy. The worst-case expected regret of π is bounded from below by

$$(h+p)\frac{0.98}{2\sqrt{T}} \max\left\{ \sum_{t=1}^{T} \mathbb{P}_a^\pi \left(S_t^\pi(\omega) > \frac{1}{2} \right), \ \sum_{t=1}^{T} \mathbb{P}_b^\pi \left(S_t^\pi(\omega) \le \frac{1}{2} \right) \right\},$$

which can be further bounded from below by

$$(h+p)\frac{0.98}{4\sqrt{T}}\sum_{t=1}^{T}\max\left\{\mathbb{P}_a^\pi\left(S_t^\pi(\omega)>\frac{1}{2}\right),\ \mathbb{P}_b^\pi\left(S_t^\pi(\omega)\le\frac{1}{2}\right)\right\}. \tag{3.8}$$

By Theorem 2.2 in Tsybakov (2009), we have

$$\max\left\{\mathbb{P}_a^\pi\left(S_t^\pi(\omega)>\frac{1}{2}\right),\mathbb{P}_b^\pi\left(S_t^\pi(\omega)\le\frac{1}{2}\right)\right\}\ge\frac{1}{4}\cdot\exp\{-\mathcal{K}_{t-1}(\mathbb{P}_a,\mathbb{P}_b)\}, \tag{3.9}$$

where

$$\mathcal{K}_t(\mathbb{P}_a,\mathbb{P}_b)=\mathbf{E}_a\left[\log\frac{\mathbf{P}_a(D_1,\ldots,D_t)}{\mathbf{P}_b(D_1,\ldots,D_t)}\right]$$

is the Kullback-Leibler divergence (see Kullback and Leibler (1951)) between the distributions of $\{D_1,\ldots,D_t\}$ under F_a and under F_b, which is equal to

$$\mathcal{K}_t(\mathbb{P}_a,\mathbb{P}_b)=t\left[\left(\frac{1}{2}+\frac{1}{\sqrt{T}}\right)\log\left(\frac{1+\frac{2}{\sqrt{T}}}{1-\frac{2}{\sqrt{T}}}\right)+\left(\frac{1}{2}-\frac{1}{\sqrt{T}}\right)\log\left(\frac{1-\frac{2}{\sqrt{T}}}{1+\frac{2}{\sqrt{T}}}\right)\right].$$

It is a simple exercise to show that $2x\le\log\frac{1+x}{1-x}\le 2x+2x^2$ for $x\in(0,1/2)$. Substituting the inequality to the equation above we obtain $\mathcal{K}_t(\mathbb{P}_a,\mathbb{P}_b)\le\frac{7t}{T}$. Plugging this into (3.9) yields

$$\max\left\{\mathbb{P}_a^\pi\left(S_t^\pi(\omega)>\frac{1}{2}\right),\ \mathbb{P}_b^\pi\left(S_t^\pi(\omega)\le\frac{1}{2}\right)\right\}\ge\frac{1}{4}\exp\left\{-\frac{7(t-1)}{T}\right\}\ge\frac{1}{4}e^{-7}.$$

Consequently, (3.8) is bounded from below by

$$(h+p)\frac{0.98}{4\sqrt{T}}\sum_{t=1}^{T}\frac{1}{4}e^{-7}=\frac{2\cdot 0.98}{16}e^{-7}\sqrt{T}.$$

This completes the proof of Proposition 3.4. □

3.4.2 Building Blocks for Regret Analysis

To prove our main result (i.e.,Theorem 3.3), we first need to establish several important building blocks for the regret analysis, which are presented below.

The first result ensures that the cycles used in designing the SCU algorithm are well defined: By maintaining no less on-hand inventory in SCU-System than in the $\underline{S}$-system, the system dynamics of the $\underline{S}$-system can always be correctly simulated.

Lemma 3.6. *The SCU-system always has no less on-hand inventory than the simulated $\underline{S}$-system.*

Proof. It suffices to prove that for every sample path and in every period, after dropping the withheld on-hand inventory, every entry of the inventory vector of the SCU-system is no lower than that of the simulated $\underline{S}$-system.

From Theorem 1 in Huh et al. (2009a), we know that the inventory vector of a system operating under a base-stock level $S \geq \underline{S}$ is always no lower than that of the $\underline{S}$-system in all the entries. During the first cycle, since the SCU-system is the same as the base-stock system with $S_1 \geq \underline{S}$, the result clearly holds for the first cycle.

We prove the result for other periods using induction. Suppose the claim holds true from the first cycle to the $(k-1)$-th cycle for some $k \geq 2$, which is from period 1 to period $\tau_k - 1$. Then, we want to prove that the result is also true from period 1 to period $\tau_{k+1} - 1$.

Since the $\underline{S}$-system has no lost-sales from period $\tau_k - L$ to period $\tau_k - 1$, and the SCU-system has more on-hand inventory than the $\underline{S}$-system in these periods, then the pipeline inventory of the SCU-system at τ_k must be of the form $[\cdot, d_{\tau_k-2}, \ldots, d_{\tau_k-L}]$, where the first entry (which is the order quantity in period τ_k) remains to be specified. There are two possible cases: 1) $I^{SCU}_{\tau_k} \geq I^{S_k}_{\tau_k} = S_k - \sum_{i=\tau_k-L}^{\tau_k-1} d_i$ and 2) $I^{SCU}_{\tau_k} < I^{S_k}_{\tau_k}$.

In Case 1), the SCU algorithm marks $I_{\tau_k} - S_k + \sum_{i=\tau_k-L}^{\tau_k-1} d_i$ amount of the on-hand inventory as withheld and orders d_{τ_k-1} in period τ_k. By the SCU algorithm, the regular (or non-withheld) inventory vector during this cycle in the SCU-system is the same as that of the base-stock system with the base-stock level S_k, which we shall refer to as the S_k-system. Now, comparing the S_k-system and the $\underline{S}$-system at the beginning of period τ_k, the only difference lies in their on-hand inventory levels, namely, $S_k - \sum_{i=\tau_k-L}^{\tau_k-1} d_\ell$ and $\underline{S} - \sum_{i=\tau_k-L}^{\tau_k-1} d_\ell$, which are achieved in period τ_k when both systems started out empty in period 1 and followed their own base-stock policies. By the monotonicity result in Theorem 1 of Huh et al. (2009a), this implies that the inventory vector for the S_k-system is no lower than that of the $\underline{S}$-system between period τ_k and period $\tau_{k+1} - 1$. Hence, the result follows from the fact that the inventory vector of the SCU-system is no lower than that of the S_k-system, and that the inventory vector of the S_k-system is no lower than that of the $\underline{S}$-system during the k-th cycle.

In Case 2), the SCU algorithm orders

$$d_{\tau_k-1} + S_k - \sum_{i=\tau_k-L}^{\tau_k-1} d_i - I_{\tau_k} = S_k - \sum_{i=\tau_k-L}^{\tau_k-2} d_i - I_{\tau_k}$$

in period τ_k to bring the inventory position up to S_k. Now consider another system that starts in period τ_k with the same inventory vector as the SCU-system, but implements a base-stock policy S_{k-1} between period τ_k and period $\tau_{k+1}-1$. With a slight abuse of notation, call the latter system the S_{k-1}-system. Then, it can be seen that the inventory vector of the SCU-system between period τ_k and period $\tau_{k+1}-1$ is always no lower than that of the S_{k-1}-system. On the other hand, by applying Theorem 1 in Huh et al. (2009a) to the S_{k-1}-system and similar arguments as in Case 1) above, we can show that the inventory vector in the S_{k-1}-system is no lower than that of the $\underline{S}$-system between period τ_k and period $\tau_{k+1}-1$. This proves that the inventory vector of the SCU-system is no lower than that of the $\underline{S}$-system during cycle k.

Thus, the result holds for both cases during the k-th cycle. This completes the induction argument and the proof of Lemma 3.6. □

The next result ensures that the gradient of the G-system, which is used in the SCU algorithm, can indeed be computed using (censored) demand data collected from the SCU-system. Recall that only the gradient in the second phase of each cycle $k \geq 2$ is computed and used in the SCU algorithm.

Lemma 3.7. *For the SCU algorithm, in each period of the second phase of any cycle $k \geq 2$, the SCU-system has no less on-hand inventory than the G-system.*

Proof. We consider the following two cases: 1) $I_{\tau_k}^{SCU} \geq I_{\tau_k}^{S_k}$, and 2) $I_{\tau_k}^{SCU} < I_{\tau_k}^{S_k}$, separately.

First suppose that $I_{\tau_k}^{SCU} \geq I_{\tau_k}^{S_k}$. In this case, it follows from $\hat{I}_{\tau_k}^{SCU} := \left(\hat{I}_{\tau_k}^{SCU} - (S_k - S_{k-1})\right)^+$ that there are two sources of excess withheld on-hand inventory at the beginning of cycle k in the SCU-system. The first is inherited from the previous cycle, and the second is the newly created ones due to a decrease in the target base-stock level. Since the pipeline inventories at the beginning of cycle k in both SCU- and G-systems are equal (as both experienced no stockouts for L consecutive periods), the SCU-system and the G-system would be the same in period τ_k if the withheld on-hand inventory is ignored. Furthermore, since the withheld on-hand inventory is not counted when making ordering decisions in SCU-system, the ordering quantities of the two systems would be the same within this cycle, and the SCU system will always have no lower on-hand inventory than the G-system for each period in both the first and second phase of cycle k.

Next suppose that $I_{\tau_k}^{SCU} < I_{\tau_k}^{S_k}$. In this case, the sales data collected from SCU-system may not allow us to simulate the G-system as it has lower on-hand inventory in period τ_k (and maybe also in some subsequent periods). Since both SCU- and G-systems implement the base-stock level S_k, it follows from Lemma 3.6 above and Huh et al. (2009a) that both

systems would have experienced no stockouts for L consecutive periods before the triggering period τ'_k. This implies that, the inventory state of SCU- and G- systems, after placing order, will be both equal to $\left(d_{\tau_k-1}, \ldots, d_{\tau'_k-L}, S_k - \sum_{t=\tau'_k-L}^{\tau'_k-1} d_t\right)$. This show that the SCU- and G-systems would be identical during the second phase of cycle k, and in particular, they will have the same on-hand inventory level in each period of the second phase of cycle k.

Combining the two cases, we complete the proof of Lemma 3.7. □

The following two lemmas delineate the relationships between demand characteristics and lost-sales events in the lost-sales inventory system, and they will play important roles in the proof of our main result. They also explain why Assumption 3.1 is needed for the main result to hold.

Lemma 3.8. *For the simulated $\underline{S}$-system, if $d_k \le \frac{\underline{S}}{L+1}$ for consecutive $2L$ periods $k = t$ to $t + 2L - 1$, then there is no lost-sales in the simulated $\underline{S}$-system from period $t + L$ to period $t + 2L - 1$.*

Proof. Denote the lost-sales from period a to period b in $\underline{S}$ system as $m^{\underline{S}}[a, b]$, then under the stated condition, we have, for any $k = t + L, \ldots, t + 2L - 1$,

$$I_k^{\underline{S}} = \underline{S} - d_{[k-L,k-1]} + m^{\underline{S}}[a, b] \ge \underline{S} - d_{[k-L,k-1]} \ge \frac{\underline{S}}{L+1} \ge d_k.$$

This implies that there will be no lost-sales from period $t + L$ to $t + 2L - 1$. □

Lemma 3.9. *For the SCU-system, if $d_k > \frac{\bar{S}}{L+1}$ for $k = t, \ldots, t + L$, then there is at least one lost-sales period from period t to period $t + L$.*

Proof. First note that by the design of the SCU algorithm, the inventory position of the SCU-system is never more than $\bar{S}$.

We prove by contradiction. Suppose the opposite it true, i.e., $d_k > \frac{\bar{S}}{L+1}$ for $k = t$ to $t+L$, but there is no lost sales from period t to period $t + L$. Because the inventory position at the beginning of period t is at most $\bar{S}$ and all will have arrived by period $t+L$, the on-hand inventory level at the beginning of period $t + L$ is no more than $\bar{S} - \frac{\bar{S}L}{L+1} = \frac{\bar{S}}{L+1}$. Because $d_{t+L} > \frac{\bar{S}}{L+1}$, this would imply that a lost-sale event occurs in period $t + L$, which leads to a contradiction. This proves Lemma 3.9. □

Following Lemmas 3.8 and 3.9, we define, for any period t, two random variables:

$$\underline{t} = \min_k \left\{ k \ge t + 2L - 1 : \max_{k-2L+1 \le i \le k} d_i \le \frac{\underline{S}}{L+1} \right\}, \tag{3.10}$$

$$\bar{t} = \min_k \left\{ k \ge t + L - 1 : \min_{k-L+1 \le i \le k} d_i \ge \frac{\bar{S}}{L+1} \right\}. \tag{3.11}$$

By Assumption 3.1, both $\underline{t}$ and $\bar{t}$ are well-defined. In fact, $\underline{t} - t$ and $\bar{t} - t$ are known as geometric random variables of orders $2L$ and L respectively (see Philippou et al. (1983)). They represent the number of periods it takes after t such that demand is no more than $\underline{S}/(L+1)$ for $2L$ consecutive periods for the first time, and no less than $\bar{S}/(L+1)$ for L consecutive periods for the first time, respectively. By Lemma 3.8, between t and $\underline{t}$, there must exist L consecutive periods such that the $\underline{S}$-system has no lost-sales. Similarly, by Lemma 3.9, between t and $\bar{t}$, the SCU-system must have at least one lost-sales period.

The following lemma discusses the impact of perturbing the initial inventory vector in an inventory system that implements a base-stock policy, and it will be used in comparing the SCU- and G- systems during a cycle. It states that the perturbation does not amplify during the cycle.

Lemma 3.10. *Fix a sample path of demand process and consider two systems, referred to as the original system and β-system respectively, both operating the same base-stock policy S, but their states at the beginning of the first period are $[q_1, q_0, \ldots, q_{2-L}, I_1]$ and $[q_1 + \beta, q_0, \ldots, q_{2-L}, I_1 - \beta]$, with $0 \le \beta \le I_1$. Then, we have $\left|I_t^o - I_t^\beta\right| \le \beta$ for all $t \ge 1$, where I_t^o and I_t^β are the on-hand inventory levels of the original system and the β-system, respectively.*

Proof. Since

$$|I_t^o - I_t^\beta| = (I_t^o - I_t^\beta)^+ + (I_t^\beta - I_t^o)^+,$$

and at most one term on the right hand side can be positive, it suffices to prove, for all $t \ge 1$,

$$(I_t^o - I_t^\beta)^+ \le \beta, \quad (I_t^\beta - I_t^o)^+ \le \beta.$$

In the following, we prove, by induction, that much stronger results hold: For all $t \ge 1$,

$$(I_t^o - I_t^\beta)^+ + \sum_{i=0}^{L-1} (q_{t-i}^o - q_{t-i}^\beta)^+ \le \beta, \tag{3.12}$$

$$(I_t^\beta - I_t^o)^+ + \sum_{i=0}^{L-1} (q_{t-i}^\beta - q_{t-i}^o)^+ \le \beta. \tag{3.13}$$

By our definition of the original and β-system, (3.12) and (3.13) are clearly satisfied when $t = 1$. Suppose (3.12) and (3.13) hold at t, we will show that (3.12) and (3.13) continue to hold at $t + 1$.

We first focus on (3.12). By the system dynamics of base-stock policy S, we have

$$
\begin{aligned}
&(I^o_{t+1} - I^\beta_{t+1})^+ + \sum_{i=0}^{L-1} (q^o_{t+1-i} - q^\beta_{t+1-i})^+ \qquad (3.14)\\
&= \big((I^o_t - d_t)^+ - (I^\beta_t - d_t)^+ + q^o_{t-L+1} - q^\beta_{t-L+1}\big)^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+ + \big(\min(I^o_t, d_t) - \min(I^\beta_t, d_t)\big)^+.
\end{aligned}
$$

We prove (3.12) holds by considering four cases separately: 1) $d_t \geq \max(I^o_t, I^\beta_t)$, 2) $d_t \leq \min(I^o_t, I^\beta_t)$, 3) $I^o_t \leq d_t \leq I^\beta_t$, and 4) $I^\beta_t \leq d_t \leq I^o_t$.

Case 1): By (3.14), the left hand side of (3.12) at $t+1$ is

$$
\begin{aligned}
&\big((I^o_t - d_t)^+ - (I^\beta_t - d_t)^+ + q^o_{t-L+1} - q^\beta_{t-L+1}\big)^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+ + \Big[\min(I^o_t, d_t) - \min(I^\beta_t, d_t)\Big]^+\\
=\ & (q^o_{t-L+1} - q^\beta_{t-L+1})^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+ + (I^o_t - I^\beta_t)^+\\
=\ & (I^o_t - I^\beta_t)^+ + \sum_{i=0}^{L-1} (q^o_{t-i} - q^\beta_{t-i})^+\\
\leq\ & \beta,
\end{aligned}
$$

where the inequality follows from the inductive assumption.

Case 2): In this case, we have, by (3.14),

$$
\begin{aligned}
&\big((I^o_t - d_t)^+ - (I^\beta_t - d_t)^+ + q^o_{t-L+1} - q^\beta_{t-L+1}\big)^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+ + \big(\min(I^o_t, d_t) - \min(I^\beta_t, d_t)\big)^+\\
=\ & \big(I^o_t - I^\beta_t + q^o_{t-L+1} - q^\beta_{t-L+1}\big)^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+\\
\leq\ & (I^o_t - I^\beta_t)^+ + (q^o_{t-L+1} - q^\beta_{t-L+1})^+ + \sum_{i=0}^{L-2} (q^o_{t-i} - q^\beta_{t-i})^+\\
=\ & (I^o_t - I^\beta_t)^+ + \sum_{i=0}^{L-1} (q^o_{t-i} - q^\beta_{t-i})^+\\
\leq\ & \beta,
\end{aligned}
$$

where the first inequality follows from $(a+b)^+ \leq a^+ + b^+$ for any real numbers a and b, and the second inequality follows from the inductive assumption.

Case 3): This case can happen only when $I_t^o \le I_t^\beta$. We have

$$\begin{aligned}
& \big((I_t^o - d_t)^+ - (I_t^\beta - d_t)^+ + q_{t-L+1}^o - q_{t-L+1}^\beta\big)^+ + \sum_{i=0}^{L-2}(q_{t-i}^o - q_{t-i}^\beta)^+ + \big(\min(I_t^o, d_t) - \min(I_t^\beta, d_t)\big)^+ \\
= \;& \big(q_{t-L+1}^o - q_{t-L+1}^\beta - (I_t^\beta - d_t)^+\big)^+ + \sum_{i=0}^{L-2}(q_{t-i}^o - q_{t-i}^\beta)^+ \\
\le \;& \sum_{i=0}^{L-1}(q_{t-i}^o - q_{t-i}^\beta)^+ \\
\le \;& \beta,
\end{aligned}$$

where the equality follows from $\min(I_t^o, d_t) - \min(I_t^\beta, d_t) \le 0$ (and hence the last term is 0), the first inequality follows from $(a-b)^+ \le a^+$ for any real numbers a and $b \ge 0$, and the second inequality follows from the inductive assumption.

Case 4): This last case can occur when $I_t^o \ge I_t^\beta$, and we have

$$\begin{aligned}
& \big((I_t^o - d_t)^+ - (I_t^\beta - d_t)^+ + q_{t-L+1}^o - q_{t-L+1}^\beta\big)^+ + \sum_{i=0}^{L-2}(q_{t-i}^o - q_{t-i}^\beta)^+ + \big(\min(I_t^o, d_t) - \min(I_t^\beta, d_t)\big)^+ \\
= \;& \big((I_t^o - d_t) + q_{t-L+1}^o - q_{t-L+1}^\beta\big)^+ + \sum_{i=0}^{L-2}(q_{t-i}^o - q_{t-i}^\beta)^+ + (d_t - I_t^\beta) \\
\le \;& (I_t^o - d_t) + (q_{t-L+1}^o - q_{t-L+1}^\beta)^+ + \sum_{i=0}^{L-2}(q_{t-i}^o - q_{t-i}^\beta)^+ + (d_t - I_t^\beta) \\
= \;& (I_t^o - I_t^\beta)^+ + \sum_{i=0}^{L-1}(q_{t-i}^o - q_{t-i}^\beta)^+ \\
\le \;& \beta,
\end{aligned}$$

where again we used $(a+b)^+ \le a^+ + b^+$ in the first inequality, and the second inequality is by the inductive assumption.

Hence (3.12) is satisfied for $t+1$ as well. Similar argument proves (3.13) for $t+1$. This finishes the induction proof and the proof of Lemma 3.10. □

3.4.3 Proof of the Main Result

In what follows, we prove Theorem 3.3 based on Lemmas 3.6–3.10 established above. The proof makes use of three bridging systems, and we will show that the cost difference between any two adjacent systems is on the order $O(\sqrt{T})$.

The three bridging systems are the $\underline{S}$-system, the G-system, and the $\overline{\text{SCU}}$-system which

is defined as the SCU-system that ignores all the withheld on-hand inventory. Note that the $\overline{\text{SCU}}$-system can also be considered as the SCU-system that does not incur any holding cost for the withheld on-hand inventory. Figure 3.3 shows the roadmap of our proof.

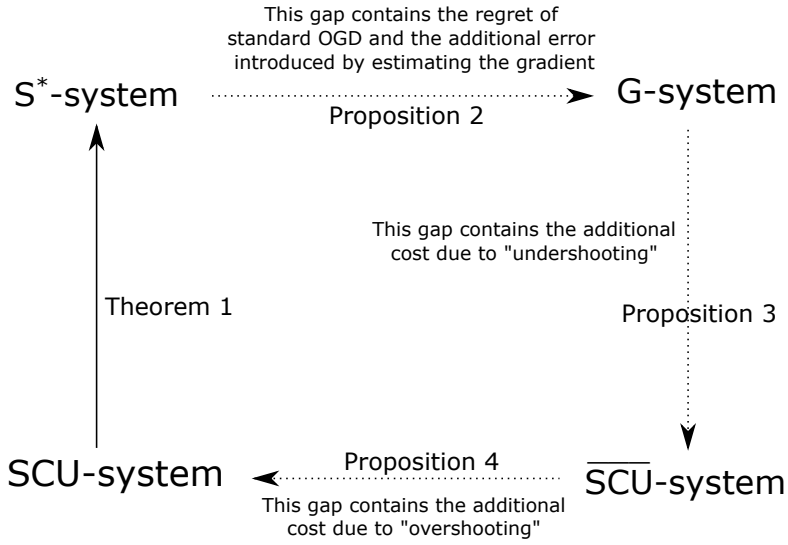

Figure 3.3: A roadmap for the proof of Theorem 3.3

For a fixed T, we let N denote the number of cycles (including the last possibly incomplete one). With a slight abuse of notation, we let $\tau_{N+1} = T + 1$ so the last cycle ends at T. Note that N is a random variable that depends on the demand process.

In the following, we first show that the difference between the expected costs of S^*-system and G-system is upper bounded by $O(\sqrt{T})$.

