MULTI-PRODUCT INVENTORY SYSTEMS WITH CONSIDERATION OF FULFILLMENT DYNAMICS: WITH AND WITHOUT DEMAND LEARNING

YANG LIUTAO

PhD

The Hong Kong Polytechnic University

2025

The Hong Kong Polytechnic University

Department of Logistics and Maritime Studies

Multi-Product Inventory Systems with Consideration of Fulfillment Dynamics: with and without Demand Learning

Yang Liutao

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

March 2025

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

________________________ (Signed)

Yang Liutao (Name of student)

Dedication

Dedicated to my mom for her endless love; to my mentors, Dr. Shining Wu and Prof. Li Jiang, for their invaluable guidance; and to my friends, Yuting Yang and Jiaqi Li, for their unwavering support. Each of you has played a crucial role in my journey.

> "*A mother's love is the fuel that enables a normal human being to do the impossible.*"
>
> *Marion C. Garretty*

> "*Mentoring is a brain to pick, an ear to listen, and a push in the right direction.*"
>
> *John C. Crosby*

> "*A friend is someone who knows all about you and still loves you.*"
>
> *Elbert Hubbard*

Abstract

How does the disconnect between back-end inventory control and front-end order fulfillment strategies in e-commerce–stemming from a common organizational structure where an inventory planning team manages the inflow of goods to the warehouse and a separate fulfillment team oversees the outflow of goods to customers–lead to significant financial losses? Additionally, how can a firm control its inventory by accounting for fulfillment dynamics where prior demand information is unknown? In this work, we delve into multi-product inventory systems by accounting for fulfillment dynamics. *Order consolidation* is implemented by the firm to reduce shipping costs, which take up a significant proportion of the total expenses in e-commerce. We first develop deterministic demand models to derive closed-form results and obtain additional managerial insights with more general features incorporated, e.g., under either a partial fulfillment policy or a *whole-order fulfillment* policy, and with ordering cost and endogenized reorder cycle lengths. In the subsequent study, we turn to a stochastic inventory model for the problem under a *partial fulfillment* policy. With known demand information, we formulate the cost function, derive its structural properties–including convexity and submodularity–and characterize the optimal inventory policy. Leveraging these structural properties, we devise asymptotically optimal Parallel Implementation and Optimization (PIO) algorithms that iteratively determine the base-stock levels for each period under unknown prior demand. Furthermore, we conduct numerical experiments to show the performance of our PIO algorithm and demonstrate that neglecting fulfillment dynamics and their associated costs can lead to significant losses. We further make extensions to the case where the commonly assumed lost-sales indicator functions are no longer necessary by developing efficient and effective Cycle-based (Cyc-) PIO and Estimator-based (Est-) PIO algorithms. These extensions address the challenges posed by the lack of information regarding lost-sales indicator function(s), a common issue in existing learning literature.

This study contributes to inventory management by highlighting the importance of accounting for fulfillment dynamics in inventory planning for e-commerce companies and provides valuable inventory control algorithms for multiple products under a partial fulfillment policy.

Acknowledgments

I extend my deepest gratitude to the external examiners, Prof. Xiting Gong and Prof. Jin Qi, for their insightful and invaluable feedback on this thesis. I am also grateful to Prof. Jinzhi Bu for attending my presentation and to Prof. Xiuli Chao for his thoughtful suggestions. Our discussions sparked new perspectives and enriched my work.

My heartfelt thanks go to my supervisor, Dr. Shining Wu, whose guidance and unwavering support have been instrumental throughout my research journey. During moments of uncertainty about my path, Dr. Wu provided clarity and encouragement, fostering both my academic growth and confidence. His wisdom inspired me to explore innovative ideas, particularly during challenging phases. It has been a privilege to witness his milestoneshis promotion to associate professor, the arrival of his daughter, and his mentorship of new studentswhile he, in turn, has guided my transformation from a novice to a self-motivated researcher. His mentorship will forever hold a cherished place in my life.

I am equally grateful to my co-supervisor, Prof. Li Jiang, for his nurturing mentorship and warm support. His encouragement created an environment where I could thrive, and I deeply appreciate the opportunity to serve as his teaching assistant. Lecturing junior PhD students not only deepened my understanding of the material but also honed my teaching skills. Thank you, Prof. Jiang, for your steadfast support.

I owe a special debt of gratitude to my seniors, Dr. Yanli Tang, Dr. Hanxiang Zhang, Dr. Lingxiao Wu, and Dr. Ran Yan, whose guidance illuminated my path during difficult times. Their support brought warmth and inspiration when I needed it most.

To my dear friends, Yuting Yang and Jiaqi Li, your companionship has been a powerful source of motivation throughout this journey. I cherish the memories weve created together. I also fondly recall the joyful times spent with Hang Yu, Jingmai Wang, and Lingyun Chen, which remain etched in my heart. A special thank you to Shaoqun Lin, my hiking partner, for bringing joy and lightness during demanding periods.

To those not named here, please know that my gratitude for your friendship is no less profound. The list of those who have supported me is vast, and each of you holds a meaningful place in my journey.

Finally, I owe an immeasurable debt to my mother, whose unwavering love and encouragement have been my bedrock. Her belief in me has sustained me through challenges and inspired me to persevere. Thank you all for your invaluable support.

Contents

I Introduction **2**

1 Introduction **3**

1.1 Background . . . 3
1.2 Fulfillment Policies . . . 4
1.3 Modeling Framework . . . 5
1.4 Contributions . . . 6
1.5 Outline . . . 7
1.6 General Notations . . . 8

2 Literature Review **9**

2.1 Inventory Control With Known Demand Information . . . 9
2.2 Inventory Control With Unknown Demand Information . . . 17
2.3 Methodologies . . . 20
2.3.1 Online Convex Optimization . . . 20
2.3.2 Regret Analysis . . . 21
2.3.3 Discrete Convexity . . . 23

3 Problem Formulation **25**

3.1 Deterministic Demand . . . 25
3.1.1 The Economic Order Quantity Model for a Single Product . . . 26
3.1.2 The Two-Product Model . . . 27
3.2 Stochastic Demand . . . 27

II Deterministic-Demand Models **30**

4 Single-product Problem with Deterministic Demand (The EOQ Model) **31**

4.1 The Case of Backlog . . . 31
4.2 The Case of Lost Sales . . . 32

5 Two-Product Problem with Deterministic Demand and Partial Fulfillment 35

5.1 The Case of Backlog . . . 35

5.2 The Case of Lost Sales . . . 37

6 Two-Product Problem with Deterministic Demand and Whole-Order Fulfillment 40

6.1 The Case of Backlog . . . 40

6.2 The Case of Lost Sales . . . 43

7 Conclusions 46

III Stochastic-Demand Models under Partial Fulfillment 47

8 Partial Fulfillment with Known Demand Information 48

8.1 The Case of Backlog . . . 49

8.2 The Case of Lost Sales . . . 54

9 Partial Fulfillment with Learning 57

9.1 The Learning Algorithm . . . 59

9.1.1 Calculating Subgradient and Updating $\vec{Z}_t$ . . . 61

9.1.2 Probabilistic Rounding for an Implementable $\vec{Y}_{t+1}$ . . . 65

9.2 Performance of the Algorithm . . . 66

9.2.1 Regret Upper Bound . . . 66

9.2.2 Regret Lower Bound . . . 69

9.3 The Case of Lost-Sales . . . 70

9.3.1 The Learning Algorithm . . . 70

9.3.2 Algorithm Performance . . . 72

10 Numerical Experiment 74

10.1 Performance of the PIO Algorithm . . . 74

10.2 The Value of Accounting for the Fulfillment Process . . . 75

11 Extension: Without Lost-Sales Indicator Function(s) 79

11.1 Cyc-PIO Algorithm: $O(T^{2/3})$ Regret . . . 80

11.1.1 Single-Product Case . . . 80

11.1.2 Two-Product Case with Fulfillment Dynamics . . . 81

11.2 Est-PIO Algorithm: $O(\sqrt{T})$ Regret . . . 84

11.2.1 Single-Product Case . . . 84

11.2.2 Two-Product Case with Fulfillment Dynamics . . . 87

12 Discussions and Conclusions **93**

Bibliography **94**

A Proof for the Convexity of the Expected Cost Function **101**
A.1 The Case of Backlog 101
A.2 The Case of Lost Sales 105

B Proofs for Regret Upper Bound **108**
B.1 Proof for Lemma 2 108
B.2 Proof for Lemma 3 109
B.3 Proof for Lemma 4 111
B.4 Proof for Lemma 5 111
B.5 Proof for Lemma 7 112

C Proofs for the Regret Lower Bound **114**
C.1 The Case of Backlog 114
C.1.1 Construction 114
C.1.2 Proof for Theorem 3 Under Our Construction 116
C.2 The Case of Lost Sales 117
C.2.1 Construction 117
C.2.2 Proof for Theorem 5 Under Our Construction 119

D Proofs for Cyc-PIO and Est-PIO Algorithms **120**
D.1 Regret Upper Bound for Cyc-PIO Algorithm 120
D.1.1 Single-Product Case with Lost Sales 120
D.1.2 Two-Product Case with Lost Sales and Fulfillment Dynamics . . . 124
D.2 Regret Upper Bound for Est-PIO Algorithm 126
D.2.1 Single-Product Case with Lost Sales 126
D.2.2 Two-Product Case with Lost Sales and Fulfillment Dynamics . . . 128

List of Figures

3.1 Sequence of events 29

4.1 Inventory position as a function of time in backlog case 31
4.2 Inventory position as a function of time in lost sales case 33

5.1 Fluid model for the two-product problem with partial fulfillment under backorder setting 36
5.2 Fluid model for the two-product problem with partial fulfillment under lost sales setting 38

6.1 Fluid model for the two-product problem with whole-order backlogged . 41
6.2 Fluid model for the two-product problem with whole-order lost 43

8.1 Contour map of costs with different demand correlation 55

9.1 A road map for the algorithm 60
9.2 Two different patterns of the cost function 62
9.3 Illustration of probabilistic rounding 65

10.1 Cumulative regret of the PIO algorithm 75
10.2 Converging performance of the PIO algorithm 76

11.1 Cumulative regret of single-product Cyc-PIO algorithm 82
11.2 Evolution of decision variables over time in the two-product problem . . 83
11.3 Comparison of Est-PIO and PIO algorithms in single-product problem . 88
11.4 Comparison of Est-PIO and PIO algorithms in two-product problem with lost sales 91
11.5 Comparison of Est-PIO and PIO algorithms in two-product problem with backlog 92

A.1 Sequence correspondence 107

List of Tables

2.1 Classification of literature on inventory problem with correlated demand under known demand information 16

10.1 Relative cost difference of the myopic and strategic firms (the backlog case) 77

10.2 Relative cost difference of the myopic and strategic firms (the lost-sales case) 78

11.1 BCO performance 80

Part I

Introduction

Chapter 1

Introduction

1.1 Background

The field of inventory management has produced a wealth of literature on optimal inventory control across various contexts since the inception of operations research. The primary goal is to identify inventory policies that minimize costs associated with holding inventory, backlogs, and lost sales. These costs include handling, obsolescence, spoilage, and the capital costs of holding inventory, as well as expedited shipping, loss of goodwill, and profit loss linked to backlogs or lost sales. However, fulfillment costs, or shipping costs, are generally classified as selling and administrative expenses rather than daily operational costs by standard accounting practices. This separation often leads e-commerce firms to treat back-end inventory control and front-end order fulfillment as distinct processes, excluding freight-out shipping costs from back-end inventory planning. Most existing studies also adopt this "divide-and-conquer" approach, overlooking the fulfillment process and its associated shipping costs in inventory decision-making.

Such an oversight could hinder cost reduction in the realm of e-commerce, where the availability of inventory directly affects fulfillment options and their associated costs. Order consolidation is a key strategy in this context. By shipping multiple items in one package, e-commerce firms can lower fulfillment costs and maintain service quality. For instance, Capital One Shopping Research reveals that the average order during an Amazon Prime event consists of 1.88 items. This purchasing behavior, characterized by multiple products in a single transaction, presents an opportunity for fulfillment through order consolidation.

Given that fulfillment costs constitute a significant portion of total operational expenses in e-commerce, the prevailing industry practice–overlooking the influence of the fulfillment process on inventory decisions–risks forgoing substantial savings that could be achieved through effective coordination of inventory levels. For example, Amazon reported total shipping costs of $89.5 billion and profits of $30.4 billion in 2023. A mere

10% reduction in shipments could yield billions in savings and potentially increase profits by around 30%.

In summary, while the prevalent divide-and-conquer practice simplifies decision-making in inventory planning and order fulfillment, firms may incur substantial losses by neglecting the potential for joint optimization of these costs. This paper investigates the critical yet under-explored issue of incorporating fulfillment dynamics into inventory planning within the context of multi-product scenarios and order consolidation, which are common in e-commerce.

1.1.1 Research Questions

Given this context, we examine the following research questions:

- How should the firm adjust its inventory decisions when their impact on the subsequent fulfillment process is taken into account?
- How much reduction in the total cost can be achieved by joint optimization of both inventory and fulfillment costs in inventory planning?
- How can the firm make online inventory decisions for such a joint optimization problem when the demand information is unknown *a priori* and needs to be learned?