Proposition 3.11. *There exists some positive constant K_1, such that*

$$\mathbf{E}\left[\sum_{k=1}^{N} G(S_k, \tau_k, \tau_{k+1} - 1)\right] - \mathbf{E}\left[\sum_{t=1}^{T} C_t^{S^*}\right] \leq K_1\sqrt{T}.$$

Proof. For the first cycle, we have

$$\begin{aligned}
\mathbf{E}\left[G(S_1, \tau_1, \tau_2 - 1)\right] - \mathbf{E}\left[\sum_{t=\tau_1}^{\tau_2-1} C_t^{S^*}\right] &\leq \mathbf{E}\left[\tau_2 - \tau_1\right] \cdot \max(h, p)\left(\bar{S} - \underline{S}\right) \\
&\leq \mathbf{E}\left[\underline{t} - \tau_1\right] \cdot \max(h, p)\left(\bar{S} - \underline{S}\right) \\
&= \frac{1 - c_1^{2L}}{(1 - c_1)c_1^{2L}} \cdot \max(h, p)\left(\bar{S} - \underline{S}\right),
\end{aligned}$$

where the last equality follows from Proposition 2.1 of Philippou et al. (1983). Similarly, for

the last cycle, we have

$$\mathbf{E}\left[G(S_N,\tau_N,T)\right]-\mathbf{E}\left[\sum_{t=\tau_N}^{T}C_t^{S^*}\right]\leq 2\cdot\frac{1-c_1^{2L}}{(1-c_1)c_1^{2L}}\cdot\max(h,p)\left(\bar{S}-\underline{S}\right).$$

Every cycle $1<k<N$ contains two phases. The first phase is from period τ_k to period $\tau_k'-1$, and the second phase is from period τ_k' to period $\tau_{k+1}-1$. Recall that the cost gradient for the first phase cannot be evaluated due to lack of demand information. To complete the proof of Proposition 3.11, we first claim Equation (3.15) is correct, which shows that, the gradient for the first phase approaches that of the second phase in expectation, which can be evaluated, when k increases.

$$\left|\mathbf{E}\left[\nabla G(S_k,\tau_k,\tau_k'-1)\right]-\mathbf{E}\left[\nabla G(S_k,\tau_k',\tau_{k+1}-1)\right]\right|=o\left(1/\sqrt{k}\right). \tag{3.15}$$

To prove this equation, first recall that τ_k, τ_k' and τ_{k+1} are three adjacent triggering periods defined by the $\underline{S}$-system. Thus, $\nabla G(S_k,\tau_k,\tau_k'-1)$ and $\nabla G(S_k,\tau_k',\tau_{k+1}-1)$ are determined by $[d_{\tau_k-1},d_{\tau_k-2},\ldots,d_{\tau_k-L}]$ and $[d_{\tau_k'-1},d_{\tau_k'-2},\ldots,d_{\tau_k'-L}]$, respectively. If we re-index $\{\tau_1,\tau_2,\tau_2',\tau_3,\tau_3',\ldots\}$ as $\{r(1),r(2),r(3),r(4),r(5),\ldots\}$, then the process $\{\mathbf{d}_i=[d_{r(i)-1},d_{r(i)-2},\ldots,d_{r(i)-L}];i\geq 1\}$ is a Markov chain on a general state space (or a Harris chain). It is important to keep in mind that this Markov chain is solely determined by the $\underline{S}$-system, and it is not affected by the SCU- or the G-system.

We show that, under Assumption 3.1, $\{\mathbf{d}_i;i\geq 1\}$ is ergodic and converges to a stationary distribution $\mathbf{d}_\infty$ exponentially fast. Following the approach in Huh et al. (2009a), we use uniform ergodicity to prove this result. A measurable set $\mathbf{U}\subseteq\mathbb{R}_+^L$ is called a *small set* with respect to a nontrivial measure ν, if there exists an $i^*>0$ such that for any $\mathbf{d}\in\mathbf{U}$ and any measurable set $B=(B_1,\ldots,B_L)\subseteq\mathbb{R}_+^L$, it holds that $\mathbf{P}(\mathbf{d}_{i^*}\in B\mid\mathbf{d}_1=\mathbf{d})\geq\nu(B)$. By Theorem 16.0.2 of Meyn and Tweedie (1993), if $\mathbf{U}$ is a small set with respect to ν, then there exists a stationary random variable $\mathbf{d}_\infty$ such that for any $\mathbf{d}\in\mathbf{U}$ and $i\geq i^*$, it satisfies $\delta_{i+1}(\mathbf{d})\leq(1-\nu(\mathbb{R}_+^L))^{\frac{i}{i^*-1}}$, where

$$\delta_i(\mathbf{d})=\sup_B\left\{\left|\mathbf{P}(\mathbf{d}_i\in B\mid\mathbf{d}_1=\mathbf{d})-\mathbf{P}(\mathbf{d}_\infty\in B)\right|:\text{ measurable set }B\subseteq\mathbb{R}_+^L\right\}.$$

By the Scheffe's theorem, we have

$$\delta_i(\mathbf{d})=\frac{1}{2}\int_{\mathbf{z}}\left|\left(P(\mathbf{d}_i\in d\mathbf{z}|\mathbf{d}_1=\mathbf{d})-P(\mathbf{d}_\infty\in d\mathbf{z})\right)\right|. \tag{3.16}$$

The first step is to define $\mathbf{U}$, B, ν and i^* for our Markov chain. Since $\mathbf{d}$ is the pipeline inventory of the $\underline{S}$-system at the beginning of a triggering period, we must have $\mathbf{d} \cdot \mathbf{1}^L \leq \underline{S}$, where $\mathbf{d} \cdot \mathbf{1}^L$ is the sum of all entries of $\mathbf{d}$. Let $\mathbf{U} = \{\mathbf{d} \in \mathbb{R}_+^L \mid \mathbf{d} \cdot \mathbf{1}^L \leq \underline{S}\}$, and B_k be any measurable set in $\mathbb{R}_+$ for $k = 1, \ldots, L$, and denote $B = B_1 \times \cdots \times B_L$. Define

$$\nu(B) = \left(\mathbf{P}\left(D \in \left(\cap_{k=1}^L B_k \right) \cap \left[0, \frac{\underline{S}}{L+1} \right] \right) \right)^{2L},$$

where D represents a generic demand. We now prove that $\mathbf{U}$ is a small set with respect to ν and $i^* = 2$, i.e., for any $\mathbf{d} \in \mathbf{U}$ and $B \in \mathbb{R}_+^L$, we have $\mathbf{P}\big(\mathbf{d}_3 \in B \mid \mathbf{d}_1 = \mathbf{d}\big) \geq \nu(B)$.

Consider the event that the demands in periods $1, 2, \ldots, 2L$ satisfy

$$E = \left\{ D_k \in \left(\cap_{k=1}^L B_k \right) \cap \left[0, \frac{\underline{S}}{L+1} \right], \text{ for } k = 1, 2, \ldots, 2L \right\}.$$

By Lemma 3.8, for any initial state $\mathbf{d}_1 = \mathbf{d}$, there is no lost sales in the $\underline{S}$-system from periods L to $2L$, implying $r(2) \leq 2L$. Moreover, on the event E, it is seen that the pipeline inventory at the beginning of period $r(2)$ satisfies $\mathbf{d}_2 \in B$. Hence, for any $\mathbf{d} \in B$, we have

$$\mathbf{P}(\mathbf{d}_2 \in B \mid \mathbf{d}_1 = \mathbf{d}) \geq \mathbf{P}(E) = \nu(B).$$

This shows that the Markov chain $\{\mathbf{d}_k;\ k \geq 1\}$ is uniformly ergodic. Applying Theorem 16.0.2 of Meyn and Tweedie (1993), we obtain, for all $i > 2$,

$$\delta_i(\mathbf{d}) \leq (1 - c_1^{2L})^{\frac{i+1}{2}}. \tag{3.17}$$

For notational convenience, we define

$$H(d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L}) = \mathbf{E}[\nabla G(S_k, \tau_k, \tau_k' - 1) \big| d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L}].$$

Then for any $[d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L}]$, $H(d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L})$ is upper bounded by

$$H(d_{\tau_k - 1}, d_{\tau_k - 2}, \ldots, d_{\tau_k - L}) \leq \frac{1 - c_1^L}{(1 - c_1) c_1^L} \max(h, b). \tag{3.18}$$

Therefore, we have

$$\begin{aligned}
\big|\mathbf{E}[H(\mathbf{d}_k) - \mathbf{E}[H(\mathbf{d}_\infty)]\big| &= \Big|\int_{\mathbf{z}} H(\mathbf{z})\big(P(\mathbf{d}_i \in d\mathbf{z}|\mathbf{d}_1 = \mathbf{d}) - P(\mathbf{d}_\infty \in d\mathbf{z})\big)\Big| \\
&\leq \int_{\mathbf{z}} H(\mathbf{z})\big|\big(P(\mathbf{d}_i \in d\mathbf{z}|\mathbf{d}_1 = \mathbf{d}) - P(\mathbf{d}_\infty \in d\mathbf{z})\big)\big| \\
&\leq \frac{1-c_1^L}{(1-c_1)c_1^L}\max(h,b)\int_{\mathbf{z}}\big|\big(P(\mathbf{d}_i \in d\mathbf{z}|\mathbf{d}_1 = \mathbf{d}) - P(\mathbf{d}_\infty \in d\mathbf{z})\big)\big| \\
&= \frac{1-c_1^L}{(1-c_1)c_1^L}\max(h,b)\cdot\delta_k(\mathbf{d}) \\
&\leq \frac{1-c_1^L}{(1-c_1)c_1^L}\max(h,b)(1-c_1^{2L})^{\frac{k+1}{2}} \\
&= o\Big(\frac{1}{\sqrt{k}}\Big),
\end{aligned}$$

where the second inequality follows (3.18), the second equality follows from the Scheffe's theorem (3.16), the last ineqaulity follows from (3.17), and the last equality follows from the fact that ρ^k tends to 0 faster than $1/\sqrt{k}$ for any $\rho \in (0,1)$.

Applying the above result, we obtain

$$\begin{aligned}
&\Big|\mathbf{E}[\nabla G(S_k,\tau_k,\tau_k'-1)] - \mathbf{E}[\nabla G(S_k,\tau_k',\tau_{k+1}-1)]\Big| \\
=\ &\Big|\mathbf{E}\big[\mathbf{E}[\nabla G(S_k,\tau_k,\tau_k'-1)]\big|\mathbf{d}_k)\big] - \mathbf{E}\big[\mathbf{E}[\nabla G(S_k,\tau_k',\tau_{k+1}-1)]\big|\mathbf{d}_\infty)\big]\Big| \\
=\ &\big|\mathbf{E}[H(\mathbf{d}_k)] - \mathbf{E}[H(\mathbf{d}_{k+1})]\big| \\
=\ &o(1/\sqrt{k}).
\end{aligned}$$

This completes the proof of Equation (3.15).

Equation (3.15) shows that, although a biased gradient is used in the SCU algorithm in the search for the best base-stock level, it is close to the true gradient when k is large and

converges at a rate faster than $o(1/\sqrt{k})$. Applying this result, we obtain

$$\begin{aligned}
& \mathbf{E}\left[\sum_{k=1}^{N} G(S_i, \tau_k, \tau_{k+1}-1) - \sum_{t=1}^{T} C_t^{S^*}\right] \\
= \;& \mathbf{E}\left[\sum_{k=1}^{N} \left(G(S_k, \tau_k, \tau_{k+1}-1) - G(S^*, \tau_k, \tau_{k+1}-1)\right)\right] \\
\leq \;& \mathbf{E}\left[\sum_{k=1}^{N} \nabla G(S_k, \tau_k, \tau_{k+1}-1)(S_k - S^*)\right] \\
\leq \;& \mathbf{E}\left[\sum_{k=2}^{N-1} \nabla G(S_k, \tau_k, \tau_{k+1}-1)(S_k - S^*)\right] + 3 \cdot \frac{1-c_1^{2L}}{(1-c_1)c_1^{2L}} \cdot \max(h,p)(\bar{S}-\underline{S}) \\
\leq \;& \mathbf{E}\left[\sum_{k=2}^{N-1} \left(2\nabla G(S_k, \tau_k', \tau_{k+1}-1)(S_i - S^*) + o(1/\sqrt{k})\right)\right] + 3 \cdot \frac{1-c_1^{2L}}{(1-c_1)c_1^{2L}} \cdot \max(h,p)(\bar{S}-\underline{S}) \\
\leq \;& \mathbf{E}\left[\sum_{k=2}^{N-1} \left(\frac{\sqrt{k}}{2\gamma}\left((S_k - S^*)^2 - (S_{k+1} - S^*)^2\right)\right)\right] + \mathbf{E}\left[\sum_{k=2}^{N-1} \frac{2\gamma \nabla G(S_k, \tau_k', \tau_{k+1}-1)^2}{\sqrt{k}}\right] + o\left(\sqrt{T}\right),
\end{aligned} \tag{3.19}$$

where the first inequality follows from the convexity of function $G(S, \tau_k, \tau_{k+1}-1)$ in S, the third inequality is due to (3.15), and the last inequality is because of, by our SCU algorithm,

$$(S_{k+1} - S^*)^2 \leq (S_k - S^*)^2 - \frac{4\gamma}{\sqrt{k}}(S_k - S^*)\nabla G(S_k, \tau_k', \tau_{k+1}-1) + \frac{4\gamma\left(\nabla G(S_k, \tau_k', \tau_{k+1}-1)\right)^2}{k}.$$

We evaluate the first term on the right hand side of (3.19) as follows:

$$\begin{aligned}
& \mathbf{E}\sum_{k=2}^{N-1}\left[\frac{\sqrt{k}}{2\gamma}\left((S_k - S^*)^2 - (S_{k+1} - S^*)^2\right)\right] \\
\leq \;& \frac{1}{\gamma}\mathbf{E}\left[\frac{\sqrt{2}}{2}(S_2 - S^*)^2 - \frac{\sqrt{N-1}}{2}(S_N - S^*)^2\right] + \frac{1}{2\gamma}\mathbf{E}\sum_{k=3}^{N-1}\left[(\sqrt{k} - \sqrt{k-1})(S_k - S^*)^2\right] \\
\leq \;& \frac{\sqrt{2}}{2\gamma}(\bar{S}-\underline{S})^2 + \frac{1}{2\gamma}\mathbf{E}\left[\sum_{k=3}^{T}(\sqrt{k} - \sqrt{k-1})(\bar{S}-\underline{S})^2\right] \\
= \;& \frac{\sqrt{T}}{2\gamma}(\bar{S}-\underline{S})^2.
\end{aligned} \tag{3.20}$$

To evaluate the second term on the right hand side of (3.19), we first focus on the term $\mathbf{E}[(\nabla G(S_k, \tau_k', \tau_{k+1}-1))^2]$. From Lemma 3.8, $\tau_{k+1}-1$ is no larger than $\underline{\tau_k'}$ with probability

1. Therefore, we have

$$\begin{aligned}\mathbf{E}\left[\nabla G(S_k,\tau_k',\tau_{k+1}-1)\right]^2 &\leq \left[\max(h,p)\left(\bar{S}-\underline{S}\right)\right]^2\cdot\mathbf{E}\left[\left(\underline{\tau_k'}-\tau_k'\right)^2\right] \qquad (3.21)\\ &= \left[\max(h,p)\left(\bar{S}-\underline{S}\right)\right]^2\cdot\frac{2+(4L-1)c_1^{2L}-(4L+1)c_1^{2L+1}+c_1^{4L}-c_1^{4L+1}}{c_1^{4L}-c_1^{4L+2}},\end{aligned}$$

where the equality above follows from Proposition 2.1 in Philippou et al. (1983). Thus, we obtain, for some constant K_1,

$$\mathbf{E}\left[\sum_{k=2}^{N}\frac{2\gamma\nabla G(S_k,\tau_k',\tau_{k+1}-1)^2}{\sqrt{k}}\right]\leq\mathbf{E}\left[\sum_{k=1}^{T}\frac{2\gamma\nabla G(S_k,\tau_k',\tau_{k+1}-1)^2}{\sqrt{k}}\right]\leq K_1\cdot\sqrt{T}. \qquad (3.22)$$

Combining (3.19), (3.21) and (3.22), we complete the proof of Proposition 3.11. □

We next compare the G-system with the $\overline{\text{SCU}}$-system. The difference between these two systems lies in the "undershooting" of the $\overline{\text{SCU}}$-system. That is, both systems operate under the same base-stock level, but at the beginning of each cycle, the $\overline{\text{SCU}}$-system potentially has less on-hand inventory and has to order more to keep the same inventory position as G-system. We will show that the cost difference created by "undershooting" the target levels is upper bounded by $O(\sqrt{T})$ in expectation.

Proposition 3.12. *There exists some positive constant K_2, such that*

$$\mathbf{E}\left[\sum_{t=1}^{T}C_t^{\overline{SCU}}\right]-\mathbf{E}\left[\sum_{k=1}^{N}G(S_k,\tau_k,\tau_{k+1}-1)\right]\leq K_2\sqrt{T}.$$

Proof. Let $\tilde{I}_t^{SCU}$ and I_t^G denote the on-hand inventory levels of $\overline{\text{SCU}}$- and G- systems, respectively. Then, for every sample path, we have

$$\begin{aligned}\sum_{t=1}^{T}C_t^{\overline{\text{SCU}}}-\sum_{k=1}^{N}G(S_k,\tau_k,\tau_{k+1}-1) &\leq \sum_{t=1}^{T}\max(h,p)\left|I_t^G-\tilde{I}_t^{SCU}\right|\\ &= \sum_{k=1}^{N}\sum_{t=\tau_i}^{\tau_{k+1}-1}\max(h,p)\left|I_t^G-\tilde{I}_t^{SCU}\right|. \qquad (3.23)\end{aligned}$$

For the first cycle, we have $I_t^G=\tilde{I}_t^{SCU}$ for every period t. For cycle $k\geq 2$, if $S_k<S_{k-1}$, then by the construction of the SCU algorithm, $I_t^G=\tilde{I}_t^{SCU}$ for every period t in cycle k; if $S_k\geq S_{k-1}$, then I_t^G may differ from $\tilde{I}_t^{SCU}$ for t in the first phase of the cycle (i.e., periods from τ_k to $\tau_k'-1$), and they will become the same from τ_k' until $\tau_{k+1}-1$.

Suppose $S_k \geq S_{k-1}$, we will show that

$$\left|I_t^G - \tilde{I}_t^{SCU}\right| \leq I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU} \leq S_k - S_{k-1}, \text{ for periods } t = \tau_k, \ldots, \tau_k' - 1.$$

We first prove the second inequality $I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU} \leq S_k - S_{k-1}$. For the G-system, its inventory vector at period τ_k is $\left[d_{\tau_k-1}, d_{\tau_k-2}, \ldots, d_{\tau_k-L}, S_k - \sum_{i=\tau_k-L}^{\tau_k-1} d_i\right]$. For the SCU-system, if $\hat{I}_{\tau_k-1} = 0$, then the inventory vector at τ_k would be $\left[d_{\tau_k-1} - S_{k-1} + S_k, d_{\tau_k-2}, \ldots, d_{\tau_k-L}, S_{k-1} - \sum_{i=\tau_k-L}^{\tau_k-1} d_i\right]$. In this case we have

$$I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU} = I_{\tau_k}^G - I_{\tau_k}^{SCU} = S_k - S_{k-1}.$$

If $\hat{I}_{\tau_k-1} > 0$, then some of the withheld on-hand inventory will be included back to the regular on-hand inventory by equation $\hat{I}_{\tau_k} = (\hat{I}_{\tau_k} - (S_k - S_{k-1}))^+$, and as a result, the regular on-hand inventory in the SCU-system may be higher and we will have $I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU} < S_k - S_{k-1}$. Hence in all cases $I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU} \leq S_k - S_{k-1}$ is satisfied. Then, we apply Lemma 3.10 to obtain $\left|I_t^G - \tilde{I}_t^{SCU}\right| \leq I_{\tau_k}^G - \tilde{I}_{\tau_k}^{SCU}$ for all $t = \tau_k, \ldots, \tau_k'$. Combining the two scenarios shows $\left|I_t^G - \tilde{I}_t^{SCU}\right| \leq |S_k - S_{k-1}|$ for $t = \tau_k, \ldots, \tau_k' - 1$ and $\left|I_t^G - \tilde{I}_t^{SCU}\right| = 0$ for $t = \tau_k', \ldots, \tau_{k+1} - 1$.

Taking expectation on both sides of (3.23), we obtain

$$\begin{aligned}
&\mathbf{E}\left[\sum_{t=1}^{T} C_t^{\overline{\text{SCU}}} - \sum_{k=1}^{N} G(S_k, \tau_k, \tau_{k+1} - 1)\right] \leq \max(h,p)\mathbf{E}\left[\sum_{k=1}^{N}\sum_{t=\tau_k}^{\tau_k'}\left|I_t^G - \tilde{I}_t^{SCU}\right|\right] \\
&\leq \max(h,p)\mathbf{E}\left[\sum_{k=2}^{N}(\tau_k' - \tau_k + 1)\left|S_k - S_{k-1}\right|\right] \leq \max(h,p)\mathbf{E}\left[\sum_{k=2}^{T}(\tau_k' - \tau_k + 1)\left|S_k - S_{k-1}\right|\right] \\
&= \max(h,p)\mathbf{E}\left[\sum_{k=2}^{T}\left|S_k - S_{k-1}\right|\right]\mathbf{E}\left[(\tau_k' - \tau_k + 1)\right] = \max(h,p)\mathbf{E}\left[\sum_{k=2}^{N}\left|S_k - S_{k-1}\right|\right]\frac{1 - c_1^{2L}}{(1-c_1)c_1^{2L}} \\
&\leq \max(h,p)^2\mathbf{E}\left[\sum_{k=2}^{N}\frac{2\gamma}{\sqrt{k}}(\tau_{k-1}' - \tau_{k-1})\right]\frac{1 - c_1^{2L}}{(1-c_1)c_1^{2L}} \leq \left(\max(h,p)\frac{1 - c_1^{2L}}{(1-c_1)c_1^{2L}}\right)^2\sum_{k=1}^{T}\frac{2\gamma}{\sqrt{k}} \leq K_2 \cdot \sqrt{T}
\end{aligned}$$

for some constant K_2. The first equality above is by the independence of $|S_k - S_{k-1}|$ and $\tau_k - \tau_k$, and the second equality and the inequality after that are by Proposition 2.1 in Philippou et al. (1983). This completes the proof of Proposition 3.12. □

Following the roadmap in Figure 3.3, the last part of the regret analysis is to bound the gap between the SCU-system and the $\overline{\text{SCU}}$-system. The cost difference between these two systems is upper bounded by the total holding cost of the withheld on-hand inventory. The following lemma shows that this part is also bounded by $O(\sqrt{T})$ in expectation.

Proposition 3.13. *There exists some positive constant K_3 such that*

$$\mathbf{E}\left[\sum_{t=1}^{T} C_t^{SCU}\right] - \mathbf{E}\left[\sum_{t=1}^{T} C_t^{\overline{SCU}}\right] \leq K_3\sqrt{T}.$$

Proof. First, we have

$$\mathbf{E}\left[\sum_{t=1}^{T} C_t^{SCU}\right] - \mathbf{E}\left[\sum_{t=1}^{T} C_t^{\overline{SCU}}\right] \leq h\sum_{t=1}^{T}\mathbf{E}\left[\hat{I}_t\right]. \tag{3.24}$$

According to the SCU algorithm, we have $\hat{I}_{t+1} = (\hat{I}_t - (d_t - \tilde{I}_t)^+)^+$ from period to period during the cycle, thus the withheld inventory is gradually consumed by demand; and when going from one cycle to the next, new withheld inventory may be created by $\hat{I}_{\tau_{k+1}} := (\hat{I}_{\tau_{k+1}} - (S_{k+1} - S_k))^+$ at the beginning of cycle $k+1$. This shows that, the only source of new withheld inventory is generated at the beginning of the cycles, and the maximum added at the beginning of cycle $k+1$ is $(S_k - S_{k+1})^+$, $k = 1, 2, \ldots, N$.