1.2 Fulfillment Policies

We begin our study with a two-product setting in that it is the most representative multi-product scenario in practice based on empirical data. For example, a recent survey by Capital One Shopping Research reveals that 92% of orders on Amazon contain 1-2 items, while only 8% include 3 or more items. The demand, in the form of orders, can be categorized into three different types: type-1 orders require only item 1, type-2 orders require only item 2, and type-12 orders require both items 1 and 2. The orders can be fulfilled either with a partial fulfillment policy or a whole-order fulfillment policy, as introduced as follows.

1.2.1 Partial Fulfillment

As in the common e-commerce setting, customers' orders are fulfilled by express shipping. In our problem, the firm is committed to providing high-quality logistics services and thereby adopts a partial fulfillment practice. This practice stipulates that the available part of an order is shipped **immediately** upon order placement and the unavailable part is either backlogged until inventory replenishment (in the case of backlog) or lost in the case where backorders are not accepted (in the case of lost sales). Furthermore, when multiple items need to be shipped to a customer at a given moment, they are grouped and shipped by a **single** package, which saves cost without sacrificing the service quality. For example, if a type-12 order arrives and both products are available, then one unit of each product will be shipped immediately to the customer in one package. Another example is when a type-12 order arrives but only product 1 is available, one unit of product 1 will be shipped immediately to the customer. Then, depending on which case we are studying, the demand for product 2 would be **either** backlogged and shipped separately to the customer when inventory is replenished later, rendering the order satisfiable by two separate shipments, **or** lost if backorders are not allowed, implying that this order requires one shipment but is only partly satisfied. For ease of reference, we refer to the policy of the backlog case as "*Partial Fulfillment with Backlog*" and that of the lost-sales case as "*Partial Fulfillment with Lost Sales*".

1.2.2 Whole-Order Fulfillment

Another common practice among e-commerce companies is the "whole-order fulfillment" policy. In this approach, if all items in an order are in stock, the order is shipped immediately to the customer. Conversely, if any item is out of stock, the entire order is delayed until all items can be fulfilled or may even be lost. For example, if a type-12 order arrives and both products are available, then one unit of each product will be shipped immediately to the customer. Another example is when a type-12 order arrives but only product 1 is available, then depending on which case we are studying, **either** the two products in this order would be both backlogged and shipped in one order to the customer when the inventory is replenished later, **or** the demand for this order is totally lost if backorders are not allowed. For ease of reference, we refer to the policy of the backlog case as "*whole-order fulfillment with backlog*" and that of the lost-sales case as "*whole-order fulfillment with lost sales*".

While this strategy helps reduce shipping costs, the resulting delays can lead to lower customer satisfaction, and the loss of entire orders can significantly impact the firm's profitability.

1.3 Modeling Framework

Under each fulfillment policy, e-commerce companies may encounter two distinct scenarios: backlog and lost sales. This leads to the four aforementioned settings:

- Partial fulfillment with backlog.
- Partial fulfillment with lost sales.

- Whole-order fulfillment with backlog.
- Whole-order fulfillment with lost sales.

We study a periodic review inventory control problem where the demand (in the form of customer orders) for two products is complementary. Unsold items from one period can carry over to the next. We categorize the demand into three types: type-1 orders require only item 1, type-2 orders require only item 2, and type-12 orders require both items 1 and 2.

We begin by formulating a deterministic demand model, where the demand rate for each type of order is constant. For an exogenously given inventory cycle length, we solve for the optimal order-up-to levels for both products under the four aforementioned settings. Next, we endogenize the inventory cycle length and solve for it optimally. By comparing the results with the single-product Economic Order Quantity (EOQ) model as a benchmark, we find that the optimal inventory decisions for the two-product case differ from those of the benchmark. This indicates that we cannot simply adapt the single-product solution to the two products separately, highlighting the importance of addressing our new problem.

Next, we explore the stochastic demand model, in which customer orders arrive sequentially, with each order requiring at most one unit of each item. First, we characterize the expected cost function and determine the optimal inventory policy under known demand information while incorporating fulfillment dynamics. We compare the optimal inventory decision to the one derived by ignoring fulfillment dynamics, revealing significant differences between the two approaches. Neglecting fulfillment dynamics can lead to a substantial increase in total costs. Next, we address the inventory control problem under unknown prior demand by devising an efficient and effective online learning algorithm. In the stochastic model, we adopt a reinforcement learning methodology, specifically an online subgradient descent method. It is important to note that our approach differs significantly from existing methods due to the complexities of fulfillment dynamics and the discrete nature of our problem.

1.4 Contributions

We address the inventory control problem within the context of multiple products, demand complementarity, and fulfillment policies that include shipping costs. Specifically, in the stochastic demand model, we focus on the partial fulfillment policy, tackling the inventory control problem under both known demand information and unknown prior demand. We develop an efficient and effective online learning algorithm tailored to the partial fulfillment policy. Additionally, we present numerical comparisons of the relative cost differences between firms that incorporate fulfillment dynamics and those that overlook them in their inventory decision-making. For the deterministic demand model, we fully solve the problem under all four settings with known demand rates. We further consider endogenized reorder cycle length and setup cost, solving the problems to optimality. The focus on fulfillment dynamics is essential as the shipping cost is significant in e-commerce companies.

By comparing the optimal costs when incorporating fulfillment dynamics with those that ignore these factors, our study highlights the necessity of considering front-end fulfillment dynamics in inventory decision-making.

1.5 Outline

In the remainder of this part, we structure the content as follows: In Chapter 2, we review relevant literature, focusing on the availability of demand information and the methodologies used in inventory control studies. In Chapter 3, we first formulate the inventory control problem with stochastic demand, and then transition to the formulation of the problem with deterministic demand.

In the second part, which addresses deterministic demand, the content is organized as follows: In Chapter 4, we differentiate between scenarios where unfilled demand is backlogged or lost, and we sequentially solve the single-product problem with exogenous and endogenous inventory cycle lengths (abbreviated as w/ExEnCs). In Chapter 5, we again differentiate between the two handling mechanisms for unfilled demand, solving the two-product problem w/ExEnCs. In Chapter 6, we explore the two handling mechanisms for unfilled orders and solve the two-product problem w/ExEnCs. Finally, in Chapter 7, we summarize the main results and conclude our study.

In the third part, which addresses stochastic demand and partial fulfillment policy, the content is structured as follows: In Chapter 8, we focus on formulating the cost function and provide convexity results under the scenarios of unfilled demand being backlogged or lost. This allows us to identify the type of optimal policy. In Chapter 9, we develop online learning algorithms tailored for multi-product inventory control, taking fulfillment dynamics into account. We also present results on asymptotic optimality and the best achievable regret for our algorithm. In Chapter 10, we provide numerical evidence to demonstrate the value of incorporating fulfillment dynamics into inventory decision-making. Finally, in Chapter 12, we summarize the main contributions and outline future research directions.

1.6 General Notations

Some general notations will be used. Given a function $f : \mathbb{R}^n \to \mathbb{R}$, let $\Gamma_f(\vec{x}) := (\vec{x}, f(\vec{x}))$ denote *an element of the graph of* f *at* $\vec{x}$. Let $\mathbb{I}\{\cdot\}$ be the *indicator function* where $\mathbb{I}\{A\}$ equals 1 if the event A is true and 0 otherwise. The *Euclidean projection function* $\mathbf{P}_V(\vec{X}) := \arg\min_{\vec{P}\in V} ||\vec{X} - \vec{P}||_2$ projects the point $\vec{X}$ onto the nearest point in domain V with respect to the Euclidean norm (2-norm). The *convex hull* of a set S is denoted by $\mathrm{Hull}(S)$. Let $x_1 \vee \cdots \vee x_k := \max\{x_1, \cdots, x_k\}$ denote the *join* function defined on multiple variables. The expressions $(x)^+ := \max\{x, 0\}$ and $(x)^- := \max\{-x, 0\}$, sometimes written as x^+ and x^- for brevity, denote the positive and negative parts of a given number x, respectively. Furthermore, element-wise operations are performed when the $\vee$, $(\cdot)^+$, $(\cdot)^-$ operators are applied to vectors; for example, the expression $(\vec{x})^+$ returns a vector consisting of the positive parts of each element of the original vector $\vec{x}$. The vector $\vec{e}_i$ is defined as the *unit vector* with a value of 1 at the ith entry and 0 elsewhere. The dimension of $\vec{e}_i$ is determined by the specific context in which it is used.

Chapter 2

Literature Review

In order to develop a comprehensive understanding of the inventory control problem, we begin by conducting a thorough review of the literature on inventory control models that assume known demand information. This review will provide a solid foundation for our subsequent discussions. Next, we shift our focus to the methodologies adopted in this paper. Firstly, we will review the concept of discrete convexity, which originates from discrete mathematics and is commonly used in operations management to establish the convexity of discrete functions. Understanding this concept will be important in proving the convexity of the proposed models in this paper. Secondly, we will provide a review of the "online convex optimization method", which is a widely recognized optimization technique utilized to tackle the challenges posed by censored demand data. This method has shown promise in handling scenarios where demand information is unknown beforehand, making it highly relevant to our problem. Thirdly, we will review the methods used to establish bounds in the regret analysis. These techniques are crucial in assessing the performance of inventory control policies and understanding the trade-offs involved. By building upon these existing theories and methodologies, we aim to address the complexities and uncertainties associated with scenarios where demand information is unknown. This progression will enable us to develop effective inventory control approaches while accounting for the challenges posed by unknown demand information, as reviewed in the third part.

2.1 Inventory Control With Known Demand Information

Inventory management as one of the traditional and most important topics in operations management has been extensively studied over the past century.

Assuming that the demand distribution is known or has been accurately estimated by the firm, researchers have looked into *single-product inventory systems* under various settings and developed policies for managing them, including classical results like economic order quantity (EOQ) (Harris, 1915) and (Q, r) model (Galliher et al., 1959) for continuous-review systems, and (S, s) policy (Scarf et al., 1960) for periodic-review systems. Noting that these policies all belong to a base-stock type family, and have gained widespread adoption in practice due to their simplicity and insightfulness. These policies (of based-stock type) have been shown to be optimal or near-optimal (for cases with lost sales and lead time) for nonperishable inventory systems under very mild conditions (see Simchi-Levi et al. 2013). Relevant costs considered in these studies and existing literature mainly include inventory holding, backlog, loss of profit, and loss of goodwill. Fulfillment cost is in general not taken into account and is usually viewed as overhead that is independent of inventory decisions.

When researchers extend to study inventory control for multiple products, the multi-product features that are incorporated into inventory models mainly include location-specific products, demand substitution, Poisson demand distribution, etc. In other words, they do not the influence of inventory decisions on the order fulfillment process and the costs generated therein.

By thoroughly reviewing these inventory control models with known demand information, we are able to lay a solid foundation for our subsequent discusssions on inventory control problems involving unknown demand information.

2.1.1 Single-product inventory control problem.

2.1.1.1 Deterministic demand model.

The Economic Order Quantity (EOQ) model, originally proposed by Harris (1915) and later utilized by Wilson (1934), stands as the pioneering mathematical inventory control model. This model assumes a constant and known demand rate, an infinite planning horizon, and instantaneous replenishment without allowing any shortages. The EOQ model effectively illustrates the tradeoff between inventory holding costs and ordering costs, providing insights into cost optimization. As inventory control research progressed, the extended EOQ model emerged to address additional complexities. For instance, the extended model incorporates positive lead time (Liberatore, 1979), capacity constraint (Eisenhut, 1975; Blackburn and Millen, 1984; Erdem et al., 2006), and shortage (Liberatore, 1979; Pentico and Drake, 2009). These extensions allow for a more realistic representation of inventory management scenarios and provide insights into managing inventory under various constraints.

While the EOQ model provides a straightforward framework for inventory control, it does not capture the inherent fluctuation property of real-world demand. The model's simplicity limits its ability to account for demand variability and any resulting uncertainties. Recognizing this limitation, researchers have developed stochastic demand models to accommodate various practical considerations.

2.1.1.2 Stochastic demand model.

The stochastic demand models take into account the probabilistic nature of demand, allowing for a more realistic representation of the uncertainties associated with inventory management. By considering the probabilistic nature of demand, the stochastic model acknowledges that demand can vary from one period to another, allowing for a more accurate representation of real-world scenarios. This variability is essential to account for the uncertainties and fluctuations that businesses often encounter in their day-to-day operations. By incorporating the stochastic demand model, businesses can make more informed decisions regarding inventory management.

Single-period inventory control. The newsvendor model, initially proposed and examined by Edgeworth (1888) and Arrow et al. (1951), represents a single-period inventory planning model that allows for inventory shortages. In this model, the manager must strike a balance between potential overage (procurement cost minus salvage value) and underage costs (price minus procurement cost). The practical relevance and versatility fo the newsvendor model have led to its widespread applications and various generalizations. Researchers have explored extensions such as the price-setting newsboy problem (Whitin, 1955; Mills, 1959; Karlin and Carr, 1962; Petruzzi and Dada, 1999; Qin et al., 2011; Schulte and Sachs, 2020), which considers setting prices for perishable goods. Additionally, the Poisson demand newsvendor problem (Schulte and Sachs, 2020) addresses situations where demand follows a Poisson distribution. The behavioral aspects of the newsvendor problem have also been investigated (Kazaz and Webster, 2015; Bai et al., 2019), taking into account decision-making biases and psychological factors that can influence inventory planning and ordering decisions. Furthermore, the multi-location newsvendor problem (Xiao and Wang, 2023) explores scenarios where inventory planning must incorporate the complexities associated with coordinating inventory across different locations. These extensions and generalizations of the classic newsvendor problem have expanded its applicability and provided insights into various real-world inventory management scenarios, allowing for more informed decision-making and optimization strategies.