When evaluating the RHS of (3.24), instead of evaluating it vertically by finding the total amount of $\hat{I}_t$ in every period and then adding up over periods, we compute it horizontally by identifying the total number of periods in which an withheld inventory unit stays in the system and then adding up over all the withheld inventory units. At the beginning of period τ_{k+1}, a maximum of $(S_k - S_{k+1})^+$ units of the withheld inventory are created in the system, and from $\hat{I}_{t+1} = (\hat{I}_t - (d_t - \tilde{I}_t)^+)^+$, it is seen that within a cycle, the amount of the withheld on-hand inventory is non-increasing, and dropping to 0 when the SCU-system experiences a lost-sale. By Lemma 3.9, we know that there is at least one lost-sale from τ_{k+1} to $\bar{\tau}_{k+1}$ (defined similarly as in (3.11)). Thus the maximum of $(S_k - S_{k+1})^+$ units of the withheld inventory will be consumed by external demand by time $\bar{\tau}_{k+1}$. That is, these units incur a maximum holding cost of $h(S_k - S_{k+1})^+(\bar{\tau}_{k+1} - \tau_{k+1})$. Therefore,

$$h\sum_{t=1}^{T}\hat{I}_t \;\leq h\sum_{k=1}^{N}(S_k - S_{k+1})^+\left(\bar{\tau}_{k+1} - \tau_{k+1}\right) \leq h\sum_{k=1}^{N}\left|S_k - S_{k+1}\right|\left(\bar{\tau}_{k+1} - \tau_{k+1}\right).$$

Taking expectation yields, for some constant K_3,

$$\begin{aligned}
\mathbf{E}\left[\sum_{t=1}^{T} C_t^{SCU}\right] - \mathbf{E}\left[\sum_{t=1}^{T} C_t^{\overline{\text{SCU}}}\right] &\leq h\mathbf{E}\left[\sum_{k=2}^{N} |S_{k-1} - S_k| \left(\bar{\tau}_k - \tau_k\right)\right] \\
&= h\mathbf{E}\left[\sum_{k=2}^{N} |S_{k-1} - S_k|\right] \cdot \mathbf{E}\left[\left(\bar{\tau}_k - \tau_k\right)\right] \\
&\leq h\mathbf{E}\left[\sum_{k=2}^{N} |S_{k-1} - S_k|\right] \cdot \frac{1 - c_2^L}{(1 - c_2)c_2^L} \\
&\leq h \max(h, p) \sum_{k=2}^{N} \frac{2\gamma}{\sqrt{k}} \mathbf{E}\left[\tau_k - \tau_{k-1}\right] \cdot \frac{1 - c_2^L}{(1 - c_2)c_2^L} \\
&\leq \sum_{k=1}^{T} \frac{2\gamma}{\sqrt{k}} \left[h \cdot \max(h, p) \cdot \frac{2(1 - c_1^{2L})}{(1 - c_1)c_1^{2L}} \cdot \frac{1 - c_2^L}{(1 - c_2)c_2^L}\right] \\
&\leq K_3 \cdot \sqrt{T},
\end{aligned}$$

where the first equality holds by independence, and the second and third inequalities are by Proposition 2.1 in Philippou et al. (1983). This completes the proof of Proposition 3.13. □

Combining Propositions 3.11, 3.12 and 3.13, we complete the proof of Theorem 3.3.

3.4.4 The SCU Algorithm for Uncensored Demand

The censored demand assumption in this chapter brings two main challenges in the development of our learning algorithm. The first challenge is to guarantee the correct simulation of the $\underline{S}$-system, which requires our system to always have no less on-hand inventory than the $\underline{S}$-system. To achieve that, we have to dynamically modify the target base-stock levels. This is the reason for introducing the concept of withheld on-hand inventory. When demand is uncensored, this is not necessary, as the $\underline{S}$-system can always be simulated in each and every period. Thus, we can simply apply the target base-stock level S_k for every cycle k. The second challenge is the evaluation of gradient for our learning algorithm. With censored demand, we cannot evaluate the gradient of function G when the SCU-system less on-hand inventory than the G-system. To overcome this issue, we design two phases in each cycle $k \geq 2$ where the gradient of G-system can always be evaluated in the second phase, and that is then used to estimate the gradient for cycle k. (We double the gradient of the second phase to provide a close yet biased estimate for the gradient of the whole cycle.) With uncensored demand, this is again not necessary, as the gradient of G can always be computed.

These observations lead to a much simpler SCU algorithm for the case with uncensored

demand in which neither the withheld inventory nor an additional phase for the learning cycle is needed. Denote this modified SCU algorithm for the uncensored demand case by SCU-UN, and we formally present it below. In this algorithm, the gradient $\nabla G(S_k, \tau_k, \tau_{k+1} - 1)$ for cycle k is computed following the same procedure in §3.3.2 but using uncensored demand data in cycle k.

Algorithm 2: SCU-UN

Step 0 (**Initialization**): Start with an arbitrary target base-stock level $S_1 \in [\underline{S}, \bar{S}]$. Set $\tau_1 = 1$, cycle number $k = 1$. Let step size $\eta_k = \frac{\gamma}{\sqrt{k}}$ for $k = 1, 2, \ldots$, for some positive constant γ. Set the consecutive no lost-sales indicator $\psi := 0$. Set $t = 1$, and the initial inventory of SCU-UN-system and simulated $\underline{S}$-system to $\mathbf{x}_1^{SCU-UN} = \mathbf{x}_1^{\underline{S}} = \mathbf{0}$.

Step 1 In each period t, do the following:

(a) For SCU-UN-system, order $q_t^{SCU-UN} = \left(S_k - \sum \mathbf{x}_t^{SCU-UN}\right)^+$; for the simulated $\underline{S}$-system, order $q_t^{\underline{S}} = (\underline{S} - \sum \mathbf{x}_t^{\underline{S}})^+$.

(b) Observe demand d_t, and update the states of SCU-UN-system and $\underline{S}$-system according to system dynamics (3.1).

(c) If there is no lost-sales in the $\underline{S}$-system, then set $\psi := \psi + 1$. Otherwise set $\psi := 0$.

(d) If $\psi = L$, then label period $t + 1$ as a triggering period. Set $\tau_{k+1} = t + 1$, and $\psi := 0$. Update the target base-stock level for the next cycle as

$$S_{k+1} = \mathbf{P}_{[\underline{S}, \bar{S}]} \left(S_k - \eta_k \nabla G \left(S_k, \tau_k, \tau_{k+1} - 1\right)\right).$$

Set $t := t + 1$ and $k := k + 1$, and repeat Step 1.

This concludes the description of the SCU-UN algorithm.

Although the SCU-UN algorithm is much simpler than the SCU algorithm, the essential idea of (random) cycle-updating rule based on the simulated $\underline{S}$-system remains the same. The performance analysis of SCU-UN is similar and simpler compared to SCU, and under the same Assumption 1, it achieves a regret of rate $O(\sqrt{T})$ for uncensored demand case. We omit the details of performance analysis for the SCU-UN algorithm.

One purpose of introducing the SCU-UN algorithm is to study the value of observing lost-demand information. That is, how much cost savings can be resulted from knowing the lost-demand information? In the next section, we will conduct a numerical study of both SCU and SCU-UN to investigate the performance gap between these two cases.

3.5 Numerical Experiments

We conduct a numerical study on the performance of the SCU algorithm. We first test against the algorithm proposed in Huh et al. (2009a), denoted by HJMR. We then test against the SCU-UN algorithm described in §3.4.4, which will reveal the value of censored demand information. The performances of the algorithms are evaluated by the percentage of total cost increase (over the planning horizon) compared with that of the clairvoyant optimal base-stock policy, i.e.,

$$\kappa = \frac{\mathbf{E}[\mathcal{R}_t]}{\mathbf{E}\left[\sum_{t=1}^{T} C_t^{S^*}\right]} \times 100\%,$$

where S^* is the clairvoyant optimal base-stock level.

Design of experiments. We present the design of our numerical experiments. Three lead times are considered $L \in \{2, 4, 6\}$. For the cost structure, we normalize the holding cost to $h = 1$, and consider four lost sales costs $p \in \{20, 30, 40, 50\}$. For the demand, we consider gamma distribution with mean 10 and different shape parameters $\alpha \in \{2, 3, 4\}$. We let $\underline{S} = 8 \cdot L + 1$ and $\bar{S} = 30 \cdot L + 1$ for SCU, HJMR and SCU-UN. The optimal solution is always contained in $[\underline{S}, \bar{S}]$. We set the step-size parameter $\gamma = 2^{-4}$ in SCU and SCU-UN. We consider three planning horizons $T \in \{100, 200, 500, 1000\}$. The system starts with 0 initial inventory. For each testing instance, we generate 5000 sample paths of the random demand process, and use that to compute the average cost of the learning algorithm.

Numerical results. The performance of SCU, HJMR and SCU-UN under all the testing instances are summarized in Tables 3.1, 3.2 and 3.3. We first compare the performances between SCU and HJMR. It can be seen that SCU generally performs better. Under some instances, HJMR converges faster at the beginning but gets dominated as T grows. We note that SCU algorithm is quite robust when the lead time increases but it seems that HJMR's performance gets worse when the lead time becomes longer.

We then compare the performances between SCU and SCU-UN. As expected, SCU-UN indeed performs better, demonstrating the value of the censored demand information. On average, SCU has 26.5% of cost increase (over the clairvoyant optimal base-stock solution) when $T = 100$, while SCU-UN has 21.1% of cost increase. And when $T = 1000$, SCU has 5.2% of cost increase while SCU-UN has 3.1% of cost increase. This shows that the value of uncensored demand information is quite significant. Thus, if feasible, it pays for the inventory manager to invest the necessary resource to collect such lost-demand information.

For the SCU (and SCU-UN) algorithm, another question apart from its expected average regret is the policy update frequency, which is determined by the number of periods between

L consecutive periods of no lost-sales (without overlapping) of the $\underline{S}$ system. The cycle length only depends on the demand process, and is independent of SCU policy or the cost parameters. In our theoretical analysis, we upper bound the time between triggering periods by a geometric distribution of order $2L$. By Philippou et al. (1983), for our testing instances, the expected value of this upper bound is about 50 for $L = 2$, about 1500 for $L = 4$, and about 44000 for $L = 6$. In our numerical examples, we find that the average number of periods between triggering periods is around 5 period for $L = 2$, around 12 periods when $L = 4$, and around 25 periods when $L = 6$. Hence the actual cycle length is much shorter than the theoretical upper bound.

Table 3.1: Performances (κ in %) of SCU, HJMR and SCU-UN when $L = 2$

	shape	2				3				4			
p	t	100	200	500	1000	100	200	500	1000	100	200	500	1000
20	SCU	29.0	22.7	9.0	5.4	26.7	20.5	7.5	4.3	24.6	18.2	6.2	3.6
	HJMR	64.9	46.0	16.8	10.8	93.0	71.4	24.7	15.4	113.3	97.7	32.7	20.0
	SCU-UN	22.4	16.5	5.5	3.1	19.4	13.7	4.2	2.4	17.4	11.8	3.5	2.0
30	SCU	38.0	29.0	10.7	6.3	34.4	25.5	8.6	4.9	30.6	22.7	7.6	4.3
	HJMR	52.7	47.2	20.8	13.0	76.3	78.0	35.6	21.1	95.1	105.1	51.2	29.5
	SCU-UN	27.9	19.4	6.2	3.5	23.8	16.2	4.8	2.7	20.6	13.7	4.0	2.3
40	SCU	42.3	32.2	11.4	6.6	35.8	26.5	8.8	5.0	31.3	22.8	7.6	4.3
	HJMR	43.5	45.5	25.2	16.4	65.8	76.7	46.4	28.8	81.0	101.6	68.3	43.2
	SCU-UN	31.0	21.5	6.7	3.8	25.5	17.4	5.3	3.0	22.3	15.1	4.6	2.6
50	SCU	44.2	33.2	11.7	6.8	38.3	27.9	9.4	5.4	33.5	24.7	8.2	4.7
	HJMR	36.9	42.5	28.0	19.1	57.9	72.9	52.4	36.5	72.3	98.3	77.0	56.0
	SCU-UN	32.4	22.4	6.9	3.9	27.3	18.4	5.6	3.3	23.2	16.0	5.1	3.0

Table 3.2: Performances (κ in %) of SCU, HJMR and SCU-UN when $L = 4$

	shape	2				3				4			
p	t	100	200	500	1000	100	200	500	1000	100	200	500	1000
20	SCU	17.7	15.9	7.5	4.7	17.0	14.8	6.2	3.7	15.2	13.4	5.3	3.1
	HJMR	113.6	105.6	37.9	22.3	141.5	145.0	52.6	30.6	161.4	177.7	66.8	38.8
	SCU-UN	15.3	12.1	4.6	2.7	14.0	10.9	3.7	2.1	13.5	9.9	3.2	1.8
30	SCU	27.8	24.0	9.9	5.8	24.2	20.6	8.1	4.7	21.9	18.3	6.9	4.1
	HJMR	89.3	104.4	54.9	31.9	110.4	142.6	84.6	49.0	127.9	176.8	111.8	64.5
	SCU-UN	22.7	17.3	5.9	3.3	19.8	14.8	4.7	2.6	18.0	13.3	4.2	2.4
40	SCU	33.0	27.7	10.7	6.2	29.3	24.4	9.0	5.2	26.2	22.2	8.3	5.0
	HJMR	73.2	98.5	68.1	44.8	93.7	137.1	105.9	72.3	105.9	164.8	137.3	97.2
	SCU-UN	27.4	20.2	6.6	3.7	23.9	17.5	5.5	3.0	21.2	15.7	5.1	2.9
50	SCU	37.1	30.9	12.0	7.0	33.9	27.3	9.9	5.8	29.8	24.9	10.2	6.6
	HJMR	63.5	93.2	74.9	55.1	80.1	128.8	115.4	89.5	90.9	154.9	147.7	119.3
	SCU-UN	30.6	22.5	7.2	4.0	25.6	19.2	6.2	3.5	23.4	17.8	6.7	4.0

Table 3.3: Performances (κ in %) of SCU, HJMR and SCU-UN when $L = 6$

	shape	2				3				4			
p	t	100	200	500	1000	100	200	500	1000	100	200	500	1000
20	SCU	10.8	10.6	5.7	3.6	10.5	10.1	5.1	3.1	9.7	9.4	4.6	2.8
	HJMR	144.0	160.8	62.8	35.8	169.2	209.0	86.1	48.6	186.9	244.7	105.6	59.6
	SCU-UN	9.7	8.7	3.7	2.2	9.3	8.3	3.1	1.9	9.0	7.8	2.9	1.7
30	SCU	19.8	18.1	8.8	5.3	17.6	16.4	7.2	4.3	16.2	15.6	7.0	4.3
	HJMR	111.3	152.5	96.0	56.8	131.4	199.2	138.0	82.3	143.8	233.4	173.4	105.1
	SCU-UN	17.3	14.5	5.4	3.1	16.2	13.1	4.5	2.6	14.6	12.4	4.4	2.6
40	SCU	25.3	23.6	10.5	6.2	22.8	21.5	9.0	5.4	21.2	20.0	9.0	6.0
	HJMR	91.1	142.2	114.7	80.5	107.3	183.6	161.9	119.4	117.7	213.9	200.5	152.8
	SCU-UN	22.4	18.5	6.4	3.6	20.9	17.0	5.9	3.3	19.6	16.4	6.9	4.2
50	SCU	29.5	27.1	11.3	6.6	25.9	23.7	10.3	6.2	23.3	22.9	11.7	8.2
	HJMR	78.1	133.7	122.6	97.0	90.5	168.8	170.5	140.2	100.7	197.4	210.5	178.6
	SCU-UN	25.8	20.6	7.1	4.0	24.1	19.5	7.3	4.2	23.3	20.0	8.9	5.3

3.6 Conclusions

In this chapter, we proposed an improved nonparametric learning algorithm for the fundamental lost-sales inventory problem with positive lead times and censored demand, the simulated cycle-update (SCU) algorithm, and showed that its regret rate is $O(T^{1/2})$, which matches the lower bound of regret for any learning algorithms and closes the gap left in Huh et al. (2009a).

As its name suggests, the SCU algorithm constructs (random) cycles using a simulated system, and updates base-stock level at the beginning of each cycle. To overcome the challenges introduced by positive lead time and censored demand, we instituted two key ideas, namely, the withheld on-hand inventory and the double-phase gradient estimation. To analyze the performance of SCU algorithm, we introduced several bridging systems between the SCU-system and the optimal clairvoyant system. We also presented a simplified algorithm for the problem when the demand data is uncensored. Our numerical results demonstrated the effectiveness of the SCU algorithm.

CHAPTER IV

Approximation Algorithms for Perishable Inventory Systems with Correlated Demand

4.1 Introduction

In this chapter, we study the classic periodic-review stochastic inventory systems with perishable products. The product lifetime is known and fixed. Initial interest in these systems was sparked by blood bank applications (see, e.g., Prastacos (1984), Pierskalla (2004), Karaesmen et al. (2011)), but the scope of applications is far greater. For example, perishable products such as food items and pharmaceuticals constitute the majority of sales revenue of grocery retailing industry. Food Market Institute (2012) reported that perishables accounted for 52.63% of the 2011 total supermarket sales of about $459 billion, and mismanagement of perishable products (such as spoilage and shrinkage) represents a major threat to the profitability of companies in grocery retailing industry. A survey by the National Supermarket Research Group reported an average loss of $34 million a year due to spoilage in one major 300-store grocery chain. Thus, finding effective inventory management policies for perishable products has always been of great interest to both practitioners and academic researchers.

We restrict our attention to the first-in-first-out (FIFO) issuing policy which is commonly adopted in the literature (see, e.g., Nahmias (2011) and Karaesmen et al. (2011)), i.e., the oldest inventory is consumed first when demand arrives. This assumption is reasonable for blood inventory systems and online grocery network (e.g., AmazonFresh). It also applies to the retailers who display only the oldest items on the shelves. The demands in different periods in our model can be non-stationary (or time-dependent) and correlated over time, capturing such demand features as seasonality and forecast updates as well as many other demand processes of practical interest. Both backlog model and lost-sales model will be

studied.

These systems are fundamental but notoriously hard to analyze in both theory and computation. As seen from our literature review below, the optimal policy is very complex even when the demands are independent and identically distributed. The optimal order quantity depends on both the age distribution of the on-hand inventory and the length till the end of the planning horizon. The computation of the optimal policy using dynamic programming is in general intractable due to the *"curse-of-dimensionality"*. Thus, many researchers turned to seek effective heuristic policies for these problems. To the best of our knowledge, almost all heuristics developed thus far have been focused on the case with independent and identically distributed demands. Moreover, none of the heuristic policies in the literature admits provably worst-case performance guarantees. In this chapter, we propose the first approximation algorithm with a worst-case performance guarantee of 2 for these important systems when the demands in different periods are independent and stochastically non-decreasing over time.

The demand processes in practical settings are often seasonal, forecast-based, or driven by the state of the economy (e.g., Markov modulated demand processes). The demands of these processes are correlated over time. For example, many firms employ forecasting methods to learn about the future demands and periodically update their forecasts; and such forecast-based demand processes can often be modeled by the martingale model of forecast evolution (MMFE for short, see, e.g., Graves et al. (1986), and Heath and Jackson (1994)), in which the updated forecast is the original forecast plus an adjustment (or random error) with mean 0. In addition, in practice firms often receive advance demand information (ADI) from some customers for the future periods, so managers have to periodically incorporate such ADI in the future demands (see, e.g., Gallego and Özer (2001)). These models are practically important, but finding the optimal policies using brute-force dynamic programming is computationally intractable, since the state space of the corresponding dynamic programs is usually large (see, e.g., Lu et al. (2006)). The computational burden is even more severe for perishable inventory systems, given the fact that the age distribution of on-hand inventory has to be tracked too.

To overcome this prohibitive computational challenge, we propose another approximation algorithm for perishable inventory systems with arbitrarily correlated demand processes, and show that it admits a worst-case performance guarantee less than 3. To the best of our knowledge, no effective heuristic policy has ever been developed in the literature for perishable inventory systems with correlated demand processes.

4.1.1 Main Results and Contributions

The main results and contributions of this chapter are summarized as follows.

Algorithms.

We develop two approximation algorithms which admit provably worst-case performance guarantees for perishable inventory systems.

Firstly, we develop a proportional-balancing (PB) policy for the perishable inventory systems with product lifetime of m (≥ 2) periods under an arbitrarily non-stationary and correlated demand process. We show that the PB policy has a worst-case performance guarantee of $\left(2 + \frac{(m-2)h}{mh+\theta}\right)$, where $h = \hat{h} + (1-\alpha)\hat{c}$, $\theta = \hat{\theta} + \alpha\hat{c}$, and $\hat{c}$, $\hat{h}$, $\hat{\theta}$, and α are the per-unit ordering, holding, outdating costs, and one-period discount factor, respectively. That is, for any instance of the problem, the expected cost of the PB policy is at most $\left(2 + \frac{(m-2)h}{mh+\theta}\right)$ times the expected cost of an optimal policy. Therefore, the theoretical worst-case performance guarantee is between 2 and 3 and it equals 2 when the product lifetime $m = 2$.

Secondly, when the demand process is independent and stochastically non-decreasing over time, we develop a dual-balancing (DB) policy which has a worst-case performance guarantee of 2, i.e., for any instance of the problem, the expected cost of the DB policy is at most twice the expected cost of an optimal policy.

To the best of our knowledge, our proposed policies are the first set of computationally efficient policies with worst-case performance guarantees in stochastic periodic-review perishable inventory systems. In contrast, computing the exact optimal policy using dynamic programming suffers from the well-known "curse of dimensionality" and is intractable even with short product lifetimes (e.g., $m = 4$) and under independent and identically distributed demand processes.

Worst-case analysis.

In our algorithmic design, we develop a nested marginal cost accounting scheme for perishable inventory systems. This scheme is similar in spirit to that developed in Levi et al. (2007a), but has a much more complex and nested structure due to the multi-dimensional inventory state representing the age distribution of on-hand inventory. The main idea of this approach is to associate the costs with each ordering decision instead of each period. However, the techniques developed for our worst-case performance analysis depart significantly from those in the previous studies (e.g., Levi et al. (2007a), Levi et al. (2008a)), which heavily

rely on the existence of a one-to-one matching between the supply and demand units when the inventory units are consumed in an first-in-first-out manner. That is, when analyzing the performances of the approximation algorithms, all the previous studies "*geometrically*" match product units in a one-to-one manner for the systems operating under two different policies; and the costs for each pair of matched units can be readily compared. However, the perishability of products destroys this matching mechanism, thus the existing techniques developed for non-perishable inventory systems are no longer applicable. To overcome this difficulty, we introduce a key new concept, called the "*trimmed on-hand inventory level*", defined as the part of on-hand inventory units ordered before a particular time. This key concept enables us to compare the costs of two perishable inventory systems operating under two different policies. Compared with the previous geometric approach, our new approach is purely algebraic and we expect it to be useful in studying other perishable inventory systems.

Empirical performances.

Our extensive computational studies show that our proposed policies perform consistently near-optimal for all the tested instances, which are significantly better than the theoretical worst-case performance guarantees. More specifically, for independent and identically distributed demand processes and short product lifetimes (for which the optimal policies can be numerically computed), our numerical results are comparable to those reported in Nahmias (1976, 1977b) and are very close to the optimal (around 0.3% above the optimal cost); for long product lifetimes for which computing optimal policies is intractable, we compare the performance of our methods with Nahmias (1976, 1977b); the overall performance of our policies is also comparable to those of Nahmias (1976, 1977b) and it improves as the frequency of oudating increases. For non-stationary and correlated demand processes including Markov modulated demand process, Martingale models of forecast evolution (MMFE), autoregressive (AR) models, and models with advance demand information (ADI), the proposed algorithms also perform close to optimal for all the problem instances tested – with maximum performance error below 3%, and average error below 1%, of the optimal costs.

4.1.2 Literature Review

The problem of managing stochastic periodic-review inventory systems with perishable products has attracted the attention of many researchers over the years. The dominant paradigm in the existing literature has been to formulate these models using a dynamic programming framework. Nahmias and Pierskalla (1973) characterized the structure of the

optimal ordering policy for a two-period product lifetime problem. Nahmias (1975b) and Fries (1975) then, independently, studied the optimal policy for the general lifetime problem with independent and identically distributed (i.i.d.) demands, in a backlogging model and a lost-sales model, respectively. They showed that the optimal ordering quantity depends on both the age distribution of the current inventory and the remaining time until the end of planning horizon. Thus, computing the optimal policy using brute-force dynamic programming is intractable due to the multi-dimensional state space. The complexity of this problem is later reinforced by Cohen (1976) who characterized the stationary distribution of inventory for the two-period lifetime problem, and showed that the optimal policy is a state-dependent policy. Recently, Li and Yu (2014) revisited the structural properties of the optimal inventory policy in perishable inventory systems by employing the "*multimodularity*" concept; and Chen et al. (2014b) studied the joint inventory control and pricing problem for perishable inventory systems and characterized the structural properties of the optimal inventory and pricing policies.