Continuous-review inventory control. The (Q, r) policy assumes that the seller continuously monitors the inventory position, which is defined as the stock on hand plus stock on order minus any backorders (Galliher et al., 1959; Browne and Zipkin, 1991;

Federgruen and Zheng, 1992). When the inventory position falls below a predetermined threshold, known as the reorder point, denoted as r, the seller replenishes the stock in a lot size of Q. Browne and Zipkin (1991) conducted a study on the (Q, r) model with continuous demand driven by a stochastic process. They established the insensitivity property under certain assumptions, highlighting its robustness in managing inventory. Federgruen and Zheng (1992) provided an efficient algorithm with linear complexity to determine the optimal reorder quantity Q^*. This algorithm assists in determining both the optimal reorder point and reorder quantity for scenarios with discrete demand. Kouki et al. (2015) investigated a perishable inventory system characterized by stochastic continuous demand, constant lifetime, and deterministic lead time. They developed an algorithm to determine the optimal values of the reorder point r and reorder quantity Q that minimize the total cost in such a system. These studies contribute to the understanding and optimization of inventory management under various demand scenarios, providing valuable insights for determining optimal reorder points and quantities, thereby minimizing costs and improving operational efficiency.

Periodic-review inventory control. The (s, S) policy is a periodic inventory review approach where two thresholds, denoted as s and S, are set. In this policy, the seller replenishes the stock up to level S whenever the ending inventory in a cycle falls below s. Arrow et al. (1951) discovered that the specific values of the two thresholds in the (s, S) policy depend on factors such as the demand distribution, setup cost, and cost of depletion penalty. Dvoretzky et al. (1953) and Scarf et al. (1960) contribute to understanding the conditions under which the (s, S) policy is the optimal choice. Dvoretzky et al. (1953) investigated the necessary and sufficient conditions, in terms of demand distribution, that validate the optimality of the (s, S) policy for certain inventory problems. Scarf et al. (1960) demonstrated the optimality of the (s, S) policy when setup costs and reorder costs are present, and the holding and shortage costs exhibit linear format. Veinott Jr and Wagner (1965) and Zheng and Federgruen (1991) delve into finding the optimal values for s and S in the (s, S) policy. Veinott Jr and Wagner (1965) developed an efficient algorithm for calculating the two thresholds, s and S, in the presence of setup costs, discount factors, and positive lead times. Zheng and Federgruen (1991) made further contributions to this field by introducing a simple algorithm to find the upper bound for S^* (the optimal value of S) and the lower bound for s^* (the optimal value of s) in (s, S) policy. These studies have significantly advanced our understanding of the (s, S) policy and provided valuable insights for determining the optimal thresholds values. As a result, businesses can effectively manage their inventory levels, minimize costs, and improve overall operational efficiency.

In addition, there is the base stock policy, which is a special case of the (s, S) policy. Under the base stock policy, replenishment is triggered whenever the inventory drops below a predetermined level S. Veinott Jr (1965b) conducted a study on a multiperiod single-product captive-demand inventory control problem and established conditions under which the base stock policy is optimal. They also demonstrated that the base stock level can be easily determined based on these conditions. Federgruen and Zipkin (1986) expanded on the base stock policy by introducing a modified version that is proven to be optimal for both finite- and infinite-period planning problems with continuous and stationary demand distributions. Their proof of optimality relies on leveraging the convexity of the one-period cost function. Feng et al. (2006) provided intuitive proof of the optimality of the base stock policy in certain delivery modes. They considered the separability of the cost-to-go function with respect to the post-order inventory positions. These studies have made valuable contributions to the strand of literature on base stock policy, shedding light on determining the appropriate base stock level and demonstrating the effectiveness of the policy in a wide range of inventory management scenarios.

2.1.2 Multi-product inventory control problem.

The research on inventory control of multiple products is relatively limited, and there is a need for further exploration in this area. While Turken et al. (2012) conducted a comprehensive literature review focusing on the multi-product single-period newsvendor problem, there is still much to be explored in the context of the multi-product multi-period inventory control problem. Specifically, more attention should be given to understanding the correlation or lack thereof in demands for different products. This area of research presents an opportunity to investigate optimal inventory control strategies for managing multiple products over a planning horizon spanning multiple time periods. By delving deeper into this problem domain, researchers can broaden the scope of developing strategies to address inventory control challenges involving multiple products. This expansion of knowledge will not only contribute to the academic literature but also have practical implications for businesses.

2.1.2.1 Uncorrelated demand.

The multiproduct inventory control problem with uncorrelated demand assumes that the demands for different products exhibit zero Pearson correlation. In other words, it assumes that there is no linear relationship between the demands of different products. This assumption allows researchers to model each product's demand independently, without considering the interdependencies or common factors that may influence the demands of multiple products simultaneously, simplifying the problem and decision-making processes.

Continuous-review. In their study, Roundy (1986) introduced a novel class of policies for an infinite-horizon inventory control problem. They specifically focused on a scenario where a single item held at different locations is treated as a distinct product. Notably, each policy proposed in their work achieves a performance level within 2% of the optimal solution. This highlights the effectiveness and efficiency of their approach in managing inventory across multiple locations. In Hill and Pakkala (2007), they develop a solution procedure for determining the best base-stock policy with independent Poisson demand under the setting of partial fulfillment with delivery cost. Different from theirs, we consider correlated demand problems under a general demand distribution and provide an online learning algorithm to iteratively determine the optimal base-stock levels.

For a comprehensive overview of multi-product inventory control problems, including those with single or multiple constraints, Haksever and Moussourakis (2005) and the references cited therein provide a detailed review of the literature. The reader interested in exploring this topic further will find this comprehensive review to be a valuable resource.

Periodic-review. In their research, Veinott Jr (1965a) identified conditions that allow for expressing the optimal policy in a simple form for a multiproduct inventory control problem with nonstationary demand. Their study sheds light on the factors that influence the structure and design of optimal policies in dynamic inventory management settings where multiple products and varying demand patterns are involved.

2.1.2.2 Correlated demand.

Complementary and substitutable products are prevalent in various real-world scenarios. A classic example of complementary products is electronic toothbrushes and their corresponding replacement heads, which are typically purchased together. The demand for one product, such as the electronic toothbrush, naturally generates a need for the other product, the replacement heads, highlighting their complementary relationship. This example exemplifies how the demand for one product drives the demand for another, illustrating the interdependency and interconnectedness of complementary products in consumer behavior and purchasing patterns.