Due to the complexity of the optimal policies for perishable inventory systems, a lot of efforts have been dedicated to the design of efficient heuristic policies for both backlogging and lost-sales models. When the demands are i.i.d., Nahmias (1976) constructed a bound on the outdating cost which is a function of only the total on-hand inventory, and developed a base-stock myopic policy using this bound. Nahmias (1977a) employed the same myopic policies to compare two dynamic perishable inventory models developed by Nahmias (1975b) and Fries (1975). Subsequently, Nahmias (1977b) used a more refined approximate state transition function which treats the product lifetime as if it were only two periods, thereby resulting in a one-dimensional state variable. The numerical results for problems with product lifetimes of 2 and 3 periods are near-optimal under i.i.d. demands. Nandakumar and Morton (1993) derived upper and lower bounds for the dynamic programming formulation of the lost-sales model, and used a weighted average of the lower and upper bounds to construct an efficient heuristic. The numerical results showed that the heuristic again performs close to optimal for short product lifetimes and i.i.d. demands. Cooper (2001) derived bounds on the stationary distribution of the number of outdated units in each period, under a fixed critical number order policy. Recently, following the approximation scheme of the outdating costs developed by Nahmias (1976), Chen et al. (2014b) proposed two heuristic policies for the joint inventory control and pricing models. Since the future demands depend on the future prices, they approximated the expected demands and prices in the future periods by solving the corresponding deterministic models. Their numerical study showed that the two heuristic policies perform very well when the demands are i.i.d., but the performance error

could go up to 15% for the independent but time-varying demands. Li et al. (2009) also designed an effective heuristic following the approximation method of Nahmias (1976) for the joint inventory control and pricing model. Li et al. (2013) proposed two myopic heuristics for perishable inventory systems with last-in-first-out issuing policy and clearance sales. To reduce the state space, the first heuristic treats all on-hand inventories as if they would expire in one period whereas the second heuristic keeps track of the total inventory level and the inventory level of items whose remaining lifetime is one period. Their numerical results showed that these heuristics perform very close to optimal under i.i.d. demands. As seen from the literature above, almost all the existing heuristic policies have been focused on i.i.d. demands, and none of them admits provably worst-case performance guarantees.

There is also a large body of literature discussing other aspects of perishable inventory systems. We partition these studies into the following categories (our list below is by no means exhaustive): (a) continuous-review perishable inventory systems (see, e.g., Weiss (1980), Goh et al. (1993), Liu and Lian (1999), Perry (1999)); (b) perishable inventory systems with multiple products or demands (see, e.g., Nahmias and Pierskalla (1976), Deuermeyer (1979), Ferguson and Koenigsberg (2007), Deniz et al. (2010), Cai and Zhou (2014)); (c) joint inventory and pricing control of perishables (see, e.g., Li et al. (2009), Chen and Sapra (2013), Chen et al. (2014b)); (d) perishable inventory systems with depletion or clearance sales (see, e.g., Cai et al. (2009), Xue et al. (2011), Li and Yu (2014), Li et al. (2013)); and (f) blood banks and health-care applications (see, e.g., Prastacos (1984), Haijema et al. (2005, 2007), Zhou et al. (2011)). We also refer interested readers to Nahmias (1982, 2011), Goyal and Giri (2001), and Karaesmen et al. (2011) for comprehensive literature reviews.

Our work is also closely related to the recent stream of literature on approximation algorithms for stochastic periodic-review inventory systems pioneered by Levi et al. (2007a). The systems allow for correlated stochastic demand processes, including all of the known approaches to model dynamic demand forecast updates (e.g., Gallego and Özer (2001), Iida and Zipkin (2006), and Lu et al. (2006)). Levi et al. (2007a) first introduced the concept of marginal cost accounting which associates the costs with each decision a particular policy makes. They proposed a 2-approximation policy which admits a worst-case performance guarantee of 2 for the backlogging model with generally correlated demands. Subsequently, Levi et al. (2008a) proposed a 2-approximation algorithm for the lost-sales models under independent demand processes; and Levi et al. (2008b) introduced the concept of forced marginal backlogging cost accounting and proposed a 2-approximation algorithm for the capacitated systems with backlogging. More recently, Levi and Shi (2013) and Shi et al. (2014) proposed two approximation algorithms for the lot-sizing backlogging models without and with

capacity constraints, respectively; and Tao and Zhou (2014) proposed a 2-approximation algorithm for inventory systems with remanufacturing. All these previous studies assume that the inventory is non-perishable, and therefore there exists a one-to-one matching between the supply and demand when the inventory issuing policy is FIFO. Perishability, however, destroys this matching mechanism and the existing techniques for the worst-case analysis cannot be applied to perishable inventory systems.

4.1.3 Structure

The remainder of this chapter is organized as follows. In §4.2, we present the mathematical formulation for the backlogging model. In §4.3, we design a nested marginal cost accounting scheme for perishable inventory systems. In §4.4, we construct the proportional-balancing policy and the dual-balancing policy, and provide their worst-case performance guarantees. In §4.5, we provide the main proofs, while leaving some of the more involved technical details in the supplemental material. In §4.6, we conduct an extensive numerical study on our proposed policies. Finally, we conclude the chapter in §4.7. Throughout this chapter, for any real numbers x and y, we denote $x^+ = \max\{x, 0\}$, $x \vee y = \max\{x, y\}$, and $x \wedge y = \min\{x, y\}$. In addition, for a sequence $x_1, x_2, \ldots$ and any integers t and s with $t \leq s$, we denote $x_{[t,s]} = \sum_{j=t}^{s} x_j$ and $x_{[t,s)} = \sum_{j=t}^{s-1} x_j$. Also, we use "expiration", "outdating", and "perishing", interchangeably.

4.2 Stochastic Periodic-Review Perishable Inventory Control Problem

In this section, we provide the mathematical formulation of the stochastic periodic-review perishable inventory system over a planning horizon of T (possibly infinite) periods, indexed by $t = 1, \ldots, T$. The lifetime of the product is m periods, i.e., a product perishes after staying m periods in stock. Our model allows for a non-stationary and generally correlated demand process. We assume that the order lead time is zero, i.e., an order placed at the beginning of a period can be used in the same period. This is a common assumption in the perishable inventory literature (see Karaesmen et al. (2011)). We shall focus our presentation on the backlogging model but our results can also be easily extended to lost-sales models.

Demand structure. The demands over the planning horizon are random, denoted by $D_1, \ldots, D_T$. The demands in different periods can be non-stationary and correlated over time. At the beginning of each period t, there is an observed *information set* denoted by f_t, which contains all of the information accumulated up to period t. More specifically, the

information set f_t consists of the realized demands $d_1, \ldots, d_{t-1}$ in the first $t-1$ periods, and possibly some exogenous information denoted by $(w_1, \ldots, w_t)$. The information set f_t in period t is one specific realization in the set of all possible realizations of the random vector $F_t = (D_1, \ldots, D_{t-1}, W_1, \ldots, W_t)$. The set of all possible realizations is denoted by $\mathscr{F}_t$. With the information set f_t, the conditional joint distribution of the future demands $(D_t, \ldots, D_T)$ is known. We assume that the conditional expectations, given f_t, are well defined. Note that this demand structure is very general, that includes a wide range of demand processes such as Markov modulated demand process (see, e.g. Song and Zipkin (1993) and Sethi and Cheng (1997)), MMFE (see, e.g., Heath and Jackson (1994)), AR(p), ARMA(p, q), ARIMA(p, r, q), (see, e.g., Mills (1990)), and models with advance demand information (ADI) (see, e.g., Gallego and Özer (2001)), among others.

Cost structure. In each period t, four types of costs may occur: a unit ordering cost $\hat{c}$, a unit holding cost $\hat{h}$ for leftover inventory, a unit backlogging cost $\hat{b}$ for unsatisfied demand, and a unit outdating cost $\hat{\theta}$ for expired products. There is also a one-period discount factor α, with $0 < \alpha \leq 1$ when $T < \infty$ and $0 < \alpha < 1$ when $T = \infty$. We assume that $\hat{b} > (1-\alpha)\hat{c}$ and $\hat{\theta} + \alpha\hat{c} \geq 0$. Thus, $\hat{\theta}$ can be negative, and in this case it can be interpreted as unit salvage value. Following Nahmias (1975b) we assume that any remaining inventory at the end of the planning horizon can be salvaged with a return of $\hat{c}$ per unit and unsatisfied demand can be satisfied by an emergency order at a cost of $\hat{c}$ per unit. We note that our analysis can be extended to the case with a unit salvage value $\hat{v}$ for any on-hand inventory and a unit penalty cost $\hat{p}$ for any unsatisfied demand at the end of the planning horizon, as long as $\hat{v} \leq \hat{c}$ and $\hat{b} + \alpha\hat{p} > \hat{c}$. However, our analysis cannot be directly extended to the case with age-dependent salvage values.

System dynamics. For each period t, the sequence of events is as follows. First, at the beginning of period t, the information set $f_t \in \mathscr{F}_t$ and the inventory vector

$$\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,m-1}) \tag{4.1}$$

are observed, where $x_{t,i}$ is the quantity of on-hand products whose remaining lifetime is i periods, $i = 1, \ldots, m-2$, and $x_{t,m-1}$ is the quantity of on-hand products whose remaining lifetime is $m-1$ periods minus the quantity of backlogged demands (if any). Thus, $x_{t,1}, \ldots, x_{t,m-2}$ are always nonnegative; while $x_{t,m-1}$ can be positive or negative. For simplicity, we assume that the inventory system is initially empty at the beginning of period 1, i.e., $x_{1,i} = 0$, for all $i = 1, \ldots, m-1$; but our analysis and results can be extended to the case with an arbitrary initial state.

Second, an order with quantity q_t is placed, incurring an ordering cost $\hat{c}q_t$. Following the discussions in Levi et al. (2007a), we assume that q_t is a continuous decision variable, but it can be extended to the case of integer values. Denote y_t as the total inventory level after receiving the order in period t. Then, $y_t = \sum_{i=1}^{m-1} x_{t,i} + q_t$.

Third, the demand in period t is realized and satisfied as much as possible by the on-hand inventory using the FIFO issuing policy, i.e., the oldest inventory is consumed first when demand arrives. At the end of period t, if $y_t - D_t \geq 0$, then the excess inventory incurs a holding cost $\hat{h}(y_t - D_t)$. Following Nahmias (1975b), we assume that all excess inventory (including the inventory which perishes at the end of this period) incurs a holding cost. On the other hand, if $y_t - D_t < 0$, then the system incurs a backlogging cost $\hat{b}(D_t - y_t)$. Furthermore, if the inventory with one period remaining life $x_{t,1} > D_t$, then $e_t := (x_{t,1} - D_t)^+$ units perish and incur an outdating cost $\hat{\theta}e_t$.

Finally, the system proceeds to the subsequent period $t+1$. By the definition of the inventory vector $\mathbf{x}_t$ and the FIFO issuing policy, we obtain the following state transition from $\mathbf{x}_t$ to $\mathbf{x}_{t+1}$:

$$\begin{aligned} x_{t+1,j} &= \left(x_{t,j+1} - \left(D_t - \sum_{i=1}^{j} x_{t,i} \right)^+ \right)^+, \text{ for } 1 \leq j \leq m-2, \\ x_{t+1,m-1} &= q_t - \left(D_t - \sum_{i=1}^{m-1} x_{t,i} \right)^+. \end{aligned} \tag{4.2}$$

We remark that in defining the inventory state $\mathbf{x}_t$ in (4.1), it is convenient and natural to combine the inventory having $m-1$ periods of remaining life with the number of backlogs in $x_{t,m-1}$. This is because when demand arrives, by the FIFO issuing policy, it is first met by $x_{t,1}$, and when $x_{t,1}$ is consumed then the remaining demand is met by $x_{t,2}$. This process continues and when (and if) $x_{t,m-2}$ also depletes to 0, the remaining demand will be satisfied by $x_{t,m-1}$. Clearly, when the demand is large, this last number will continue to go down after reaching 0, representing the backlog level. We also note that inventory only outdates through the first dimension, $x_{t,1}$, of vector $\mathbf{x}_t$, while backlogs always stay in the last dimension, $x_{t,m-1}$ (hence backlogs will *not* disappear after m periods). Moreover, if in period t there are backlogs (thus $x_{t,m-1}$ is negative and $x_{t,j} = 0$ for $j = 1, \ldots, m-2$), then by (4.2), in the next period $x_{t+1,j}$ will be equal to 0 for all $j = 1, \ldots, m-2$, but $x_{t+1,m-1}$ can be positive or negative, depending on whether q_t is greater or less than $D_t - x_{t,m-1}$.

Objective. For clarity, we often distinguish between a random variable and its realization using a capital letter and a lowercase letter, respectively. Then the expected total

discounted cost incurred under a given policy P that orders q_t in period t can be written as

$$\mathscr{C}(P) = \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(\hat{c}q_t + \hat{h}(Y_t - D_t)^+ + \hat{b}(D_t - Y_t)^+ + \hat{\theta}e_t\right) - \alpha^T \hat{c}\sum_{i=1}^{m-1} X_{T+1,i}\right]. \quad (4.3)$$

Note that, the quantities q_t, Y_t, e_t, and X_t all depend on the policy P; and whenever necessary, we shall make the dependency explicit, i.e., write them as q_t^P, Y_t^P, e_t^P, and X_t^P, respectively.

The objective is to coordinate the sequence of orders to minimize the expected total discounted cost. As discussed in Section 4.1, it is known that finding the exact optimal policy using dynamic programming is computationally intractable. Thus, our focus in this chapter is to design easy-to-compute and near-optimal approximation algorithms.

Approximation policy assessment. To measure the effectiveness of an approximation algorithm P, we define its performance measure by the ratio $\mathscr{C}(P)/\mathscr{C}(OPT)$, where $\mathscr{C}(OPT)$ is the cost under an optimal policy. Clearly, the value of this ratio depends on the problem instance, and is at least 1. If under algorithm P this ratio is always equal to 1 for all problem instances, then P is an optimal policy. Otherwise, if there exists some number r (> 1) such that this ratio is less than or equal to r for any problem instance, then we say that the algorithm admits a worst-case performance guarantee of r, or simply call it an r-approximation algorithm. As mentioned earlier, we will present efficient approximation algorithms for the perishable inventory systems with worst-case performance guarantees of 2 and 3, respectively.

Cost transformation. Next we carry out a cost transformation to obtain an equivalent model with the unit ordering cost equal to 0. This will enable us to assume, without loss of generality, that the unit ordering cost is 0 in the subsequent analysis.

Proposition 4.1. *For every perishable inventory system with cost parameters $\hat{c}$, $\hat{h}$, $\hat{b}$ and $\hat{\theta}$, there is an equivalent system with non-negative cost parameters $c = 0$, $h = \hat{h} + (1-\alpha)\hat{c}$, $b = \hat{b} - (1-\alpha)\hat{c}$ and $\theta = \hat{\theta} + \alpha\hat{c}$. And the expected total discounted cost can be rewritten as*

$$\mathscr{C}(P) = \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(h(Y_t - D_t)^+ + b(D_t - Y_t)^+ + \theta e_t\right)\right] + \sum_{t=1}^{T} \alpha^{t-1}\hat{c}\mathsf{E}\left[D_t\right]. \quad (4.4)$$

Proof. Recall that the amount of outdating products in period t is

$$e_t := (x_{t,1} - D_t)^+.$$

The starting inventory level in period $t+1$ is equal to the ending inventory level in period t

minus the demand and also the outdated units in period t, i.e.,

$$\sum_{i=1}^{m-1} x_{t+1,i} = Y_t - D_t - e_t. \tag{4.5}$$

Hence using the relationship

$$q_t = Y_t - \sum_{i=1}^{m-1} x_{t,i} = (Y_t - D_t)^+ - (D_t - Y_t)^+ + D_t - \sum_{i=1}^{m-1} x_{t,i}, \tag{4.6}$$

we can rewrite the cost $\mathscr{C}(P)$ in (4.3) as

$$\begin{aligned}
\mathscr{C}(P) &= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(\hat{c}q_t + \hat{h}(Y_t - D_t)^+ + \hat{b}(D_t - Y_t)^+ + \hat{\theta}e_t\right) - \alpha^T \hat{c}\sum_{i=1}^{m-1} X_{T+1,i}\right] \\
&= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(\hat{c}D_t - \hat{c}\sum_{i=1}^{m-1} x_{t,i} + (\hat{h}+\hat{c})(Y_t - D_t)^+ + (\hat{b}-\hat{c})(D_t - Y_t)^+ + \hat{\theta}e_t\right) - \alpha^T \hat{c}\sum_{i=1}^{m-1} X_{T+1,i}\right] \\
&= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(-\alpha\hat{c}\sum_{i=1}^{m-1} x_{t+1,i} + (\hat{h}+\hat{c})(Y_t - D_t)^+ + (\hat{b}-\hat{c})(D_t - Y_t)^+ + \hat{\theta}e_t\right)\right] + R \\
&= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(-\alpha\hat{c}(Y_t - D_t) + (\hat{h}+\hat{c})(Y_t - D_t)^+ + (\hat{b}-\hat{c})(D_t - Y_t)^+ + (\hat{\theta}+\alpha\hat{c})e_t\right)\right] + R \\
&= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left((\hat{h}+\hat{c}-\alpha\hat{c})(Y_t - D_t)^+ + (\hat{b}-\hat{c}+\alpha\hat{c})(D_t - Y_t)^+ + (\hat{\theta}+\alpha\hat{c})e_t\right)\right] + R \\
&= \mathsf{E}\left[\sum_{t=1}^{T} \alpha^{t-1}\left(h(Y_t - D_t)^+ + b(D_t - Y_t)^+ + \theta e_t\right)\right] + R,
\end{aligned}$$

where the second equality follows from (4.6), the fourth equality follows from (4.5), $h = \hat{h} + (1-\alpha)\hat{c}$, $b = \hat{b} - (1-\alpha)\hat{c}$, $\theta = \hat{\theta} + \alpha\hat{c}$, and

$$R = -\hat{c}\sum_{i=1}^{m-1} x_{1,i} + \sum_{t=1}^{T} \alpha^{t-1}\hat{c}\mathsf{E}[D_t] = \sum_{t=1}^{T} \alpha^{t-1}\hat{c}\mathsf{E}[D_t].$$

Note that we have used the assumption that the inventory system is initially empty, i.e., $x_{1,i} = 0$, for all $i = 1, \ldots, m-1$. The proof is complete. □

4.3 Nested Marginal Cost Accounting Scheme

The traditional cost accounting scheme given in (4.3) decomposes the total cost by periods. Levi et al. (2007a) presented a marginal cost accounting scheme for the classical

non-perishable inventory systems. In this section, we develop a marginal cost accounting scheme for perishable inventory systems, similar in spirit to that in Levi et al. (2007a). Our marginal cost accounting scheme exhibits a nested structure due to the multi-dimensionality of system state. The main idea underlying this approach is to decompose the total cost in terms of the marginal costs of individual decisions. That is, we associate the decision in period t with its affiliated cost contributions to the system. These marginal costs may include costs (associated with the decision) incurred in both the current and subsequent periods.

Given the inventory vector $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,m-1})$ at the beginning of period t, and that a policy P orders q_t, we aim to compute the marginal cost contributions to the system by these q_t units on the holding, outdating, and backlogging costs. To this end, for $i = 1, \ldots, m-1$, we let $B_t(\mathbf{x}_t, i)$ denote the number of outdated units in periods $[t, t+i-1]$ given that the inventory vector at the beginning of period t is $\mathbf{x}_t$, with the convention that $B_t(\mathbf{x}_t, 0) \equiv 0$. Then, for $1 \leq i \leq m-1$, we have

$$B_t(\mathbf{x}_t, i) = \max\Big\{ \sum_{j=1}^{i} x_{t,j} - D_{[t,t+i-1]},\ \ B_t(\mathbf{x}_t, i-1)\Big\}. \tag{4.7}$$

To see why this is true, note that $\sum_{j=1}^{i} x_{t,j} - B_t(\mathbf{x}_t, i-1)$ is the number of non-expired units in $x_{t,1}, \ldots, x_{t,i}$ that would meet demands in periods $[t, t+i-1]$. These units, if not consumed, will expire at the end of period $t+i-1$. Thus $\big(\sum_{j=1}^{i} x_{t,j} - B_t(\mathbf{x}_t, i-1) - D_{[t,t+i-1]}\big)^+$, if positive, would be the number of units that will expire at the end of period $t+i-1$. Adding $B_t(\mathbf{x}_t, i-1)$ to it gives the total number of expired units in $[t, t+i-1]$, which is (4.7).

The nested structure in the auxiliary function $B_t(\cdot, \cdot)$ follows from the fact that some inventory units reach their expiration date before meeting the demand, and have to be discarded from the on-hand inventory. Using this auxiliary function, the number of outdated units in period $t+i-1$, for $1 \leq i \leq m-1$, is given as

$$e_{t+i-1} = \Big(\sum_{j=1}^{i} x_{t,j} - B_t(\mathbf{x}_t, i-1) - D_{[t,t+i-1]}\Big)^+,$$

and the number of oudated units in period $t+m-1$ is

$$e_{t+m-1} = \Big(q_t + \sum_{j=1}^{m-1} x_{t,j} - B_t(\mathbf{x}_t, m-1) - D_{[t,t+m-1]}\Big)^+.$$

4.3.1 Nested Marginal Holding Cost Accounting

We first focus on the marginal holding cost accounting of a given policy P. The holding cost for the q_t units ordered in period t may be incurred in any period from t to $t+m-1$ (after which the remaining ones will perish), or T, whichever is smaller. Let $H_t^P(q_t)$ be the discounted marginal holding cost (to period 1) incurred by these q_t units. Then it follows from the FIFO issuing policy that

$$H_t^P(q_t) \quad := \quad h \sum_{i=t}^{(t+m-1)\wedge T} \alpha^{i-1} \left(q_t - \left(D_{[t,i]} + B_t(\mathbf{x}_t, i-t) - \sum_{j=1}^{m-1} x_{t,j} \right)^+ \right)^+, \qquad (4.8)$$

where the auxiliary function $B_t(\mathbf{x}_t, i)$ is given recursively via (4.7). To see why (4.8) is valid, note that the total number of units in $\mathbf{x}_t$ that do not expire until $t+i$ is $\sum_{j=1}^{m-1} x_{t,j} - B_t(\mathbf{x}_t, i)$, thus the net demand after consuming the units in $\mathbf{x}_t$ is $\left(D_{[t,t+i]} - (\sum_{j=1}^{m-1} x_{t,j} - B_t(\mathbf{x}_t, i)) \right)^+$. Hence, the number of unconsumed units from q_t at the end of period $t+i$ is $\left(q_t - \left(D_{[t,t+i]} + B_t(\mathbf{x}_t, i) - \sum_{j=1}^{m-1} x_{t,j} \right)^+ \right)^+$.

Because the marginal holding cost is computed based on the nested structure of the auxiliary function $B_t(\cdot,\cdot)$, we call it *the nested marginal holding cost accounting*. It is important to note that the marginal holding cost associated with the q_t units ordered in period t is only affected by the future demands but not by the future decisions.

4.3.2 Nested Marginal Outdating Cost Accounting

Similarly, we can compute the marginal outdating cost associated with the q_t units ordered by policy P in period t using the following nested scheme. For $t = 1, \ldots, T-m+1$,

$$\Theta_t^P(q_t) \; := \alpha^{t+m-2}\, \theta\, e_{t+m-1} = \alpha^{t+m-2}\, \theta \left(q_t + \sum_{j=1}^{m-1} x_{t,j} - B_t(\mathbf{x}_t, m-1) - D_{[t,t+m-1]} \right)^+, \qquad (4.9)$$

where $B_t(\cdot,\cdot)$ is defined in (4.7); and for $t = T-m+2, \ldots, T$, we have $\Theta_t^P \equiv 0$ since the ordered units do not expire within the planning horizon.

4.3.3 Marginal Backlogging Cost Accounting

For each period $t = 1, \ldots, T$, the discounted (to period 1) marginal backlogging cost of the q_t units ordered in period t by policy P can be expressed as

$$\Pi_t^P(q_t) := \alpha^{t-1}\, b \left(D_t - \sum_{i=1}^{m-1} x_{t,i} - q_t\right)^+, \tag{4.10}$$

which is exactly the same as the traditional backlogging cost using the period-by-period accounting scheme. The intuition is that any negative consequence of under-ordering can be corrected by placing an order in the next period; thus it suffices to only consider the backlogging cost incurred in the current period.

4.3.4 Total Cost of a Given Policy

Note that the marginal costs defined above, $H_t^P(q_t)$, $\Theta_t^P(q_t)$, and $\Pi_t^P(q_t)$, are random as they depend on the future demands. Since the system is initially empty, the expected total system cost $\mathscr{C}(P)$ of policy P can be obtained by summing (4.8), (4.9) and (4.10) over t from 1 to T, and then taking expectations. Thus, by (4.4) we have

$$\mathscr{C}(P) \;=\; \mathsf{E}\Big[\sum_{t=1}^{T}\Big(H_t^P(q_t) + \Pi_t^P(q_t) + \Theta_t^P(q_t)\Big)\Big] + \sum_{t=1}^{T} \alpha^{t-1}\hat{c}\mathsf{E}[D_t]. \tag{4.11}$$

If we ignore the constant terms that are independent of the policy, then we can write the effective cost of a policy P as

$$C(P) = \mathsf{E}\Big[\sum_{t=1}^{T}\Big(H_t^P(q_t) + \Pi_t^P(q_t) + \Theta_t^P(q_t)\Big)\Big]. \tag{4.12}$$

Clearly, to compare different policies, we only need to compare their effective costs.

4.4 Balancing Policies and Worst-Case Performance Guarantees

In this section, we propose two efficient cost-balancing algorithms for perishable inventory systems with general product lifetime using the nested cost accounting scheme defined in the previous section. The first one is for arbitrary non-stationary and correlated demand processes; while the second is for independent and stochastically non-decreasing demand processes. These policies will be shown to admit worst-case performance guarantees of 3 and

2, respectively.

4.4.1 Proportional-Balancing (PB) Policy

For each period $t = 1, \ldots, T$, with an observed information set $f_t \in \mathscr{F}_t$, the proportional-balancing (PB) policy orders $q_t^{PB} = q_t$ that balances a proportion of the expected marginal holding and outdating costs with the expected backlogging cost as follows:

$$\tfrac{mh+\theta}{2(m-1)h+\theta}\mathsf{E}[H_t^{PB}(q_t) + \Theta_t^{PB}(q_t) \mid f_t] = \mathsf{E}[\Pi_t^{PB}(q_t) \mid f_t]. \tag{4.13}$$

It can be verified that the left hand side (LHS) of (4.13) is an increasing convex function of the order quantity q_t, which equals 0 when $q_t = 0$ and approaches infinity when q_t tends to infinity. On the other hand, the right hand side (RHS) of (4.13) is a decreasing convex function of the order quantity q_t, which equals a non-negative number when $q_t = 0$ and tends to 0 when q_t goes to infinity. Since q_t can take any non-negative real value and both functions are continuous, q_t in (4.13) is well defined. Furthermore, since LHS minus RHS of (4.13) is increasing in q_t, q_t^{PB} can be very efficiently computed using bisection methods. It should be noted that q_t^{PB} is a function of f_t and $\mathbf{x}_t$, but for simplicity we make this dependency implicit.

For the special case where the demands in different periods are independent, q_t^{PB} does not depend on the information set f_t, and it becomes a function of only the inventory vector $\mathbf{x}_t$ at the beginning of period t. Several studies in the literature have analyzed the qualitative properties of the optimal order quantity q_t^{OPT} on the starting inventory vector of period t for the case of independent and identically distributed demands (see, e.g., Fries (1975) and Nahmias (1982)). Suppose the inventory vector at the beginning of period t is $\mathbf{x}_t = (x_1, \ldots, x_{m-1})$. It has been shown that q_t^{OPT} decreases at a rate less than one when the product inventory of any age group increases, and that it decreases more rapidly in the inventory level of newer product than that of older product. The following result shows that, the order quantity q_t^{PB} under the PB policy satisfies these properties as well.

Proposition 4.2. *For each period t, the order quantity q_t^{PB} under the PB policy satisfies*

$$-1 \leq \frac{\partial q_t^{PB}}{\partial x_{m-1}} \leq \frac{\partial q_t^{PB}}{\partial x_{m-2}} \leq \cdots \leq \frac{\partial q_t^{PB}}{\partial x_1} \leq 0.$$

Proof. The marginal costs are functions of $\mathbf{x}$ and q, thus in this proof we write them as

$H_t^{PB}(\mathbf{x}, q)$, $\Theta_t^{PB}(\mathbf{x}, q)$, and $\Pi_t^{PB}(\mathbf{x}, q)$. Define

$$h_t(\mathbf{x}, q) = \tfrac{mh+\theta}{2(m-1)h+\theta}\mathsf{E}[H_t^{PB}(\mathbf{x}, q) + \Theta_t^{PB}(\mathbf{x}, q) \mid f_t] - \mathsf{E}[\Pi_t^{PB}(\mathbf{x}, q) \mid f_t].$$

Then, q_t^{PB} is the solution of the equation $h_t(\mathbf{x}, q) = 0$. First, one can easily verify that $h_t(\mathbf{x}, q)$ is increasing in $\mathbf{x}$ and in q. Thus, it follows that q_t^{PB} is decreasing in x_i, $i = 1, \ldots, m-1$. Next, we argue that to complete the proof it suffices to show that $h_t(x_1, x_2-x_1, \ldots, x_{m-1}-x_{m-2}, \hat{q}-x_{m-1})$ is decreasing in $(x_1, \ldots, x_{m-1})$ while increasing in $\hat{q}$. Now suppose these results are true. Define $\hat{q}^*(x_1, \ldots, x_{m-1})$ as the solution of the equation $h_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q}-x_{m-1}) = 0$. Then, it follows from our assumption that $\hat{q}^*(x_1, \ldots, x_{m-1})$ is increasing in $(x_1, \ldots, x_{m-1})$. From the definition of q_t^{PB}, we must have, if we write the dependency on state variables explicit,

$$q_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}) = \hat{q}^*(x_1, \ldots, x_{m-1}) - x_{m-1}.$$

After taking derivative with respect to x_i, $i = 1, \ldots, m-1$, we obtain

$$-1 \leq \frac{\partial q_t^{PB}}{\partial x_{m-1}} \leq \frac{\partial q_t^{PB}}{\partial x_{m-2}} \leq \cdots \leq \frac{\partial q_t^{PB}}{\partial x_1}.$$

Now we prove that $h_t(x_1, x_2-x_1, \ldots, x_{m-1}-x_{m-2}, \hat{q}-x_{m-1})$ is decreasing in $(x_1, \ldots, x_{m-1})$ while increasing in $\hat{q}$. To this end, it suffices to show that, for any realizations of demands $D_t, \ldots, D_T$, $H_t^{PB}(x_1, x_2-x_1, \ldots, x_{m-1}-x_{m-2}, \hat{q}-x_{m-1})$, $\Theta_t^{PB}(x_1, x_2-x_1, \ldots, x_{m-1}-x_{m-2}, \hat{q}-x_{m-1})$ and $-\Pi_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q} - x_{m-1})$ all satisfy the above properties. First, notice that

$$-\Pi_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q} - x_{m-1}) = -\alpha^{t-1}b(D_t - \hat{q})^+.$$

Thus the desired results are trivially true.

To show that $H_t^{PB}(\cdot)$ and $\Theta_t^{PB}(\cdot)$ also satisfy the desired results, we first prove by induction that $B_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, i)$ is increasing in $(x_1, \ldots, x_i)$ and independent in $(x_{i+1}, \ldots, x_{m-1})$, $i = 1, \ldots, m-1$. When $i = 1$, the results are obviously true. Now suppose the results hold for i. For $i + 1$, we have

$$\begin{aligned} &B_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, i+1) \\ =&\max\left\{x_{i+1} - D_{[t,t+i-1]}, B_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, i)\right\}. \end{aligned}$$

Then, it follows from the above expression that $B_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, i+1)$ is increasing in $(x_1, \ldots, x_{i+1})$ and independent in $(x_{i+2}, \ldots, x_{m-1})$. Hence, by induction, the desired results on $B_t(\cdot, i)$ hold for $i = 1, \ldots, m-1$.

Now consider $\Theta_t^{PB}(\cdot)$. Since

$$\begin{aligned}&\Theta_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q} - x_{m-1})\\ =&\alpha^{t+m-2}\theta(\hat{q} - B_t(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, m-1) - D_{[t,t+m-1]})^+,\end{aligned}$$

then $\Theta_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q} - x_{m-1})$ is decreasing in $(x_1, \ldots, x_{m-1})$ while increasing in $\hat{q}$. Similarly, we can prove that the desired results hold for $H_t^{PB}(x_1, x_2 - x_1, \ldots, x_{m-1} - x_{m-2}, \hat{q} - x_{m-1})$. The proof is complete. □

The more important question is how well the PB policy performs. In what follows, we first provide a theoretical worst-case performance guarantee; then in Section 6, we will provide a comprehensive numerical study to demonstrate its empirical performance.

Theorem 4.3. *For an arbitrary non-stationary and correlated demand process, the proportional-balancing policy for the perishable inventory system with $m \geq 2$ periods of product lifetime has a worst-case performance guarantee of $\left(2 + \frac{(m-2)h}{mh+\theta}\right)$, i.e., for any instance of the problem, the expected cost of the proportional-balancing policy is at most $\left(2 + \frac{(m-2)h}{mh+\theta}\right)$ times the expected cost of an optimal policy.*

Theorem 4.3 shows that, when the product lifetime $m = 2$, the PB policy has a worst-case performance guarantee of 2; while for a general lifetime m, the PB policy has a worst-case performance guarantee between 2 and 3.

We remark that the balancing coefficient on the LHS of (4.13) is chosen so that the resulting PB policy admits our best provable worst-case performance guarantee. If we select a general positive balancing coefficient β to construct the PB policy, then we can prove that it admits a worst-case performance guarantee of $(\beta+1)/\min\{\beta, \beta_0\}$, where $\beta_0 = \frac{mh+\theta}{2(m-1)h+\theta}$. Since the worst-case performance guarantee is minimized when $\beta = \beta_0$, we construct the PB policy with this optimized parameter.

4.4.2 Dual-Balancing (DB) Policy

In this subsection, we propose another approximation policy, referred to as the dual-balancing policy, which has a worst-case performance of 2 for an arbitrary fixed lifetime m when the demands $D_1, \ldots, D_T$ are independent and stochastically non-decreasing over

time. The random variables $D_1, \ldots, D_T$ are said to be stochastically non-decreasing if for any $1 \leq t \leq s \leq T$, D_t is less than D_s in the *usual stochastic order*, or equivalently, $\Pr(D_t > d) \leq \Pr(D_s > d)$ for all d. For more detailed discussions on stochastic orders, we refer interested readers to Shaked and Shanthikumar (2007). In the remainder of this section, we assume that the demands $D_1, \ldots, D_T$ are independent and stochastically non-decreasing.

To introduce the dual-balancing policy, we first define the discounted (to period 1) marginal holding cost for period t for an arbitrary policy P by

$$\hat{H}_t^P(q_t) \;:=\; \alpha^{t-1}\, h\, \left(\sum_{i=1}^{m-1} x_{t,i} + q_t - D_t\right)^+.$$

Then, we define the discounted marginal outdating and discounted marginal backlogging costs for a policy P in exactly the same way as those in (4.9) and (4.10). In addition, for each period t, let S_t be the solution of y to the equation $h\mathsf{E}[(y - D_t)^+] = b\mathsf{E}[(D_t - y)^+]$, which depends only on the distribution of D_t in period t. Since the demands $D_1, \ldots, D_T$ are stochastically non-decreasing, it follows that S_t is non-decreasing in t.

The *dual-balancing* (DB) policy is described as follows: Suppose at the beginning of period t the state is $\mathbf{x}_t = (x_{t,1}, \ldots, x_{t,m-1})$, the DB policy orders $q_t^{DB} = q_t$ if $\sum_{i=1}^{m-1} x_{t,i} \leq S_t$, where q_t is the solution of

$$\mathsf{E}[\hat{H}_t^{DB}(q_t) + \Theta_t^{DB}(q_t) \mid \mathbf{x}_t] = \mathsf{E}[\Pi_t^{DB}(q_t) \mid \mathbf{x}_t], \tag{4.14}$$

and $q_t^{DB} = 0$ otherwise. Note that, since the demands are independent random variables, the information set f_t can be removed, and q_t^{DB} is only a function of the inventory vector $\mathbf{x}_t$.

The q_t in (4.14) balances the expected discounted marginal holding and outdating costs with the expected marginal backlogging cost. It can be verified that the LHS of (4.14) is an increasing convex function of the order quantity q_t, which equals $\alpha^{t-1}h\mathsf{E}[(\sum_{i=1}^{m-1} x_{t,i} - D_t)^+]$ when $q_t = 0$ and approaches infinity when q_t goes to infinity. On the other hand, the RHS of (4.14) is a decreasing convex function of q_t, which equals $\alpha^{t-1}b\mathsf{E}[(D_t - \sum_{i=1}^{m-1} x_{t,i})^+]$ when $q_t = 0$ and approaches 0 when q_t goes to infinity. When $\sum_{i=1}^{m-1} x_{t,i} \leq S_t$, the quantity q_t in (4.14) is well defined with $0 \leq q_t \leq S_t - \sum_{i=1}^{m-1} x_{t,i}$. When $\sum_{i=1}^{m-1} x_{t,i} > S_t$, Equation (4.14) does not have a nonnegative solution and in this case the DB policy orders $q_t^{DB} = 0$.

It is important to note that since S_t is non-decreasing in t, if for some period t we have $\sum_{i=1}^{m-1} x_{t,i} \leq S_t$, then following the DB policy we have $\sum_{i=1}^{m-1} x_{t',i} \leq S_{t'}$ for all $t' \geq t$ regardless

of the demand realizations of $D_t, \ldots, D_T$. This is because when $\sum_{i=1}^{m-1} x_{t,i} \leq S_t$, we have

$$\sum_{i=1}^{m-1} x_{t+1,i} = \sum_{i=1}^{m-1} x_{t,i} + q_t - D_t - e_t \leq S_t - D_t - e_t \leq S_t \leq S_{t+1}.$$

This implies that when the demands are independent and stochastically non-decreasing, the DB policy can perfectly balance the expected marginal holding and outdating costs with the expected marginal backlogging cost after placing its first order. If the demands are not independent or stochastically non-decreasing, then S_t will not necessarily be monotonically non-decreasing in t and as a result, the DB policy will not have the above property. This is the reason why we need to assume that the demands are independent and stochastically non-decreasing over time.

The following result shows that, again, the desired properties exhibited by the optimal control policy for the perishable inventory system are inherited by the DB policy. Its proof is very similar to that of Proposition 4.2 and hence it is omitted.

Proposition 4.4. *For each period t, the order quantity q_t^{DB} under the DB policy satisfies*

$$-1 \leq \frac{\partial q_t^{DB}}{\partial x_{m-1}} \leq \frac{\partial q_t^{DB}}{\partial x_{m-2}} \leq \cdots \leq \frac{\partial q_t^{DB}}{\partial x_1} \leq 0.$$

The following theorem shows that, the DB policy has a worst-case performance guarantee of 2 when the demands are independent and stochastically non-decreasing over time.

Theorem 4.5. *For an arbitrary independent and stochastically non-decreasing demand process, the dual-balancing policy for the perishable inventory system with an arbitrary fixed product lifetime has a worst-case performance guarantee of 2, i.e., for any instance of the problem, the expected cost of the dual-balancing policy is at most twice the expected cost of an optimal policy.*

4.5 Worst-Case Analysis

The arguments used in the literature on proving worst-case performance guarantees for approximation algorithms utilize a "unit-matching" approach (see, e.g., Levi et al. (2007a, 2008a,b, 2012), Levi and Shi (2013)). In a sense, the approach is geometric, and it relies on the correspondence of units in the systems operating under different policies throughout the planning horizon, and then it compares the costs incurred by the matched units in different systems. However, the unit-matching approach fails to work for perishable inventory systems

because the inventory units can perish and the number of outdating units differs in systems operating under different policies.

To overcome this difficulty, we develop an algebraic approach for comparing different systems. A key concept in our approach is the *trimmed on-hand inventory level*, which is defined as the part of on-hand inventory units ordered before any given particular time. These trimmed inventory levels serve as a generalization of the traditional inventory level, as they provide critical (partial) information on the ages of the products on-hand. Due to the nature of perishable systems, it is impossible to quantify the effect of the decision made in the current period t on future costs only through the traditional total inventory level Y_t. The trimmed inventory levels provide a tractable way to analyze this effect, and also provide the *right* framework for coupling the marginal holding and outdating costs in different systems. More technically, the difference between the trimmed inventory levels of our policy and the optimal policy OPT can be bounded by the difference between the outdating units of the two policies. An essential part of this worst-case analysis presented below is based on this new concept.

The main ideas and arguments for the proofs of our key results are given below. We leave some of the more involved technical analysis in the online Appendix. For simplicity, whenever possible we will abbreviate the marginal costs $H_t^P(q_t)$, $\Theta_t^P(q_t)$ and $\Pi_t^P(q_t)$ by H_t^P, Θ_t^P and Π_t^P, respectively, i.e., we make the ordering quantity q_t in these functions implicit.

In the following, we first study the PB policy and its worst-case performance, and then study the DB policy.

4.5.1 Analysis of PB Policy

We now compare the PB policy with the optimal policy OPT. To this end, we make the dependency of the relevant quantities on the policy, PB or OPT, explicit. For each realization of demands $D_1, \ldots, D_T$ and the exogenous information $W_1, \ldots, W_T$, we compare and analyze the inventory processes of the systems operating under these two policies.

Given a realization $f_T \in \mathscr{F}_T$, let $\mathscr{T}_H$ be the set of periods in which the optimal policy has more total inventory level than the PB policy does. In other words, we denote

$$\mathscr{T}_H = \left\{ t \in [1,T] : Y_t^{OPT} \geq Y_t^{PB} \right\}.$$

In addition, we let its complement set be denoted by

$$\mathscr{T}_\Pi = \left\{ t \in [1,T] : Y_t^{OPT} < Y_t^{PB} \right\}.$$

Lemma 4.6. *For each realization $f_T \in \mathscr{F}_T$, we have*

$$\sum_{t\in\mathscr{T}_H} H_t^{PB} \le \sum_{t=1}^{T} H_t^{OPT} + (m-2)\frac{h}{\theta}\sum_{t=1}^{T}\Theta_t^{OPT}.$$

Proof. For brevity, we only prove the result for the case when the discount factor $\alpha = 1$. The general case with $\alpha \in [0,1]$ can be proved similarly. For any $t = 1,\dots,T$, denote e_t as the amount of perished products in period t. Since the lifetime of the products is m, we have $\Theta_t = \theta e_{t+m-1}$. In addition, for any t and $s \ge t \ge 0$, denote $Y_{t,s}$ as the part of on-hand inventory at the beginning of period s which is ordered in period t or earlier. Note that $Y_{t,s} - Y_{0,s}$ is the part of on-hand inventory at the beginning of period s which is ordered in between periods 1 and t (in the case of 0 initial inventory level, $Y_{0,s} \equiv 0$). Thus, $Y_{t,s} \ge Y_{0,s}$. For convenience, we denote $D_0 = e_0 = 0$. From the definition of $Y_{t,s}$, it is readily verified that

$$Y_{t,s} = (Y_t - D_{[t,s)} - e_{[t,s)})^+, \quad s = t, t+1, \dots, t+m-1, \tag{4.15}$$

and $Y_{t,s} = 0$ when $s \ge t+m$.

For any period $t = 1,\dots,T$, we define the notation $R(t)$ as follows: if the set $\{s \in \mathscr{T}_H : s > t\}$ is not empty, then $R(t) := \min\{s \in \mathscr{T}_H : s > t\}$; otherwise, $R(t) := T+1$. In addition, for any $s \ge 1$, denote $\tilde{H}_s$ as the part of holding cost incurred in period s which is associated with the products ordered in periods $\{t : t \in \mathscr{T}_H,\ t \le s\}$. Since the lifetime of the products is m, all products ordered in period t or earlier will leave the system at the end of period $t+m-1$. Then, it follows that for any $t \in \mathscr{T}_H$, $\tilde{H}_s = 0$ when $t+m \le s \le R(t)-1$. Consequently, by the definitions of H_t, $\tilde{H}_s$, and $R(t)$, we have

$$\sum_{t\in\mathscr{T}_H} H_t = \sum_{t\in\mathscr{T}_H}\sum_{s=t}^{R(t)-1}\tilde{H}_s = \sum_{t\in\mathscr{T}_H}\sum_{s=t}^{(t+m)\wedge R(t)-1}\tilde{H}_s. \tag{4.16}$$

For each period $s \in [t, R(t)-1]$, the amount of leftover inventories (after satisfying the demand D_s but before product outdating) associated with the orders in periods $1,\dots,t$ can be expressed as $(Y_{t,s} - Y_{0,s} - (D_s - Y_{0,s})^+)^+$, which also equals $(Y_{t,s} - D_s)^+ - (Y_{0,s} - D_s)^+$. It is clear that these leftover inventories consist of two parts: 1) the leftover inventories associated with the orders in periods $\{t : t \in \mathscr{T}_H,\ t \le s\}$; and 2) the leftover inventories associated with the orders in periods $\{t : t \in \mathscr{T}_{\mathrm{II}},\ t \le s\}$. Note that the outdating products during periods $s,\dots,t+m-1$ only come from the above leftover inventories. Thus, the

total outdating products in periods $\{t' : t' = s, \ldots, t+m-1,\ t'-m+1 \in \mathscr{T}_\Pi\}$ are less than or equal to the leftover inventories associated with the orders in periods $\{t : t \in \mathscr{T}_\Pi,\ t \le s\}$. For convenience, for $t = m, \ldots, T$, we denote

$$e_t^\Pi = \begin{cases} e_t, & \text{if } t-m+1 \in \mathscr{T}_\Pi; \\ 0, & \text{otherwise.} \end{cases}$$

In addition, we denote $e_t^H = e_t - e_t^\Pi$. Then, we have

$$\tilde{H}_s \le h(Y_{t,s} - D_s)^+ - h(Y_{0,s} - D_s)^+ - he^\Pi_{[s,t+m-1]}. \tag{4.17}$$

Combining (4.16) and (4.17), we obtain

$$\sum_{t\in\mathscr{T}_H} H_t \le h \sum_{t\in\mathscr{T}_H} \sum_{s=t}^{(t+m)\wedge R(t)-1} \left((Y_{t,s} - D_s)^+ - (Y_{0,s} - D_s)^+ - e^\Pi_{[s,t+m-1]}\right). \tag{4.18}$$

On the other hand, for any $t < s$, since $Y_s \ge Y_{t,s} + q_s$, it is clear that $(Y_s - D_s)^+ \ge (Y_{t,s} - D_s)^+ + e_{s+m-1}$. Therefore, we have

$$\begin{aligned} \sum_{t=1}^T H_t =& h\sum_{t=1}^T ((Y_t - D_t)^+ - (Y_{0,t} - D_t)^+) \\ \ge& h \sum_{t\in\mathscr{T}_H} \sum_{s=t}^{R(t)-1} \left((Y_{t,s} - D_s)^+ - (Y_{0,s} - D_s)^+\right) + h\sum_{s\in\mathscr{T}_\Pi} e_{s+m-1} \\ \ge& h \sum_{t\in\mathscr{T}_H} \sum_{s=t}^{(t+m)\wedge R(t)-1} \left((Y_{t,s} - D_s)^+ - (Y_{0,s} - D_s)^+\right) + h\sum_{t=m}^T e_t^\Pi. \end{aligned} \tag{4.19}$$

Note that $Y_{0,s}$ is the part of on-hand inventory at the beginning of period s which is ordered before period 1, it follows from the FIFO issuing policy that $Y_{0,s}$ is independent of the ordering policy. Hence,

$$\sum_{t\in\mathscr{T}_H} \sum_{s=t}^{(t+m)\wedge R(t)-1} (Y_{0,s}^{PB} - D_s)^+ = \sum_{t\in\mathscr{T}_H} \sum_{s=t}^{(t+m)\wedge R(t)-1} (Y_{0,s}^{OPT} - D_s)^+. \tag{4.20}$$

Combining (4.18), (4.19), and (4.20), we obtain

$$\begin{aligned}\sum_{t\in\mathscr{T}_H} H_t^{PB} - \sum_{t=1}^{T} H_t^{OPT} \leq & h\sum_{t\in\mathscr{T}_H}\sum_{s=t}^{(t+m)\wedge R(t)-1}\left((Y_{t,s}^{PB}-D_s)^+ - (Y_{t,s}^{OPT}-D_s)^+ - e_{[s,t+m-1]}^{PB,\Pi}\right)\\ & - h\sum_{t=m}^{T} e_t^{OPT,\Pi}\\ \leq & h\sum_{t\in\mathscr{T}_H}\left((Y_t^{PB}-D_t)^+ - (Y_t^{OPT}-D_t)^+ - e_{[t,t+m-1]}^{PB,\Pi}\right) - h\sum_{t=m}^{T} e_t^{OPT,\Pi}\\ & + h\sum_{t\in\mathscr{T}_H}\sum_{s=t+1}^{(t+m)\wedge R(t)-1}(e_{[t,s)}^{OPT} - e_{[t,s)}^{PB})^+, \end{aligned} \tag{4.21}$$

where the second inequality holds because

$$(Y_{t,s}^{PB}-D_s)^+-(Y_{t,s}^{OPT}-D_s)^+ \leq (Y_{t,s}^{PB}-Y_{t,s}^{OPT})^+ = (Y_t^{PB}-Y_t^{OPT}-e_{[t,s)}^{PB}+e_{[t,s)}^{OPT})^+ \leq (e_{[t,s)}^{OPT}-e_{[t,s)}^{PB})^+,$$

where the equality is from (4.15) and the second inequality holds since $Y_t^{PB} \leq Y_t^{OPT}$ when $t \in \mathscr{T}_H$.