Our model is closely aligns with previous studies such as Hausman et al. (1998) and Song (1998), which also investigate multi-product, base-stock inventory systems with correlated demand, partial shipment, order backlog, and a first-come first-served (FCFS) rule. In our study, we specifically consider a periodic-review and “unit-demand” problem, as defined by Song (1998), where demands are correlated across different types of items within a given period but are independent across time periods, and the overall demand process is stationary. Additionally, we extend the analysis to include the case where backlog is not permitted, resulting in lost sales for unfulfilled demand.

Continuous-review. Song (1998) investigate the order fill rate in a continuous-review, multi-product, base-stock inventory system with a constant lead time. They assume that the demand process is a multivariate compound Poisson process and propose procedures to determine the order fill rate in a tractable manner. Notably, they begin by considering a two-product unit-demand system and subsequently generalize to encompass a multi-product unit-demand system and finally a general demand size system. Furthermore, Hill and Pakkala (2007) propose a base-stock policy that minimizes the total cost in a continuous-review, multi-product system when demand follows Poisson distribution. In their settings, partial shipment, stochastic lead time, and time-weighted holding and backorder costs are considered. In a more recent study, Poormoaied and Atan (2020) formulated a continuous-review inventory system with two complementary products. The demand process follows Poisson distribution, and the cost components include lost sales and time-weighted holding costs. They employ a simulation-based optimization approach to determine the parameters of the optimal (Q, r) policy to maximize the expected profit rate. They focus on a scenario where the entire order would be lost if a customer demands both products simultaneously. For readers seeking a comprehensive review of inventory control problems that involve correlated demand, we recommend referring to Poormoaied and Atan (2020).

Periodic-review. Hausman et al. (1998) consider a periodic-review, multi-product, base-stock inventory system when lead time is constant. Under the assumption that the demand process follows a multivariate normal distribution, they evaluate and maximize the order fill rate. In recent research, Poormoaied (2022) investigated a periodic-review base-stock inventory system with two complementary products. The demand is Poisson and the firm adopts partial shipment. The cost components include lost sales and time-weighted holding costs. The lead time is assumed to be zero. They introduce an exhaustive search algorithm to examine how product complementarity affects the optimal base-stock level and reorder cycle time.

See Table 2.1 for a summary of the literature on inventory problems with correlated demand. By building upon these prior studies, our research contributes to the understanding of multi-product inventory control with demand complementarity, partial fulfillment, and shipping costs within the framework of a periodic-review base-stock inventory system with zero lead time, partial fulfillment, and either backlog or lost sales.

Table 2.1: Classification of literature on inventory problem with correlated demand under known demand information

Article	**Periodic-/ Continuous-review**	**Control policy**	**Lead time**	**Order fulfillment method**	**Backorder/ Lost sales**	**Demand process**	**Cost components**	**Objective**
Hausman et al. (1998)	periodic-review	base-stock policy	constant	partial shipment and complete fulfillment	backorder	multivariate normal distribution	-	Evaluating order fill rate.
Song (1998)	continuous-review	base-stock policy	constant	partial shipment	backorder	multivariate compound Poisson process	-	Evaluating order fill rate.
Poormoaied and Atan (2020)	continuous-review	(Q, r) policy	zero	complete fulfillment	lost sales	Poisson process	time-weighted holding cost, lost-sales cost, and fixed replenishment cost	Determining the optimal parameters of (Q, r) policy that minimize the total cost.
Poormoaied (2022)	periodic-review	base-stock policy	zero	complete fulfillment	lost sales	Poisson process	time-weighted holding cost, lost-sales cost, and fixed replenishment cost	Determining the optimal period length and the base-stock levels that maximize the expected profit rate.
This paper	periodic-review	base-stock policy	zero	partial fulfillment	backorder and lost sales	-	holding cost, backorder cost or lost-sales cost, and delivery cost	Determining the optimal base-stock levels that minimize the total cost.

2.2 Inventory Control With Unknown Demand Information

The classic newsvendor problem assumes prior knowledge of the demand distribution, enabling the seller to determine optimal order quantities to maximize expected profit or minimize expected cost. However, in the multi-period *censored* demand newsvendor problem, the seller does not know the demand distribution a priori. Instead, only the sales data, represented by $\min\{D, Q\}$, is observed. Because the seller does not have known information about the demand, estimating its distribution or parameters therein becomes challenging. As a result, a significant body of literature is dedicated to leveraging available information effectively to maximize expected profit or minimize expected cost with censored demand.

The emergence of online learning methods has brought increased interest in optimizing the trade-offs between exploration and exploitation when managing sequentially arrived demand. This is especially relevant in inventory control, where decisions must account for limited or censored demand data. One key consideration is balancing between maintaining sufficient inventory levels to reduce stock-out risks and avoiding overstocking, which can harm profitability. Gao and Zhang (2022b) comprehensively reviewed recent literature on inventory control problems addressing censored demand.

Building upon this, we adopt a complementary perspective and review the literature on inventory control problems in the presence of censored demand. Within this field, two main strands of research can be identified: the parametric method and the nonparametric method. The parametric method assumes models based on specific parameter sets. While the nonparametric method does not assume an explicit parametric form of the demand model and is distribution-free. Both parametric and nonparametric approaches offer valuable insights and solutions for managing inventory. By examining and synthesizing the research within these two strands, we contribute to a deeper understanding of effective inventory control in the presence of censored demand information.

2.2.1 Parametric method

Braden and Freimer (1991) focused on a pricing problem that involves a unique nonperishable good. The problem considers customers with random reservation values, where the variance is known but the mean remains unknown. These customers appear sequentially, and the study identified various families of probability distributions that effectively capture the informational dynamics resulting from censored observations. Building upon this line of research, Lu et al. (2008) introduced a suitable notation that accurately describes the dynamics of an inventory system. They proposed a sample-path analysis method to determine the optimal inventory levels for a perishable product. They adopted a Bayesian framework to periodically update the system parameters using censored demand data. By leveraging this Bayesian approach, they explicitly characterized the trade-off involved in the inventory control problem.

2.2.2 Nonparametric method

A significant body of literature studies stochastic inventory planning problems in the context of periodic review systems with a single product. These problems entail making decisions regarding inventory replenishment and order quantities within a dynamic and uncertain environment.