For convenience, we denote $\hat{e}_t$ as follows: when $1 \leq t \leq m-1$, $\hat{e}_t := 0$; and when $m \leq t \leq T$, $\hat{e}_t := e_t$. Since inventories are consumed in a first-in-first-out manner, it is seen that $e_t^{PB} = e_t^{OPT}$ for $1 \leq t \leq m-1$ for any realization of the demands. Then, for $1 \leq t \leq T$, $e_t^{OPT} - e_t^{PB} = \hat{e}_t^{OPT} - \hat{e}_t^{PB}$. Thus, we have

$$\begin{aligned}&\sum_{t\in\mathscr{T}_H}\sum_{s=t+1}^{(t+m)\wedge R(t)-1}(e_{[t,s)}^{OPT} - e_{[t,s)}^{PB})^+ = \sum_{t\in\mathscr{T}_H}\sum_{s=t+1}^{(t+m)\wedge R(t)-1}(\hat{e}_{[t,s)}^{OPT} - \hat{e}_{[t,s)}^{PB})^+\\ \leq & \sum_{\substack{t\in\mathscr{T}_H,\ t\geq m\\ R(t)\geq t+m}}(e_t^{OPT}-e_t^{PB})^+ + \sum_{\substack{t\in\mathscr{T}_H\\ R(t)\geq t+m}}\sum_{s=t+2}^{t+m-1}\hat{e}_{[t,s)}^{OPT} + \sum_{\substack{t\in\mathscr{T}_H\\ R(t)\leq t+m-1}}\sum_{s=t+1}^{R(t)-1}\hat{e}_{[t,s)}^{OPT}\\ \leq & \sum_{\substack{t\in\mathscr{T}_H,\ t\geq m\\ R(t)\geq t+m}}(e_t^{OPT}-e_t^{PB})^+ + (m-2)\sum_{\substack{t\in\mathscr{T}_H\\ R(t)\geq t+m}}\hat{e}_{[t,R(t))}^{OPT} + (m-3)\sum_{\substack{t\in\mathscr{T}_H\\ R(t)\leq t+m-1}}\hat{e}_{[t,R(t))}^{OPT} + \sum_{\substack{t\in\mathscr{T}_H\\ R(t)=t+m-1}}\hat{e}_t^{OPT}\\ = & \sum_{\substack{t\in\mathscr{T}_H,\ t\geq m\\ R(t)\geq t+m}}(e_t^{OPT}-e_t^{PB})^+ + (m-2)\sum_{t\in\mathscr{T}_H}\hat{e}_{[t,R(t))}^{OPT} - \sum_{\substack{t\in\mathscr{T}_H\\ R(t)\leq t+m-2}}\hat{e}_{[t,R(t))}^{OPT} - \sum_{\substack{t\in\mathscr{T}_H\\ R(t)=t+m-1}}\hat{e}_{[(t+1),R(t))}^{OPT}. \end{aligned} \tag{4.22}$$

Note that

$$\sum_{t\in\mathscr{T}_H}\hat{e}^{OPT}_{[t,R(t))}\le\sum_{t=1}^{T}\hat{e}^{OPT}_t=\sum_{t=m}^{T}e^{OPT}_t=\frac{1}{\theta}\sum_{t=1}^{T}\Theta^{OPT}_t. \tag{4.23}$$

Hence, according to (4.21)-(4.23), to prove the lemma, it suffices to show

$$\begin{aligned}\sum_{\substack{t\in\mathscr{T}_H,\ t\ge m\\ R(t)\ge t+m}}(e^{OPT}_t-e^{PB}_t)^+\le&\sum_{t\in\mathscr{T}_H}\left((Y^{OPT}_t-D_t)^+-(Y^{PB}_t-D_t)^++e^{PB,\Pi}_{[t,t+m-1]}\right)+\sum_{t=m}^{T}e^{OPT,\Pi}_t\\&+\sum_{\substack{t\in\mathscr{T}_H\\ R(t)\le t+m-2}}\hat{e}^{OPT}_{[t,R(t))}+\sum_{\substack{t\in\mathscr{T}_H\\ R(t)=t+m-1}}\hat{e}^{OPT}_{[t+1,R(t))}.\end{aligned} \tag{4.24}$$

Denote $\hat{\mathscr{T}}_H=\{t:t\in\mathscr{T}_H,R(t)\ge t+m\}$, and define $\hat{L}(t)$ as follows: if the set $\{s\in\hat{\mathscr{T}}_H:s<t\}$ is not empty, then $\hat{L}(t):=\max\{s\in\hat{\mathscr{T}}_H:s<t\}$; otherwise, $\hat{L}(t):=0$. Then, to prove (4.24), it suffices to show that, for any $s\ge m$ and $s\in\hat{\mathscr{T}}_H$,

$$\begin{aligned}(e^{OPT}_s-e^{PB}_s)^+\le&\sum_{t\in\mathscr{T}_H,\hat{L}(s)<t\le s}\left((Y^{OPT}_t-D_t)^+-(Y^{PB}_t-D_t)^++e^{PB,\Pi}_{[t,t+m-1]}\right)+\sum_{t=\hat{L}(s)\vee m}^{s}e^{OPT,\Pi}_t\\&+\sum_{\substack{t\in\mathscr{T}_H,\hat{L}(s)<t\le s\\ R(t)\le t+m-2}}\hat{e}^{OPT}_{[t,R(t))}+\sum_{\substack{t\in\mathscr{T}_H,\hat{L}(s)<t\le s\\ R(t)=t+m-1}}\hat{e}^{OPT}_{[t+1,R(t))}.\end{aligned} \tag{4.25}$$

Furthermore, to prove (4.25), it is sufficient to show that, for any $s\ge m$,

$$\begin{aligned}(e^{OPT}_s-e^{PB}_s)^+\le&1_{\{s-m+1\in\mathscr{T}_H\}}\left((Y^{OPT}_{s-m+1}-D_{s-m+1})^+-(Y^{PB}_{s-m+1}-D_{s-m+1})^++e^{PB,\Pi}_{[s-m+1,s]}\right)\\&+e^{OPT,\Pi}_s+1_{\{R(s-2m+2)<s-m+1\in\mathscr{T}_H\}}\hat{e}^{OPT}_{[R(s-2m+1),s-m+1)}\\&+1_{\{s-2m+2\in\mathscr{T}_H,R(s-2m+2)=s-m+1\}}\left((e^{OPT}_{s-2m+2}-e^{PB}_{s-2m+2})^++\hat{e}^{OPT}_{[s-2m+3,s-m+1)}\right).\end{aligned} \tag{4.26}$$

In what follows, we prove that (4.26) is indeed true for any $s\ge m$ and $s\in\hat{\mathscr{T}}_H$, which then completes the proof of Lemma 4.6.

Note that $(e^{OPT}_s-e^{PB}_s)^+=0$ when $e^{OPT}_s\le e^{PB}_s$, and $(e^{OPT}_s-e^{PB}_s)^+\le e^{OPT,\Pi}_s$ when $s-m+1\in\mathscr{T}_\Pi$. (4.26) is obviously true in both cases. In the following, we assume $e^{OPT}_s>e^{PB}_s$ and $s-m+1\in\mathscr{T}_H$.

From the definition of Y_s and e_s, the following identity holds for any policy:

$$e_s = \left((Y_{s-m+1} - D_{s-m+1})^+ - D_{[s-m+2,t]} - e_{[s-m+1,s-1]}\right)^+. \tag{4.27}$$

Since $e_t^{OPT} > e_t^{PB}$, with some simple algebra, it follows from the above identity that

$$\begin{aligned} e_s^{OPT} - e_s^{PB} &\le (Y_{s-m+1}^{OPT} - D_{s-m+1})^+ - (Y_{s-m+1}^{PB} - D_{s-m+1})^+ + e_{[s-m+1,s-1]}^{PB} - e_{[s-m+1,s-1]}^{OPT} \\ &\le (Y_{s-m+1}^{OPT} - D_{s-m+1})^+ - (Y_{s-m+1}^{PB} - D_{s-m+1})^+ + e_{[s-m+1,t]}^{PB,\Pi} + e_{[s-m+1,s-1]}^{PB,H} - e_{[s-m+1,s-1]}^{OPT}. \end{aligned}$$

If $e_{[s-m+1,s-1]}^{PB,H} \le e_{[s-m+1,s-1]}^{OPT}$, then (4.26) is proved. Now suppose $e_{[s-m+1,s-1]}^{PB,H} > e_{[s-m+1,s-1]}^{OPT}$. In this case, there must exist a period $w(s) \in [s-m+1, s-1]$ such that $e_{w(s)}^{PB,H} > e_{w(s)}^{OPT}$, and $e_{(w(s),s-1]}^{PB,H} \le e_{(w(s),s-1]}^{OPT}$. From the definition of $e_{w(s)}^{PB,H}$, we also have $w(s) - m + 1 \in \mathscr{T}_H$. Now applying the identity (4.27) and the fact that $Y_{w(s)-m+1}^{OPT} \ge Y_{w(s)-m+1}^{PB}$, we obtain

$$\begin{aligned} e_{[s-m+1,t-1]}^{PB,H} - e_{[s-m+1,t-1]}^{OPT} \le& e_{[s-m+1,w(s)]}^{PB} - e_{[s-m+1,w(s)]}^{OPT} \le e_{[w(s)-m+1,s-m+1)}^{OPT} - e_{[w(s)-m+1,t+m+1)}^{PB} \\ =& \hat{e}_{[w(s)-m+1,s-m+1)}^{OPT} - \hat{e}_{[w(s)-m+1,t+m+1)}^{PB}. \end{aligned} \tag{4.28}$$

Since both $w(s) - m + 1$ and $s - m + 1$ belong to $\mathscr{T}_H$, $R(s - 2m + 1) \le w(s) - m + 1$, and $s - w(s) \le m - 1$, it can be verified that

$$\begin{aligned} \hat{e}_{[w(s)-m+1,s-m+1)}^{OPT} - \hat{e}_{[w(s)-m+1,s+m+1)}^{PB} \le& 1_{\{R(s-2m+2)<s-m+1\in\mathscr{T}_H\}} \hat{e}_{[R(s-2m+1),s-m+1)}^{OPT} \\ &+ 1_{\{s-2m+2\in\mathscr{T}_H, R(s-2m+2)=s-m+1\}} \left((e_{s-2m+2}^{OPT} - e_{s-2m+2}^{PB})^+ + \hat{e}_{[s-2m+3,s-m+1)}^{OPT}\right). \end{aligned} \tag{4.29}$$

Hence, (4.26) is proved, and the proof of Lemma 4.6 is complete. □

Lemma 4.7. *For each realization $f_T \in \mathscr{F}_T$, we have $\sum_{t\in\mathscr{T}_H} \Theta_t^{PB} \le \sum_{t=1}^T \Theta_t^{OPT}$.*

Proof. We first establish a preliminary result: For any period $t \in \mathscr{T}_H$, if $\Theta_t^{PB} > \Theta_t^{OPT}$, then there exists a period w_t such that $(t - m + 1) \vee 1 \le w_t < t$ and

$$\sum_{s=w_t}^{t} \Theta_s^{PB} \le \sum_{s=w_t}^{t} \Theta_s^{OPT}. \tag{4.30}$$

Suppose $t \in \mathscr{T}_H$ and $\Theta_t^{PB} > \Theta_t^{OPT}$. Since $\Theta_t = \theta\alpha^{t+m-2} e_{t+m-1}$, we have $e_{t+m-1}^{PB} > e_{t+m-1}^{OPT} \ge 0$. Recall that e_t is the number of perished products in period t, and it satisfies

the following identity under any policy:

$$e_{t+m-1} = (Y_t - D_{[t,t+m-1]} - e_{[t,t+m-1)})^+.$$

Thus it follows from $e^{PB}_{t+m-1} > 0$ that

$$\begin{aligned} e^{PB}_{[t,t+m-1]} &= \left(Y^{PB}_t - D_{[t,t+m-1]} - e^{PB}_{[t,t+m-1)}\right)^+ + e^{PB}_{[t,t+m-1)} \\ &= Y^{PB}_t - D_{[t,t+m-1]}. \end{aligned} \tag{4.31}$$

On the other hand, for the OPT policy we have

$$\begin{aligned} e^{OPT}_{[t,t+m-1]} &= \left(Y^{OPT}_t - D_{[t,t+m-1]} - e^{OPT}_{[t,t+m-1)}\right)^+ + e^{OPT}_{[t,t+m-1)} \\ &\geq Y^{OPT}_t - D_{[t,t+m-1]}, \end{aligned} \tag{4.32}$$

Subtracting (4.32) from (4.31) yields

$$e^{PB}_{[t\vee m,t+m-1]} - e^{OPT}_{[t\vee m,t+m-1]} = e^{PB}_{[t,t+m-1]} - e^{OPT}_{[t,t+m-1]} \leq Y^{PB}_t - Y^{OPT}_t \leq 0, \tag{4.33}$$

where the equality holds since $e^{PB}_t = e^{OPT}_t$ for $1 \leq t \leq m-1$ and the last inequality follows from $t \in \mathscr{T}_H$. This proves $e^{PB}_{[t\vee m,t+m-1]} < e^{OPT}_{[t\vee m,t+m-1]}$.

We argue that (4.33) implies the existence of $w_t \in [(t-m+1)\vee 1, t)$ that satisfies (4.30). To this end, we apply Abel's lemma (see e.g., Chow and Teicher (2012)) and the identity $\Theta_s = \theta\alpha^{s+m-2}e_{s+m-1}$ to obtain, under any policy, that

$$\begin{aligned} \theta e_{[t\vee m,t+m-1]} &= \sum_{s=(t-m+1)\vee 1}^{t} \alpha^{-s-m+2}\Theta_s \\ &= \alpha^{-t\vee m+1}\left[\sum_{s=(t-m+1)\vee 1}^{t}\Theta_s\right] + \sum_{t'=(t-m+1)\vee 1+1}^{t}\alpha^{-t'-m+2}(1-\alpha)\left[\sum_{s=t'}^{t}\Theta_s\right]. \end{aligned}$$

Therefore, $e^{PB}_{[t\vee m,t+m-1]} < e^{OPT}_{[t\vee m,t+m-1]}$ together with the condition that $\Theta^{PB}_t > \Theta^{OPT}_t$ imply the existence of at least one $w_t \in [(t-m+1)\vee 1, t)$, such that (4.30) holds.

We are now ready to prove Lemma 4.7. Since any products ordered after period $T-m+1$ do not perish within the planning horizon, we only need to consider periods $t = 1, \ldots, T-m+1$. For each realization of f_T and the resulting $\mathscr{T}_H$, we partition the periods $\{1, \ldots, T-m+1\}$ as follows: First, start in period $T-m+1$ and search backward for the latest period $t \in \mathscr{T}_H$ such that $\Theta^{PB}_t > \Theta^{OPT}_t$. If no such period exists, then we terminate the partition

process. Otherwise, let t' be that period, and by the preliminary result, there exists a $w_{t'} \in [(t'-m+1) \vee 1, t')$ that satisfies $\sum_{s=w_{t'}}^{t'} \Theta_s^{PB} \leq \sum_{s=w_{t'}}^{t'} \Theta_s^{OPT}$. Mark the periods $w_{t'}, w_{t'}+1, \ldots, t'$. Next, repeat the above procedure over periods $1, \ldots, w_{t'}-1$ until the remaining set of periods is empty. As a result, the above procedure partitions the periods $\{1, \ldots, T-m+1\}$ into marked and unmarked periods. Let $\mathscr{T}_M$ denote the set of marked periods.

Consider any period $t \in \mathscr{T}_H \setminus \mathscr{T}_M$. By the definition of $\mathscr{T}_M$, we have $\Theta_t^{PB} \leq \Theta_t^{OPT}$. Thus,

$$\sum_{t \in \mathscr{T}_H \setminus \mathscr{T}_M} \Theta_t^{PB} \leq \sum_{t \in \mathscr{T}_H \setminus \mathscr{T}_M} \Theta_t^{OPT}. \tag{4.34}$$

On the other hand, if a period $t \in \mathscr{T}_H$ is also in $t \in \mathscr{T}_M$, then by the construction of $\mathscr{T}_M$ it belongs to an interval of the form $[w_{t'}, \ldots, t']$, and the preliminary result above has shown $\sum_{s=w_{t'}}^{t'} \Theta_s^{PB} \leq \sum_{s=w_{t'}}^{t'} \Theta_s^{OPT}$. This proves

$$\sum_{t \in \mathscr{T}_M} \Theta_t^{PB} \leq \sum_{t \in \mathscr{T}_M} \Theta_t^{OPT}. \tag{4.35}$$

Since

$$\mathscr{T}_H \subset (\mathscr{T}_H \setminus \mathscr{T}_M) \cup \mathscr{T}_M \subset \{1, 2, \ldots, T\},$$

we obtain, using (4.34) and (4.35), that

$$\sum_{t \in \mathscr{T}_H} \Theta_t^{PB} \leq \sum_{t \in \mathscr{T}_H \setminus \mathscr{T}_M} \Theta_t^{PB} + \sum_{t \in \mathscr{T}_M} \Theta_t^{PB} \leq \sum_{t \in \mathscr{T}_H \setminus \mathscr{T}_M} \Theta_t^{OPT} + \sum_{t \in \mathscr{T}_M} \Theta_t^{OPT} \leq \sum_{t=1}^{T} \Theta_t^{OPT}.$$

This completes the proof of Lemma 4.7. □

Note that for each perished unit ordered in periods $1, \ldots, T$, it must stay in the system for exactly m periods. Thus, for any policy, we have the following inequality

$$mh \sum_{t=1}^{T} \Theta_t \leq \theta \sum_{t=1}^{T} H_t. \tag{4.36}$$

Combining this inequality with Lemmas 4.6 and 4.7 leads to the following result.

Corollary 4.8. *For each realization $f_T \in \mathscr{F}_T$, we have*

$$\sum_{t \in \mathscr{T}_H} \left(H_t^{PB} + \Theta_t^{PB}\right) \leq \left(1 + \tfrac{(m-2)h}{mh+\theta}\right) \sum_{t=1}^{T} \left(H_t^{OPT} + \Theta_t^{OPT}\right). \tag{4.37}$$

Proof. We apply Lemmas 4.6 and 4.7, and (4.36) to obtain

$$\begin{aligned}\sum_{t\in\mathscr{T}_H}\left(H_t^{PB}+\Theta_t^{PB}\right) &\le \sum_{t=1}^{T} H_t^{OPT}+\frac{(m-2)h}{\theta}\sum_{t=1}^{T}\Theta_t^{OPT}+\sum_{t=1}^{T}\Theta_t^{OPT}\\ &= \sum_{t=1}^{T}\left(H_t^{OPT}+\Theta_t^{OPT}\right)+\frac{(m-2)h}{mh+\theta}\left(1+\frac{mh}{\theta}\right)\sum_{t=1}^{T}\Theta_t^{OPT}\\ &\le \left(1+\frac{(m-2)h}{mh+\theta}\right)\sum_{t=1}^{T}\left(H_t^{OPT}+\Theta_t^{OPT}\right),\end{aligned}$$

□

Lemma 4.9. *For each realization $f_T \in \mathscr{F}_T$, we have $\sum_{t\in\mathscr{T}_\Pi}\Pi_t^{PB} \le \sum_{t=1}^{T}\Pi_t^{OPT}$.*

Proof. From the definition of Π_t and $\mathscr{T}_\Pi$, we have

$$\sum_{t\in\mathscr{T}_\Pi}\Pi_t^{PB} = b\sum_{t\in\mathscr{T}_\Pi}\alpha^{t-1}(D_t-Y_t^{PB})^+ \le b\sum_{t\in\mathscr{T}_\Pi}\alpha^{t-1}(D_t-Y_t^{OPT})^+ \le \sum_{t=1}^{T}\Pi_t^{OPT},$$

where the first inequality holds since $Y_t^{OPT} < Y_t^{PB}$ when $t \in \mathscr{T}_\Pi$. □

With the preparations above, we are now ready to prove our first main result in this chapter, i.e., Theorem 4.3.

Proof of Theorem 4.3. For each period $t = 1, \ldots, T$, denote Z_t^{PB} as the conditional expected *balanced cost* by the PB policy in period t. That is,

$$Z_t^{PB} = \tfrac{mh+\theta}{2(m-1)h+\theta}\mathsf{E}[H_t^{PB}+\Theta_t^{PB}|F_t] = \mathsf{E}[\Pi_t^{PB}|F_t].$$

Note that Z_t^{PB} is a random variable before period t; and in period t, $F_t = f_t$ is realized and its value is the expected balanced cost conditional on the observed information set f_t. Using the marginal cost accounting scheme and a standard argument of conditional expectations, we have

$$\begin{aligned}C(PB) &= \sum_{t=1}^{T}\mathsf{E}[H_t^{PB}+\Theta_t^{PB}+\Pi_t^{PB}] = \sum_{t=1}^{T}\mathsf{E}[\mathsf{E}[H_t^{PB}+\Theta_t^{PB}+\Pi_t^{PB}|F_t]]\\ &= \left(2+\tfrac{(m-2)h}{mh+\theta}\right)\sum_{t=1}^{T}\mathsf{E}[Z_t^{PB}]. \qquad (4.38)\end{aligned}$$

Applying Corollary 4.8, Lemma 4.9, and the fact that $\{t \in \mathscr{T}_H\}$ and $\{t \in \mathscr{T}_\Pi\}$ are

completely determined by F_t, we obtain

$$\begin{aligned}
C(OPT) &= \mathsf{E}\Big[\sum_{t=1}^{T}(H_t^{OPT}+\Theta_t^{OPT})+\sum_{t=1}^{T}\Pi_t^{OPT}\Big] \\
&\geq \mathsf{E}\left[\frac{1}{1+\frac{(m-2)h}{mh+\theta}}\sum_{t\in\mathscr{T}_H}(H_t^{PB}+\Theta_t^{PB})+\sum_{t\in\mathscr{T}_\Pi}\Pi_t^{PB}\right] \\
&= \sum_{t=1}^{T}\mathsf{E}\left[\frac{1}{1+\frac{(m-2)h}{mh+\theta}}\mathbf{1}(t\in\mathscr{T}_H)(H_t^{PB}+\Theta_t^{PB})+\mathbf{1}(t\in\mathscr{T}_\Pi)\Pi_t^{PB}\right] \\
&= \sum_{t=1}^{T}\mathsf{E}\left[\mathsf{E}\left[\frac{1}{1+\frac{(m-2)h}{mh+\theta}}\mathbf{1}(t\in\mathscr{T}_H)(H_t^{PB}+\Theta_t^{PB})+\mathbf{1}(t\in\mathscr{T}_\Pi)\Pi_t^{PB}|F_t\right]\right] \\
&= \sum_{t=1}^{T}\mathsf{E}\left[\frac{1}{1+\frac{(m-2)h}{mh+\theta}}\mathbf{1}(t\in\mathscr{T}_H)\mathsf{E}\big[H_t^{PB}+\Theta_t^{PB}\mid F_t\big]+\mathbf{1}(t\in\mathscr{T}_\Pi)\mathsf{E}\big[\Pi_t^{PB}|F_t\big]\right] \\
&= \sum_{t=1}^{T}\mathsf{E}\big[(\mathbf{1}(t\in\mathscr{T}_H)+\mathbf{1}(t\in\mathscr{T}_\Pi))Z_t^{PB}\big]=\sum_{t=1}^{T}\mathsf{E}[Z_t^{PB}].
\end{aligned}$$

Thus, it follows from (4.38) that $C(PB)\leq\left(2+\frac{(m-2)h}{mh+\theta}\right)C(OPT)$. The proof is complete. □

4.5.2 Analysis of DB Policy

We next analyze the DB policy for the case of independent and stochastically non-decreasing demand processes. Similar to the analysis for the PB policy, for each realization of $D_1,\ldots,D_T$ and $W_1,\ldots,W_T$, we denote $\mathscr{T}_H=\big\{t\in[1,T]:Y_t^{OPT}\geq Y_t^{DB}\big\}$ and $\mathscr{T}_\Pi=\big\{t\in[1,T]:Y_t^{OPT}<Y_t^{DB}\big\}$. Then, we have the following result.