2.2.2.1 Perishable products.

The distribution-free newsvendor problem was first examined by Scarf (1957), who developed a theory based on the assumption that future demand follows a similar pattern to past demand. Their seminal work considered different inventory policies and identified the policy that maximizes the minimum expected profit across all possible demand distributions. Similarly, Gallego and Moon (1993) studied the distribution-free newsvendor problem, where only the mean and variance of the demand are known. In the domain of adaptive algorithms, Godfrey and Powell (2001) introduced the Concave, Adaptive Value Estimation (CAVE) algorithm. CAVE constructs concave piece-wise linear functions iteratively using sample gradients to estimate the value function and optimize inventory decisions in the newsvendor problem. In multi-period inventory planning with censored demand, Huh et al. (2009) utilized the subgradient method. Their study focused on situations where replenishment lead time is an integer multiple of the review period. Examining stochastic periodic-review lost-sales inventory system, Huh and Rusmevichientong (2009) extended the work of Flaxman et al. (2004) to establish the asymptotic optimality of their proposed gradient-descent type AIM (Asymptotically Ideal Model) algorithm. Their algorithm achieved a regret on the order of $\sqrt{T}$. Zhang et al. (2018) studied a stochastic periodic-review lost-sales inventory system for divisible perishable products with known and fixed lifetimes. They introduced the cycle-update policy (CUP), an online learning algorithm that asymptotically approaches the best base-stock policy. The performance of the CUP algorithm was compared to the clairvoyant optimal policy, and a lower bound policy called Replacement of Old Inventories (ROI) was established. The regret of the CUP algorithm was shown to be eventually on the order of $\sqrt{T}$. Bu et al. (2023) investigated a periodic-review inventory system for perishable products, considering both lost or backlogged cases. They examined the asymptotic optimality of the base-stock policy as certain parameters approach infinity and explored the impact of various parameters on the optimality gap. The study also extended the asymptotic optimality result to alternative inventory management policies such as Last-In-First-Out (LIFO) or arbitrary issuance policies.

2.2.2.2 Nonperishable products.

Huh and Rusmevichientong (2009) also investigated nonperishable products. They explored the application of the stochastic gradient method to the waiting time process of a $GI/D/1$ queue. Their analysis establishes that the regret follows an order of $\sqrt{T}$. Furthermore, Huh et al. (2011) proposed a novel class of nonparametric adaptive data-driven policies for multiproduct stochastic inventory control problems. They utilize the product-limit form of the Kaplan-Meier estimator in their approach. Shi et al. (2016) focus on a periodic review multiproduct stochastic inventory planning problem in the presence of warehouse-capacity constraints. They proposed an asymptotically optimal subgradient algorithm. By connecting the demand process with a $GI/G/1$ queue, they proved a regret bound of $\sqrt{T}$. Gao and Zhang (2022a) investigated a periodic review multi-product inventory system that allows for stock-out substitutions or inventory assortment problems in the context of an urban warehouse. They employed the upper confidence bound (UCB) algorithm to decide the optimal inventory policy efficiently. Tang et al. (2022) proposed an online stochastic gradient descent with perturbed gradient (SGD-PG) algorithm to tackle a multi-product inventory control problem with lost sales. This algorithm jointly determines the order-up-to level and upgrading decisions to optimize inventory management. The cumulative regret bound for this algorithm is on the order of $\sqrt{T}$.

Our research aligns closely with the Huh and Rusmevichientong (2009) and Shi et al. (2016), which investigated nonperishable products and proposed online subgradient algorithms. However, key distinctions from their work exist. While Shi et al. (2016) focused on a multi-product inventory control problem with a single capacity constraint and fractional decision variables, our study considers demand correlation and integer decisions. Additionally, we explore the impact of shipping costs. In Huh and Rusmevichientong (2009), an integer decision variable problem is also considered. However, our research is much more complex due to higher dimensionality, the interdependence of demand, and shipping costs. Besides, we examine scenarios where unfilled demand can be backlogged, in addition to the lost sales cases considered in Huh and Rusmevichientong (2009). By incorporating these additional factors and considering various demand fulfillment options, our research extends the understanding of inventory control beyond the scope of previous studies. We provide insights into the impact of demand correlation, integer decision-making, and shipping costs on inventory management for nonperishable products.

2.3 Methodologies

2.3.1 Online Convex Optimization

In contrast to batch learning methods, which process data in batches to generate the best predictor, online learning is a dynamic, learning-while-doing approach where data arrives sequentially. It allows for continuous updates to improve the predictor for future data. The concept of "Online Convex Optimization" was initially introduced by Zinkevich (2003), highlighting the importance of making real-time decisions and optimizing in online learning scenarios. In this context, we will focus on one specific online convex optimization method, the online subgradient method, and provide a concise overview of its key aspects and functionality.

2.3.1.1 Online subgradient method.

The subgradient method is an unconstrained optimization technique initially developed by Naum Z. Shor and others. It is particularly useful for handling convex objective functions that may not be differentiable. While the subgradient method may be sensitive to problem scaling and generally slower compared to more modern approaches like interior-point methods and Newton's method, it offers several advantages such as a broader range of applicability and lower memory requirements (Boyd et al., 2003).

While the concept of "Online Convex Optimization" was formally introduced in 2003, the framework of the subgradient method itself predates this work. The initial appearance of the subgradient method can be traced back to Gordon (1999) (For a more comprehensive historical account, refer to Orabona (2019) and Hazan (2022)). The online subgradient method, also known as online subgradient descent, is a versatile online learning algorithm designed to minimize convex loss functions, even those that may be non-differentiable (Boyd et al., 2003; Hazan, 2022). By employing subgradients, which are generalizations of derivatives, this method enables us to optimize efficiently in the online learning setting.

In the study by Shalev-Shwartz et al. (2007), the projected subgradient method is employed to address a problem formulated by Support Vector Machines (SVMs). They establish that this method requires approximately $\tilde{O}(1/\epsilon)$ iterations to achieve a solution with an accuracy of ϵ. This result demonstrates the efficiency of the projected subgradient method in solving SVM-related problems. Shalev-Shwartz and Singer (2007)) also investigate a strongly convex repeated game scenario, where predictors select a sequence of vectors from a strongly convex set. They propose a family of prediction algorithms that achieve a logarithmic regret bound of the form $O(\frac{\log(T)L}{\sigma})$, where σ and L are independent of the time horizon T. This finding highlights the effectiveness of these algorithms in minimizing regret in repeated game settings. Ratliff et al. (2007) propose the use of simple subgradient-based algorithms to solve structured learning problems in the online learning setting. They demonstrate sublinear regret. Orabona et al. (2012) develop a proximal projected subgradient descent algorithm and establish a logarithmic regret bound (Lemma 2) for this approach. Akbari et al. (2015) introduce a distributed online subgradient push-sum algorithm. They prove a sublinear regret bound for individual agents within this distributed framework. Cao and Başar (2020) propose an event-triggered decentralized online subgradient descent algorithm in the full information setting. They establish a sublinear regret bound for this algorithm and further extend their analysis to the problem with bandit feedback, showing a similar regret bound in terms of order of magnitude. These algorithms prove to be efficient and effective in real practice.

2.3.2 Regret Analysis

Regret is used to quantify the cost difference between the algorithm's performance and the optimal performance achievable with full information about the underlying model, known as the oracle. In this part, we delve into the literature on operations management that focuses on providing rigorous proof for both the upper and the lower bounds of regret.