Lemma 4.10. *For each realization $f_T\in\mathscr{F}_T$, we have*

$$\sum_{t\in\mathscr{T}_H}\left(\hat{H}_t^{DB}+\Theta_t^{DB}\right)+\sum_{t\in\mathscr{T}_\Pi}\Pi_t^{DB}\leq\sum_{t=1}^{T}\left(\hat{H}_t^{OPT}+\Theta_t^{OPT}+\Pi_t^{OPT}\right).$$

Proof. By the definition of $\hat{H}_t$ and $\mathscr{T}_H$, we have

$$\sum_{t\in\mathscr{T}_H}\hat{H}_t^{DB}=h\sum_{t\in\mathscr{T}_H}\alpha^{t-1}(Y_t^{DB}-D_t)^+\leq h\sum_{t\in\mathscr{T}_H}\alpha^{t-1}(Y_t^{OPT}-D_t)^+\leq\sum_{t=1}^{T}\hat{H}_t^{OPT},$$

where the first inequality follows from $Y_t^{OPT}\geq Y_t^{DB}$ for $t\in\mathscr{T}_H$. Combining the above inequality with Lemmas 4.7 and 4.9, which can be shown to continue to hold under the DB

policy, we obtain the desired result. □

Lemma 4.11. *Suppose $D_1, \ldots, D_T$ are independent and stochastically non-decreasing. For each period t and realization $f_t \in \mathscr{F}_t$, if $t \in \mathscr{T}_\Pi$, then $\mathsf{E}[\hat{H}_t^{DB} + \Theta_t^{DB} | f_t] = \mathsf{E}[\Pi_t^{DB} | f_t]$.*

Proof. From the definition of the DB policy, to prove the lemma, we only need to prove that $\sum_{i=1}^{m-1} x_{t,i}^{DB} \leq S_t$. When the demands are independent and stochastically non-decreasing, based on our discussions in Section 4.2, it suffices to show that there exists one period t' ($\leq t$) such that $\sum_{i=1}^{m-1} x_{t',i}^{DB} \leq S_{t'}$. This is clearly true if $t \in \mathscr{T}_\Pi$ (or equivalently $Y_t^{OPT} < Y_t^{DB}$), because otherwise by definition the DB policy will not place any order in periods $1, \ldots, t$ and consequently we must have $Y_t^{OPT} \geq Y_t^{DB}$, leading to a contradiction. The proof is complete. □

We are now ready to prove our second main result, i.e., Theorem 4.5.

Proof of Theorem 4.5. For any policy P, define $\hat{C}(P) := \sum_{t=1}^{T} \mathsf{E}[\hat{H}_t^P + \Theta_t^P + \Pi_t^P]$. Then, it follows from (4.4) and (4.11) that $\mathscr{C}(P) = \hat{C}(P) + \sum_{t=1}^{T} \alpha^{t-1} \hat{c} \mathsf{E}[D_t]$. Thus, to prove $\mathscr{C}(DB) \leq 2\mathscr{C}(OPT)$, it suffices to show that $\hat{C}(DB) \leq 2\hat{C}(OPT)$. Using the marginal cost accounting scheme and a standard argument of conditional expectations, we have

$$
\begin{aligned}
\hat{C}(DB) =& \sum_{t=1}^{T} \mathsf{E}\left[\hat{H}_t^{DB} + \Theta_t^{DB} + \Pi_t^{DB}\right] \\
=& \sum_{t=1}^{T} \mathsf{E}\left[(\mathbf{1}(t \in \mathscr{T}_H) + \mathbf{1}(t \in \mathscr{T}_\Pi))(\hat{H}_t^{DB} + \Theta_t^{DB} + \Pi_t^{DB})\right] \\
\leq& \hat{C}(OPT) + \sum_{t=1}^{T} \mathsf{E}\left[\mathbf{1}(t \in \mathscr{T}_\Pi)(\hat{H}_t^{DB} + \Theta_t^{DB}) + \mathbf{1}(t \in \mathscr{T}_H)\Pi_t^{DB}\right] \\
=& \hat{C}(OPT) + \sum_{t=1}^{T} \mathsf{E}\left[\mathsf{E}\left[\mathbf{1}(t \in \mathscr{T}_\Pi)(\hat{H}_t^{DB} + \Theta_t^{DB}) + \mathbf{1}(t \in \mathscr{T}_H)\Pi_t^{DB} | F_t\right]\right] \\
\leq& \hat{C}(OPT) + \sum_{t=1}^{T} \mathsf{E}\left[\mathsf{E}\left[\mathbf{1}(t \in \mathscr{T}_\Pi)(\hat{H}_t^{DB} + \Theta_t^{DB}) + \mathbf{1}(t \in \mathscr{T}_H)(\hat{H}_t^{DB} + \Theta_t^{DB}) | F_t\right]\right] \\
=& \hat{C}(OPT) + \sum_{t=1}^{T} \mathsf{E}\left[\mathsf{E}\left[\mathbf{1}(t \in \mathscr{T}_\Pi)\Pi_t^{DB} + \mathbf{1}(t \in \mathscr{T}_H)(\hat{H}_t^{DB} + \Theta_t^{DB}) | F_t\right]\right] \\
\leq& 2\hat{C}(OPT),
\end{aligned}
$$

where the first and last inequalities follow from Lemma 4.10, the second inequality holds since $\mathsf{E}[\Pi_t^{DB} | F_t] \leq \mathsf{E}[\hat{H}_t^{DB} + \Theta_t^{DB} | F_t]$ for each period t and any realization of F_t under the

DB policy, and the fourth equality follows from Lemma 4.11. The proof of Theorem 4.5 is thus complete.

4.6 Numerical Experiments

To test the empirical performance of our proposed policies, we have conducted an extensive numerical study. The numerical results show that our proposed policies perform consistently close to optimal for a large set of demand and parameter instances.

Parameterized policies. Similar to Levi and Shi (2013), we can slightly improve the performance of the approximation algorithms by employing an instance-dependent balancing parameter if the system parameters and demand process are stationary over time, and the planning horizon is long. The parameterized policies involve a balancing parameter β. Specifically, the parameterized proportional-balancing policy (PPB) computes the balancing quantity q_t^{PPB} that solves $\beta \mathsf{E}[H_t^{PPB}(q_t^{PPB}) + \Theta_t^{PPB}(q_t^{PPB})|f_t] = \mathsf{E}[\Pi_t^{PPB}(q_t^{PPB})|f_t]$. Similarly, the parameterized dual-balancing policy (PDB) computes the balancing quantity q_t^{PDB} that solves $\hat{\beta}\mathsf{E}[\hat{H}_t^{PDB}(q_t^{PDB}) + \Theta_t^{PDB}(q_t^{PDB})] = \mathsf{E}[\Pi_t^{PDB}(q_t^{PDB})]$. In addition to the PB and DB policies, we also report the empirical performance of the PPB and PDB policies when comparing with optimal policies.

Design of experiments. In our numerical experiments, we consider five demand settings (one independent and four correlated demand settings).

(a) Independent and identically distributed (i.i.d.) demands;

(b) ADI demands with two periods of advance demand information;

(c) Autoregressive demands AR(1);

(d) MMFE demands with two periods of forecast evolution;

(e) Markov modulated demands with three states of the economy.

For the i.i.d. demand setting (a), we consider the lifetime $m = 2, 3, 4$ and 6 (see, e.g., Haijema et al. (2005) for blood bank applications where platelet pools are the most expensive and most perishable blood product having a shelf life of four to six days). When $m = 2$ and 3, computing the exact optimal policies using dynamic programming is tractable. Thus, we compare the performance of our proposed policies directly with that of the optimal policies.

In addition, we adopt the same set of numerical parameters as that in Nahmias (1976, 1977b), and also compare the performance of our proposed policies with their heuristics. When $m = 4$ and 6, computing the exact optimal policies becomes intractable (even for this i.i.d. demand case); thus we compare our policies with two other effective policies in Nahmias (1976, 1977b). (The key idea behind these heuristics in Nahmias (1976, 1977b) is to collapse the state space into a single scalar, which has also been used in Li et al. (2009) and Chen et al. (2014b) for other perishable inventory systems.) For correlated demand settings (b) to (e), we are not aware of any heuristic policies in the literature, thus we only consider the lifetime $m = 2$ and 3. Following the numerical studies in the literature on perishable inventory systems, we assume for all testing instances that the system starts empty in period 1, the unit holding cost $\hat{h}$ is normalized to 1, and the discounted factor is $\alpha = 0.95$.

Performance metrics. We use two types of performance metrics in our numerical study. First, when the product lifetime $m = 2$ or 3, we are able to compare the performance of our proposed policies with that of an optimal policy. From (4.11) and (4.12), the cost ratio is

$$\frac{\mathscr{C}(P)}{\mathscr{C}(OPT)} = \frac{C(P) + \sum_{t=1}^{T} \alpha^{t-1} \hat{c} \mathsf{E}[D_t]}{C(OPT) + \sum_{t=1}^{T} \alpha^{t-1} \hat{c} \mathsf{E}[D_t]}.$$

We define the *performance error* of an approximation policy P as the percentage of increase in the total cost of this policy over the planning horizon compared to the optimal total cost, i.e.,

$$\% \text{ error} = \left(\frac{\mathscr{C}(P)}{\mathscr{C}(OPT)} - 1 \right) \times 100\%.$$

Second, when the product lifetime $m = 4$ or 6, since computing the exact optimal solution using dynamic programming is intractable, we compare the performance of our policies against those of Nahmias (1976, 1977b). Denote the heuristic algorithms in Nahmias (1976) and Nahmias (1977b) by N_1 and N_2, respectively. We define the *performance ratio* of an approximation policy P as the ratio of Nahmias' minimum cost to the cost of P, i.e.,

$$r(P) = \frac{\min \{\mathscr{C}(N_1), \mathscr{C}(N_2)\}}{\mathscr{C}(P)}.$$

Demand setting (a). When the product lifetime $m = 2$ and 3, we adopt the same set of parameters as that in Nahmias (1976, 1977b) for our numerical test and directly use the optimal costs reported in those papers. More specifically, the planning horizon is $T = 50$

periods, the ordering cost $\hat{c} \in \{0, 5, 10\}$, the backlogging cost $\hat{b} \in \{5, 10\}$, and the outdating cost $\hat{\theta} \in \{0, 5, 10\}$. Same as Nahmias (1976, 1977b), we also test two demand distributions, i.e., exponential distribution and Erlang-2 distribution, both with mean 10. The numerical results are summarized in Table 4.1.

Table 4.1: Performance errors of heuristics for i.i.d. demands (% Errors) for $m = 2$ and $m = 3$

	$m = 2$								$m = 3$							
	Exponential Demand				Erlang-2 Demand				Exponential Demand				Erlang-2 Demand			
$\hat{c}$, $\hat{b}$, $\hat{\theta}$	PB	PPB	DB	PDB	PB	PPB	DB	PDB	PB	PPB	DB	PDB	PB	PPB	DB	PDB
0,5,5	0.89	0.21	0.82	0.18	0.44	0.35	0.27	0.11	0.95	0.36	0.63	0.27	0.80	0.59	0.65	0.28
0,10,5	0.76	0.36	0.92	0.11	0.63	0.13	0.22	0.21	0.93	0.48	0.74	0.12	1.05	0.82	0.66	0.16
0,5,10	1.37	0.81	1.41	0.42	0.73	0.13	0.59	0.27	1.12	0.88	1.40	0.52	1.63	0.59	0.56	0.47
5,10,5	0.63	0.09	0.96	0.05	0.12	0.07	0.09	0.06	1.06	0.92	1.25	0.15	0.22	0.13	0.57	0.22
5,5,10	0.49	0.36	0.69	0.34	0.35	0.03	0.47	0.28	0.57	0.31	0.62	0.29	0.17	0.12	0.22	0.13
5,5,5	0.32	0.15	0.64	0.13	0.15	0.06	0.11	0.10	0.68	0.38	0.58	0.35	0.16	0.12	0.24	0.11
5,10,0	0.28	0.10	0.18	0.11	0.36	0.11	0.15	0.11	0.58	0.39	0.56	0.21	0.45	0.18	0.89	0.24
10,10,5	0.52	0.07	0.78	0.06	0.27	0.08	0.25	0.18	0.79	0.56	1.28	0.60	0.19	0.11	0.62	0.42
10,10,10	0.80	0.18	1.38	0.21	0.09	0.08	0.24	0.15	0.92	0.74	1.07	0.15	0.06	0.04	0.17	0.04
10,5,5	0.77	0.28	0.92	0.22	0.13	0.11	0.19	0.12	0.74	0.46	0.24	0.05	0.10	0.06	0.26	0.10
10,10,0	0.08	0.06	0.54	0.14	0.07	0.05	0.19	0.03	0.44	0.25	0.47	0.13	0.14	0.09	0.46	0.17
max	1.37	0.81	1.41	0.42	0.73	0.35	0.59	0.28	1.12	0.92	1.40	0.60	1.63	0.82	0.89	0.47
mean	0.63	0.24	0.84	0.18	0.30	0.11	0.25	0.15	0.80	0.52	0.80	0.26	0.45	0.26	0.48	0.21

The empirical performance error of the DB policy does not exceed 1.41% in all test cases, with an average error of 0.84% (resp., 0.80%) under exponential demands and $m = 2$ (resp., $m = 3$), and with an average error of 0.25% (resp., 0.48%) under Erlang-2 demands and $m = 2$ (resp., $m = 3$). Hence, the average performance error is uniformly within 1%. Similar to the numerical results in Nahmias (1976, 1977b), the approximation algorithms perform better under Erlang-2 demands than exponential demands due to a smaller coefficient of variation. Furthermore, if the balancing parameter is optimized (we search for β and $\hat{\beta}$ over $\{0.5, 0.6, 0.7, \ldots, 1.8, 1.9, 2\}$), then the performance error of the PPB policy does not exceed 0.92%, with an average performance error of 0.28%; and the performance error of the PDB policy does not exceed 0.60%, with an average performance error of 0.20%.

When the product lifetime $m = 4$ and 6, computing the exact optimal policies is intractable. Thus, we compare the performance of our proposed policies with that of Nahmias (1976, 1977b), and report the performance ratios r(DB) and r(PB). We consider the problem with $T = 50$, $\hat{c} = 0$, $\hat{b} \in \{5, 10, 15\}$ and $\hat{\theta} \in \{10, 50, 100\}$. We test Erlang-2, exponential, and hyper-exponential demands with mean 10. The performance ratios are summarized in Table 4.2.

Our numerical results show that the PB policy outperforms Nahmias' policies in 54% of the cases, and the DB policy outperforms Nahmias' policies in 65% of the cases. The

Table 4.2: Performance ratios of heuristics for i.i.d. demands (r) for $m = 4$ and $m = 6$

$\hat{b}$,$\hat{\theta}$	m=4					
	Erlang-2		Exponential		Hyper-exponential	
	PB	DB	PB	DB	PB	DB
5,10	99.14%	99.52%	99.73%	101.11%	99.09%	100.56%
5,50	100.27%	100.43%	100.43%	100.96%	99.96%	100.70%
5,100	99.78%	100.29%	100.22%	100.02%	100.88%	100.59%
10,10	98.66%	99.01%	101.18%	102.43%	101.49%	102.43%
10,50	101.67%	101.80%	107.57%	106.69%	106.71%	107.43%
10,100	100.89%	101.70%	100.05%	101.07%	102.02%	102.84%
15,10	100.06%	100.28%	102.17%	103.85%	102.81%	103.35%
15,50	102.71%	102.81%	116.69%	111.59%	110.86%	111.55%
15,100	100.46%	101.06%	110.21%	108.46%	115.87%	116.65%
	m=6					
5,10	98.78%	97.80%	98.76%	98.93%	99.13%	99.23%
5,50	98.79%	98.93%	98.38%	100.06%	100.00%	100.40%
5,100	98.48%	98.79%	100.04%	100.06%	100.57%	100.60%
10,10	97.58%	97.38%	98.39%	97.97%	98.30%	98.15%
10,50	98.41%	98.84%	99.07%	99.35%	99.05%	100.31%
10,100	98.67%	98.71%	100.79%	100.66%	101.14%	101.57%
15,10	98.28%	98.19%	99.65%	99.50%	98.71%	98.60%
15,50	97.05%	97.24%	99.53%	99.79%	100.03%	100.17%
15,100	97.37%	97.57%	101.38%	102.50%	100.94%	102.08%

PB and DB policies perform 1.09% and 1.34% better on average than Nahmias' polices, respectively. From the numerical results, we can also see that the PB and DB policies have similar performance ratios, which improve as the outdating cost becomes more significant. Our explanation for this finding is as follows: When the frequency of outdating is low, the problem almost reduces to a non-perishable inventory model, for which a myopic policy is optimal in the case of i.i.d. demand. Nahmias' policies, using modified single period cost, yield near-optimal solution when outdating frequency is low.

In summary, under i.i.d. demands, for short product lifetime $m = 2$ and 3, the numerical results of our proposed policies are comparable to those reported in Nahmias (1976, 1977b), as both ours and their policies are very close to optimal. For longer product lifetime $m = 4$ and 6, the overall performance of our proposed policies is also comparable with those of Nahmias (1976, 1977b) and it improves as the frequency of oudating increases.

Since we are more interested in the performances of our policies under correlated demand processes, we conduct more comprehensive studies for this case. As mentioned earlier, when the demand process is correlated over time, the computation of exact optimal solution is intractable for reasonable problem sizes. Thus, in order to compare with the optimal cost under demand settings (b) to (e), we consider a planning horizon $T = 20$ periods and product lifetime $m = 2$ or 3 periods (the same as the majority of the literature under i.i.d. demands). More specifically, we consider $m = 2$ and 3 for the setting (e) and $m = 2$ for the other settings. The cost parameters for each demand class are $\hat{c} \in \{0, 5, 10\}$, $\hat{b} \in \{5, 10, 15\}$, and $\hat{\theta} \in \{0, 5, 10, 15\}$. For these instances, the optimal costs are computed using dynamic programming via backward induction.

Demand setting (b). For demand processes with *advance demand information* (ADI), we adopt a model proposed in Gallego and Özer (2001). We assume that customers can place orders two periods ahead. Thus, in each period t, a demand vector $(D_{t,t}, D_{t,t+1}, D_{t,t+2})$ is received, where $D_{t,s}$ is the order placed in period t for period $s \geq t$. The total demand for period t is $D_t = D_{t-2,t} + D_{t-1,t} + D_{t,t}$. We tested the cases for which each entry $D_{t,s}$ follows an exponential distribution, an Erlang-2 distribution, or a truncated normal distribution with coefficient of variation (cv) being 0.1, 0.3, or 0.5. The mean value for each $D_{t,s}$ is 3, thus the average demand for each period is 9.

To report the numerical results under ADI, we group the instances as follows. The ordering costs are L ($\hat{c} = 0$), M ($\hat{c} = 5$), and H ($\hat{c} = 10$); the outdating costs are L ($\hat{\theta} \in \{0, 5\}$), and H ($\hat{\theta} \in \{10, 15\}$). The first five rows of Table 4.3 report the numerical results under ADI.

The row corresponds to the demand processes (one for exponential demands, one for Erlang-2 demands, and three for truncated normal demands with different cv's). For each pair $(\hat{c}, \hat{\theta})$, we choose between one value of $\hat{c}$, two values of $\hat{\theta}$, and three values of $\hat{b} \in \{5, 10, 15\}$, giving 6 combinations for each pair $(\hat{b}, \hat{\theta})$. The maximum and the average performance errors for both PB and PPB policies are reported in Table 4.3. Among all the test instances under ADI, the average error of the PB (resp., PPB) policy is 0.45% (resp., 0.32%), and the maximum performance error of the PB (resp., PPB) policy is 2.65% (resp., 1.79%).

Demand setting (c). For the *autoregressive* demand model, we consider an AR(1) process $D_t = D_{t-1} + \epsilon_t$ with $D_0 = 10$, where the perturbation term ϵ_t follows a normal distribution with mean 0 and variance 1. The numerical results are reported in the sixth row of Table 4.3. The average performance error of the PB (resp., PPB) policy is 0.66% (resp., 0.40%) while the maximum performance error of the PB (resp., PPB) policy among all test instances is 2.21% (resp., 1.88%).

Demand setting (d). For demand processes of *Martingale model of forecast evolution* (MMFE), we assume that the system in each period t updates its forecast for the next-two-period demands $(D_{t,t+1}, D_{t,t+2})$. The true demand in period t is given by

$$D_t = D_{t-1,t} + \epsilon_{t,1} = D_{t-2,t} + \epsilon_{t-1,2} + \epsilon_{t-1,1},$$

where $\epsilon_{t,i}$ follows a normal distribution with mean 0 and variance 1 for all t and $i = 1, 2$, and $D_{t-2,t}$ follows a normal distribution with mean 10 and variance 10 for all t. The numerical results are reported in the last row of Table 4.3. The average performance error of the PB (resp., PPB) policy is 0.61% (resp., 0.34%); and the maximum performance error of the PB (resp., PPB) policy for all the test instances is 2.39% (resp., 1.50%).

Demand setting (e). The *Markov modulated demand process* (MMDP) is governed by the state of the economy: poor (1), fair (2), and good (3). If the state of the economy in period t is i $(i = 1, 2, 3)$, then the demand in period t is iD_t, where D_t has mean 10 and follows one of the following distributions: exponential, Erlang-2, and truncated normal with coefficient of variation being 0.1, 0.3, or 0.5. We assume that the state of the economy follows a Markov chain with transition probabilities

$$p_{11} = 0.6, p_{12} = 0.3, p_{13} = 0.1, p_{21} = 0.4, p_{22} = 0.2, p_{23} = 0.4, p_{31} = 0.1, p_{32} = 0.3, \text{ and } p_{33} = 0.6.$$

Note that this Markov chain is stochastically monotone, i.e., the state of the economy in

Table 4.3: Performance errors of heuristics for ADI, AR(1), and MMFE demands (% Errors)

	$\hat{c}$	L				M				H				All	
	$\hat{\theta}$	L		H		L		H		L		H			
demand	Policy	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max
ADI	PB	0.56	1.04	0.61	0.94	0.34	0.55	0.33	0.48	0.14	0.24	0.10	0.13	0.33	1.04
cv=0.1	PPB	0.36	0.69	0.40	0.69	0.30	0.51	0.28	0.46	0.12	0.21	0.08	0.12	0.25	0.69
ADI	PB	1.02	1.45	0.77	1.62	0.32	0.99	0.12	0.19	0.10	0.19	0.06	0.10	0.34	1.62
cv=0.3	PPB	0.74	1.39	0.60	1.30	0.21	0.67	0.07	0.11	0.07	0.13	0.05	0.07	0.25	1.39
ADI	PB	0.33	0.53	1.13	2.04	0.51	0.68	0.42	0.56	0.40	0.56	0.17	0.27	0.51	2.04
cv=0.5	PPB	0.24	0.38	0.87	1.58	0.33	0.53	0.32	0.42	0.29	0.48	0.07	0.17	0.36	1.58
ADI	PB	0.75	1.08	1.10	1.89	0.45	0.98	0.40	0.64	0.33	0.55	0.23	0.50	0.52	1.89
Erlang-2	PPB	0.61	0.91	0.79	1.49	0.19	0.55	0.19	0.33	0.15	0.32	0.15	0.28	0.32	1.49
ADI	PB	1.35	2.65	0.86	1.57	0.38	0.65	0.49	0.83	0.30	0.54	0.32	0.58	0.55	2.65
Exp.	PPB	0.97	1.79	0.74	1.49	0.31	0.50	0.37	0.73	0.22	0.46	0.17	0.33	0.42	1.79
AR(1)	PB	1.51	1.95	1.68	2.21	0.36	0.43	0.42	0.76	0.22	0.35	0.22	0.34	0.66	2.21
	PPB	1.06	1.42	1.19	1.88	0.18	0.26	0.14	0.32	0.11	0.20	0.07	0.13	0.40	1.88
MMFE	PB	1.60	2.26	1.49	2.39	0.35	0.58	0.38	0.50	0.15	0.21	0.19	0.42	0.61	2.39
	PPB	0.84	1.26	0.93	1.50	0.14	0.18	0.17	0.23	0.09	0.15	0.09	0.19	0.34	1.50

the next period is stochastically non-decreasing in the state of the economy in the current period.

The performance of our proposed policies under MMDP is reported in Table 4.4. The first column specifies the product lifetimes and demand processes. Similar to Table 4.3, $\hat{c}$ takes three possible values, denoted by L, M and H, and $\hat{\theta}$ is divided into two groups, with L standing for $\{0, 5\}$ and H standing for $\{10, 15\}$. Thus, for each demand process and each pair $(\hat{c}, \hat{\theta})$, there are 6 combinations of $\hat{\theta}$ and $\hat{b}$. The numerical results for both PB and PPB policies are reported in Table 4.4. The average performance error of the PB (resp., PPB) policy among all test instances is 0.71% (resp, 0.44%), and the maximum performance error of the PB (resp., PPB) policy is 2.94% (resp., 1.88%).

Computation time. For the instances with product lifetime $m = 2$ or 3, the optimal policies were computed using dynamic programming. For ADI and AR(1) demand settings, the average computation times for one instance are 784 seconds and 152 seconds, respectively. For MMFE, the average computation time for one instance is 1148 seconds. For MMDP, the average computation time for one instance is 94 seconds for $m = 2$ and 435 seconds for $m = 3$.

In contrast, our proposed policies (including PB, DB, PPB, and PDB policies) do not require any recursive computation and the ordering decisions can be computed in an *online* manner. In period t, the main computation effort lies in the computation of expectation

Table 4.4: Performance errors of heuristics for MMDP demands (% Errors)

	$\hat{c}$	L				M				H				All	
	$\hat{\theta}$	L		H		L		H		L		H			
case	Policy	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max	mean	max
m=2	PB	1.57	2.00	2.44	2.94	0.31	0.49	0.41	0.67	0.30	0.59	0.17	0.27	0.80	2.94
cv=0.1	PPB	1.05	1.41	1.22	1.57	0.17	0.33	0.24	0.53	0.16	0.24	0.10	0.19	0.44	1.57
m=2	PB	0.91	1.94	1.79	2.47	0.31	0.67	0.41	0.91	0.30	0.54	0.30	0.65	0.65	2.47
cv=0.3	PPB	0.57	1.27	1.09	1.74	0.16	0.52	0.21	0.38	0.19	0.48	0.14	0.24	0.38	1.74
m=2	PB	0.39	0.75	1.30	2.18	0.29	0.52	0.49	0.77	0.33	0.61	0.31	0.49	0.53	2.18
cv=0.5	PPB	0.29	0.56	0.76	1.44	0.15	0.23	0.25	0.60	0.19	0.40	0.15	0.23	0.30	1.44
m=2	PB	0.85	1.51	1.55	2.69	0.69	1.06	0.85	1.12	0.66	0.93	0.58	0.77	0.86	2.69
Erlang-2	PPB	0.70	1.20	1.00	1.51	0.44	0.68	0.54	0.74	0.42	0.76	0.34	0.57	0.56	1.51
m=2	PB	0.58	1.32	1.61	2.82	0.95	1.60	0.98	1.44	0.64	1.03	0.68	0.97	0.94	2.82
Exp.	PPB	0.45	1.03	0.89	1.59	0.51	0.91	0.64	1.15	0.27	0.42	0.46	0.77	0.55	1.59
m=3	PB	1.32	1.62	1.90	2.43	0.74	1.34	0.47	0.56	0.41	0.80	0.24	0.46	0.80	2.43
cv=0.1	PPB	0.96	1.13	1.34	1.88	0.44	1.18	0.27	0.47	0.25	0.37	0.14	0.35	0.53	1.88
m=3	PB	1.21	1.28	1.69	1.95	0.45	0.97	0.47	0.59	0.17	0.26	0.14	0.19	0.64	1.95
cv=0.3	PPB	0.61	0.89	1.11	1.62	0.32	0.75	0.40	0.54	0.10	0.17	0.10	0.15	0.43	1.62
m=3	PB	1.15	1.65	0.83	1.51	0.33	0.67	0.38	0.82	0.24	0.38	0.30	0.38	0.48	1.65
cv=0.5	PPB	0.95	1.41	0.57	1.03	0.21	0.44	0.22	0.74	0.13	0.25	0.14	0.23	0.32	1.41
m=3	PB	1.31	1.94	1.41	1.78	0.53	0.85	0.59	0.79	0.40	0.62	0.44	0.64	0.73	1.94
Erlang-2	PPB	0.73	1.04	0.98	1.58	0.38	0.84	0.38	0.51	0.27	0.42	0.27	0.43	0.48	1.58
m=3	PB	0.98	1.41	1.18	1.81	0.53	0.96	0.69	1.05	0.53	0.66	0.35	0.75	0.69	1.81
Exp.	PPB	0.61	0.69	0.68	1.51	0.35	0.88	0.43	0.64	0.32	0.52	0.22	0.56	0.42	1.51

$\mathsf{E}[H_t(q_t) + \Theta_t(q_t) \mid f_t]$ via (4.8) and (4.9). The PB policy takes on average 0.05 second to find a decision for $m = 2$ and 0.6 second for $m = 3$. For longer lifetimes, the PB policy takes on average 1.52 seconds to make a decision for $m = 4$ and 1.94 seconds for $m = 6$. We note that the heuristics proposed in Nahmias (1976, 1977b) are also very efficient as they ignore most of the inventory and future demand information when making a decision. To compute mathematical expectations, both in our procedure and in that of Nahmias (1976, 1977b), we use Monte Carlo simulation with 10000 sample paths. All computations were done using Matlab R2013a on a desktop computer with an Inter Core I7-3770 3.40GHz CPU.

4.7 Conclusions

It is well known that the optimal control policy for perishable inventory systems is complicated and computationally challenging. In this chapter, we develop two approximation algorithms for perishable inventory systems with worst-case performance guarantees. For systems with independent and stochastically non-decreasing demand processes, we propose a dual-balancing (DB) policy that admits a worst-case performance guarantee of 2; and for

systems with arbitrarily correlated demand processes, we propose a proportional-balancing (PB) policy that admits a worst-case performance guarantee between 2 and 3 (2 when the product lifetime is 2). Both policies are easy to compute and implement. More importantly, our numerical study shows that they perform consistently close to optimal for all the tested instances for which we are able to compute their optimal policies: the maximum performance error for the PB policy among all tested instances is below 3%, while the maximum error for the parameterized proportional-balancing (PPB) policy among all tested instances is below 2%. In addition, the average performance errors of the PB policy and the PPB policy among all the correlated demand processes tested, including ADI, AR(1), MMFE, and MMDP, are 0.625% and 0.397%, respectively. For the instances where we are unable to compute the optimal policies, we compare the performances of our approximation algorithms with those of the heuristic policies reported in the literature (Nahmias (1976, 1977b)), and our numerical results show that the performance of our policies is comparable with those of Nahmias (1976, 1977b) and it improves as the frequency of oudating increases.

BIBLIOGRAPHY

BIBLIOGRAPHY

Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Science* **59**(6) 1407–1424.

Bijvank, M., W.T. Huh, G. Jannakiraman, W. Kang. 2014. Robustness of order-up-to policies in lost-sales inventory systems. *Operational Research* **62**(5) 1049 – 1047.

Bijvank, M., I. F .A. Vis. 2011. Lost-sales inventory theory: A review. *European Journal of Operational Research* **215**(1) 1 – 13.

Bookbinder, J. H., A. E. Lordahl. 1989. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions* **21**(4) 302–312.

Bulinskaya, E. V. 1964. Some results concerning optimum inventory policies. *Theory of Probability and Its Applications* **9**(3) 389–402.

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

Cai, X., J. Chen, Y. Xiao, X. Xu. 2009. Optimization and coordination of fresh product supply chains with freshness-keeping effort. *Production and Operations Management* **19**(3) 261–278.

Cai, X., X. Zhou. 2014. Optimal policies for perishable products when transportation to export market is disrupted. *Production and Operations Management* **23**(5) 907–923.

Chao, X., X. Gong, C. Shi, C. Yang, H. Zhang, S. X. Zhou. 2015a. Approximation algorithms for capacitated perishable inventory systems with positive lead times. Working paper. `http://www.umich.edu/~shicong/papers/perishable2.pdf`.

Chao, X., X. Gong, C. Shi, H. Zhang. 2015b. Approximation algorithms for perishable inventory systems. *Operations Research* **63**(3) 585–601.

Chazan, D., S. Gal. 1977. A markovian model for a perishable product inventory. *Management Science* **23**(5) 512–521.

Chen, B., X. Chao, H.-S. Ahn. 2015a. Coordinating pricing and inventory replenishment with nonparametric demand learning. Working paper. Available at `http://ssrn.com/abstract=2694633`.

Chen, B., X. Chao, C. Shi. 2015b. Nonparametric algorithms for joint pricing and inventory control with lost-sales and censored demand. Working paper. Available at http://ssrn.com/abstract=2700491.

Chen, L., E. L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Operations Management* **10**(2) 236–256.

Chen, L. M., A. Sapra. 2013. Joint inventory and pricing decisions for perishable products with two-period lifetime. *Naval Research Logistics* **60**(5) 343–366.

Chen, W., M. Dawande, G. Janakiraman. 2014a. Fixed-dimensional stochastic dynamic programs: An approximation scheme and an inventory application. *Operations Research* **62**(1) 81–103.

Chen, X., Z. Pang, L. Pan. 2014b. Coordinating inventory control and pricing strategies for perishable products. *Operations Research* **62**(2) 284–300.

Chow, Y. S., H. Teicher. 2012. *Probability Theory*. Springer Series in Statistics, Spinger-Verlag.

Chu, L. Y., J. G. Shanthikumar, Z.-J. M. Shen. 2008. Solving operational statistics via a bayesian analysis. *Operations Research Letters* **36**(1) 110 – 116.

Cohen, M. 1976. Analysis of single critical number ordering policies for perishable inventories. *Operations Research* **24**(4) 726–741.

Cooper, W. L. 2001. Pathwise properties and performance bounds for a perishable inventory system. *Operations Research* **49**(3) 455–466.

Deniz, B., I. Karaesmen, A. Scheller-Wolf. 2010. Managing perishables with substitution: Inventory issuance and replenishment heuristics. *Manufacturing & Service Operations Management* **12**(2) 319–329.

Deuermeyer, B. L. 1979. A multi-type production system for perishable inventories. *Operations Research* **27**(5) 935–943.

Ferguson, M. E., O. Koenigsberg. 2007. How should a firm manage deteriorating inventory? *Production and Operations Management* **16**(3) 306–321.

Fries, B. 1975. Optimal ordering policy for a perishable commodity with fixed lifetime. *Operational Research* **23**(1) 46–61.

Gallego, G., Ö. Özer. 2001. Integrating replenishment decisions with advance demand information. *Management Science* **47**(10) 1344–1360.

Glasserman, P. 1991. *Gradient Estimation Via Perturbation Analysis*. Kluwer international series in engineering and computer science: Discrete event dynamic systems, Springer.

Godfrey, G. A., W. B. Powell. 2001. An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science* **47**(8) 1101–1112.

Goh, C. H., B. S. Greenberg, H. Matsuo. 1993. Two-stage perishable inventory models. *Management Science* **39**(5) 633–649.

Goldberg, D. A., D. A. Katz-Rogozhnikov, Y. Lu, M. Sharma, M. S. Squillante. 2016. Asymptotic optimality of constant-order policies for lost sales inventory models with large lead times. *Mathematics of Operations Research* **41**(3) 898–913.

Goyal, S. K., B. C. Giri. 2001. Recent trends in modeling of deteriorating inventory. *European Journal of Operations Research* **134**(1) 1–16.

Graves, S., H. Meal, S. Dasu, Y. Qin. 1986. Two-stage production planning in a dynamic environment. *S. Axster, C. Schneeweiss, E. Silver, eds. Multi-Stage Production Planning and Control. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, Berlin, Germany* 9–43.

Haijema, R., J. van-der Wal, N. M. van Dijk. 2007. Blood platelet production: Optimization by dynamic programming and simulation. *Computers & Operations Research* **34**(3) 760 – 779. Logistics of Health Care Management Part Special Issue: Logistics of Health Care Management.

Haijema, R., J. Wal, N. M. van Dijk. 2005. Blood platelet production: a multi-type perishable inventory problem. Hein Fleuren, Dick Hertog, Peter Kort, eds., *Operations Research Proceedings 2004*. Springer Berlin Heidelberg, 84–92.

Hazan, E. 2015. Introduction to online convex optimization. Book Draft. Computer Science, Princeton University. Available at `http://ocobook.cs.princeton.edu/OCObook.pdf`.

Heath, D. C., P. L. Jackson. 1994. Modeling the evolution of demand forecasts with application to safety stock analysis in production/distribution system. *IIE Transactions* **26**(3) 17–30.

Huh, W. H., P. Rusmevichientong. 2009. A non-parametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1) 103–123.

Huh, W. H., P. Rusmevichientong, R. Levi, J. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research* **59**(4) 929–941.

Huh, W. T., G. Janakiraman, J. A. Muckstadt, P. Rusmevichientong. 2009a. An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand. *Mathematics of Operations Research* **34**(2) 397–416.

Huh, W. T., G. Janakiraman, J. A. Muckstadt, P. Rusmevichientong. 2009b. Asymptotic optimality of order-up-to policies in lost sales inventory systems. *Management Science* **55**(3) 404–420.

Iida, T., P. Zipkin. 2006. Approximate solutions for a dynamic forecast-inventory model. *Manufacturing & Service Operations Management* **8**(4) 407–425.

Janakiraman, G., R. O. Roundy. 2004. Lost-sales problems with stochastic lead times: Convexity results for base-stock policies. *Operations Research* **52**(5) 795–803.

Janakiraman, G., S. Seshadri, J. G. Shanthikumar. 2007. A comparison of the optimal costs of two canonical inventory systems. *Operations Research* **55**(5) 866–875.

Karaesmen, I. Z., A. Scheller-Wolf, B. Deniz. 2011. Managing perishable and aging inventories: Review and future research directions. *International Series in Operations Research & Management Science* **151** 393–436.

Karlin, S., H. Scarf. 1958. *Optimal Inventory Policy for the Arrow-Harris-Marschak Dynamic Model*. Stanford University Press, Stanford, California. In K. Arrow, S. Karlin, and H. Scarf (Eds.), Studies in the Mathematical Theory of Inventory and Production.

Kleywegt, A. J., A. Shapiro, T. Homem-de Mello. 2002. The sample average approximation method for stochastic discrete optimization. *SIAM J. on Optimization* **12**(2) 479–502.

Kullback, S., R. A. Leibler. 1951. On information and sufficiency. *Ann. Math. Statist.* **22**(1) 79–86.

Lariviere, M. A., E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* **45**(3) 346–363.

Levi, R., G. Janakiraman, M. Nagarajan. 2008a. A 2-approximation algorithm for stochastic inventory control models with lost-sales. *Mathematics of Operations Research* **33**(2) 351–374.

Levi, R., M. Pál, R. O. Roundy, D. B. Shmoys. 2007a. Approximation algorithms for stochastic inventory control models. *Mathematics of Operations Research* **32**(4) 821–838.

Levi, R., G. Perakis, J. Uichanco. 2015. The data-driven newsvendor problem: New bounds and insights. *Operations Research* **63**(6) 1294–1306.

Levi, R., R. O. Roundy, D. B. Shmoys. 2007b. Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research* **32**(4) 821–839.

Levi, R., R. O. Roundy, D. B. Shmoys, V. A. Truong. 2008b. Approximation algorithms for capacitated stochastic inventory models. *Operations Research* **56**(5) 1184–1199.

Levi, R., R. O. Roundy, V. A. Truong. 2012. Provably near-optimal balancing policies for multi-echelon stochastic inventory control models. Working Paper.

Levi, R., C. Shi. 2013. Approximation algorithms for the stochastic lot-sizing problem with order lead times. *Operations Research* **61**(3) 593–602.

Li, Q., P. Yu. 2014. Multimodularity and its applications in three stochastic dynamic inventory problems. *Manufacturing & Service Operations Management* **16**(3) 455–463.

Li, Q., P. Yu, X. Wu. 2013. Managing perishable inventories in retailing: Replenishment, clearance sales, and segregation. Working Paper.

Li, Y., A. Lim, B. Rodrigues. 2009. Pricing and inventory control for a perishable product. *Manufacturing & Service Operations Management* **11**(3) 538–542.

Liu, L., Z. Lian. 1999. (s,S) continuous review models for products with fixed lifetimes. *Operations Research* **47**(1) 150–158.

Liyanage, L. H., J. G. Shanthikumar. 2005. A practical inventory control policy using operational statistics. *Operations Research Letters* **33**(4) 341 – 348.

Lu, X., J. S. Song, A. C. Regan. 2006. Inventory planning with forecast updates: approximate solutions and cost error bounds. *Operations Research* **54**(6) 1079–1097.

Lu, Y., M. S. Squillante, D. D. Yao. 2015. Matching supply and demand in production-inventory systems: Asymptotics and optimization. *arXiv preprint arXiv:1501.07141* .

Maglaras, C., S. Eren. 2015. A maximum entropy joint demand estimation and capacity control policy. *Production and Operations Management* **24**(3) 438–450.

Meyn, S.P., R.L. Tweedie. 1993. *Markov chains and stochastic stability*. Springer-Verlag, London.

Mills, T.C. 1990. Time series techniques for economists. *Cambridge University Press* **45**.

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

Nahmias, S. 1975a. A comparison of alternative approximations for ordering perishable inventory. *Information Systems and Operations Research* **13**(2) 175–184.

Nahmias, S. 1975b. Optimal ordering policies for perishable inventory-II. *Operational Research* **23**(4) 735–749.

Nahmias, S. 1976. Myopic approximations for the perishable inventory problem. *Management Science* **22**(9) 1002–1008.

Nahmias, S. 1977a. Comparison between two dynamic perishable inventory models. *Operations Research* **25**(1) 175–184.

Nahmias, S. 1977b. Higher order approximations for the perishable inventory problem. *Operations Research* **25**(4) 630–640.

Nahmias, S. 1977c. On ordering perishable inventory when both demand and lifetime are random. *Management Science* **24**(1) 82–90.

Nahmias, S. 1978. The fixed charge perishable inventory problem. *Operations Research* **26**(3) 464–481.

Nahmias, S. 1982. Perishable inventory theory: A review. *Operations Research* **30**(4) 680–708.

Nahmias, S. 2011. *Perishable Inventory Systems. International Series in Operations Research & Management Science*, vol. 160. Springer.

Nahmias, S., W. Pierskalla. 1973. Optimal ordering policies for a product that perishes in two periods subject to stochastic demand. *Naval Research Logistics Quarterly* **20**(2) 207–229.

Nahmias, S., W. P. Pierskalla. 1976. A two product perishable/non-perishable inventory problem. *SIAM Journal of Applied Mathematics* **30** 483–500.

Nandakumar, P., T. E. Morton. 1993. Near myopic heuristics for the fixed-life perishability problem. *Management Science* **39**(12) 1490–1498.

Perry, D. 1999. Analysis of a sampling control scheme for a perishable inventory system. *Operations Research* **47**(6) 966–973.

Philippou, A. N., C. Georghiou, G. N. Philippou. 1983. A generalized geometric distribution and some of its properties. *Statistics & Probability Letters* **1**(4) 171 – 175.

Pierskalla, W. P. 2004. *Supply Chain Management of Blood Banks. Oeprations Research and Health Care - A Handbook of Methods and Applications*. Kluwer Academic Publishers, New York, 104C145.

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

Prastacos, G. P. 1984. Blood inventory management: an overview of theory and practice. *Management Science* **30** 777–800.

Reiman, M. I. 2004. A new and simple policy for the continuous review lost sales inventory model. *Unpublished manuscript* .

Sethi, S., F. Cheng. 1997. Optimality of (s,S) policies in inventory models with markovian demand. *Operations Research* **45** 931–939.

Shaked, M., J. G. Shanthikumar. 2007. *Stochastic Orders*. Springer Series in Statistics, Physica-Verlag.

Shi, C., W. Chen, I. Duenyas. 2015. Nonparametric data-driven algorithms for multi-product inventory systems with censored demand. Forthcoming, Operations Research.

Shi, C., H. Zhang, X. Chao, R. Levi. 2014. Approximation algorithms for capacitated stochastic inventory systems with setup costs. *Naval Research Logistics* **61**(4) 304–319.

Song, J., P. H. Zipkin. 1993. Inventory control in a fluctuating demand environment. *Operations Research* **41**(2) 351–370.

Tao, Z., S. X. Zhou. 2014. Approximation balancing policies for inventory systems with remanufacturing. *Mathematics of Operations Research* **39**(4) 1179–1197.

Tsybakov, A.B. 2009. *Introduction to Nonparametric Estimation*. Springer-Verlag New York.

Van Zyl, G. 1964. Inventory control for perishable commodities. PhD Thesis, University of North Carolina, Chapel Hill, NC.

Veinott, A. F. 1960. Optimal ordering, issuing and disposal of inventory with known demand. PhD Thesis, Columbia University.

Weiss, H. J. 1980. Optimal ordering policies for continuous review perishable inventory models. *Operations Research* **28**(2) 365–374.

Xin, L., D. A. Goldberg. 2016. Optimality gap of constant-order policies decays exponentially in the lead time for lost sales models. Forthcoming, Operations Research.

Xue, Z., D. D. Yao, M. Ettl. 2011. Managing freshness inventory in direct store delivery supply chains. Working Paper.

Zhang, H., X. Chao, C. Shi. 2016. Nonparametric learning algorithms for optimal base-stock policy in perishable inventory systems with censored demand. Working Paper.

Zhang, H., C. Shi, X. Chao. 2015. Approximation algorithms for perishable inventory systems with setup costs. Forthcoming, Operations Research.

Zhou, D., L. C. Leung, W. P. Pierskalla. 2011. Inventory management of platelets in hospitals: Optimal inventory policy for perishable products with regular and optional expedited replenishments. *Manufacturing & Service Operations Management* **13**(4) 420–438.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.

Zipkin, P. 2008a. Old and new methods for lost-sales inventory systems. *Operations Research* **56**(5) 1256–1263.

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