By analyzing the upper bounds, researchers can guarantee the algorithm's performance, demonstrating that it performs reasonably well even without full information. On the other hand, the lower bounds offer a benchmark for the best possible performance that any algorithm can achieve, revealing the inherent limitations imposed by the problem structure and the available information.

2.3.2.1 Regret upper bound.

The pioneering works of Zinkevich (2003) and Hazan et al. (2007) were among the first to provide upper bounds for regret in the context of online convex optimization problems. In particular, Zinkevich (2003) introduced an inequality that connects decisions made in consecutive periods, enabling them to establish the $O(\sqrt{T})$ regret bound. This result was achieved by canceling out terms in the summand and leveraging the relationship between integration and summation. Building upon the methodology proposed by Zinkevich (2003), Huh and Rusmevichientong (2009) and Shi et al. (2016) derived regret upper bounds for their specific inventory management problems. They also employed a $G/G/1$ queue to establish an upper bound on the difference between the *actual implemented* order-up-to level and the *target* order-up-to level. We relax the assumption utilized in the aforementioned studies. A comprehensive review of the problem setting presented in the latter two papers can be found in Section 2.2.2, where we provide a detailed analysis of their methodologies and highlight the distinctions and contributions of our work.

2.3.2.2 Regret lower bound.

There are two commonly employed information-theoretical approaches to establish lower bounds on regret.

Kullback-Leibler divergence. The Kullback-Leibler divergence is widely used in information theory to quantify the difference between two probability distributions. It provides a way to compare the information content of one distribution with another. In the context of regret analysis, researchers have employed the Kullback-Leibler divergence to establish lower bounds on regret in various pricing and inventory management problems. Here are some examples: In their work, Besbes and Zeevi (2011) investigated a pricing problem where the customer's willingness-to-pay distribution undergoes a single change during the selling season. The seller knows the pre- and post-change distributions but does not know when the change occurs. They demonstrated that any pricing policy will inevitably make errors, which are measured in terms of regret. They proved that the best achievable regret for any policy is lower bounded by $\Omega(\sqrt{N})$ by using the Kullback-Leibler divergence, where N represents the total number of sequentially arrived customers. Similarly, Broder and Rusmevichientong (2012) delved into a dynamic pricing problem where a monopolist sets the price for a product offered to a total number of T customers. Customers' purchasing decisions are influenced by a general parametric choice model, the parameter of which is unknown to the seller. They established that the worst-case regret is lower bounded by $\Omega(\sqrt{T})$ in the general case. They also presented numerical evidence demonstrating the good performance of their pricing policies. Additionally, Besbes and Muharremoglu (2013) considered a repeated newsvendor problem, where the seller does not know the underlying demand distribution. They established lower bounds $\Omega(\log T)$ on regret for continuous and discrete demand distributions. The Kullback-Leibler divergence is employed in their analysis to gain insights into the information gap between the two specific probability distributions. By quantifying the difference between these distributions, researchers demonstrated that any algorithm would struggle to distinguish between them, leading to the errors mentioned earlier. We adopt a similar approach and utilize the Kullback-Leibler divergence to establish lower bounds on regret. The intuition behind this choice is elaborated explicitly in Section 9.2.2, where we provide a novel construction and explain in detail how the Kullback-Leibler divergence is a valuable tool in our analysis.

Van Tree inequality. In addition to Kullback-Leibler divergence, the van Tree inequality–also known as Bayesian Cramer-Rao lower bound–is utilized in certain studies to provide information-theoretical proofs of lower bounds on regret. In the case of Broder and Rusmevichientong (2012), they not only considered the general scenario but also examined a "well-separated" case where the regret is lower bounded by $\Omega(\log T)$. This case involves specific conditions that lead to a different regret lower bound. Furthermore, Keskin and Zeevi (2014) investigated a monopolist selling a set of products within a finite planning horizon T. The seller initially does not know the parameters of the linear demand curve. As demand becomes realized over time, the seller incrementally learns the demand parameters and exploits the profit accordingly. They demonstrated that when the seller has no prior information about the demand parameters, the smallest achievable regret relative to an oracle is lower bounded by $\Omega(\sqrt{T})$. However, if the seller knows the expected demand under an incumbent price, the smallest achievable regret becomes lower bounded by $\Omega(\log T)$. Both papers, following Goldenshluger and Zeevi (2009), utilize the van Tree inequality to provide information-theoretical proofs for these lower bounds. The van Tree inequality is powerful in statistics. It establishes a connection between the Fisher information and the divergence between two probability distributions. With its help, researchers can quantify the information gap and rigorously derive lower bounds on regret.

2.3.3 Discrete Convexity

Discrete convex analysis, proposed by K. Murota in 1996, is a novel paradigm for discrete optimization that brings ideas from continuous optimization and combinatorial optimization (Murota, 2003). This method has been studied and applied extensively in the operations management and operations research literature (Simchi-Levi et al., 2005; Chen, 2017; Chen and Li, 2021).

The two most commonly used concepts in discrete convex analysis are $L^\natural$-*convexity* and $M^\natural$-*convexity*. These two concepts lead to a broader class known as *convex-extensibility*. They are all important in analyzing the convexity property in discrete settings, thus enabling development of effective optimization techniques for our research objectives.

2.3.3.1 $L^\natural$-convexity.

The concept of $L^\natural$-convexity is generalized from mid-point convexity. In the case of mid-point convexity, a function g in $\mathbb{R}^n$ satisfies the condition $\frac{g(p)+g(q)}{2} \geq g(\frac{p+q}{2})$, or equivalently, $g(p) + g(q) \geq g(\frac{p+q}{2}) + g(\frac{p+q}{2})$. It is well-known that mid-point convexity is equivalent to convexity when g is a continuous function. Similarly, for a discrete function g, if it satisfies $g(p)+g(q) \geq g(\lfloor \frac{p+q}{2} \rfloor)+g(\lceil \frac{p+q}{2} \rceil)$ for $p, q \in \mathbb{Z}^n$, then we refer to g as an $L^\natural$-convex function. In their work, Moriguchi and Murota (2012) presented the necessary and sufficient conditions for $L^\natural$-convexity utilizing a generalized Hessian matrix. Furthermore, in their paper, Chen (2017) comprehensively surveyed the properties and applications of $L^\natural$-convexity. See their paper for a comprehensive survey.

2.3.3.2 $M^\natural$-convexity.

$M^\natural$-convexity is a concept derived from equidistrance convexity. A function g is considered equidistance-convex if it satisfies the condition $g(p)+g(q) \geq g(p-\alpha(p-q))+g(q+\alpha(p-q))$ for every $\alpha \in [0,1]$. $M^\natural$-convexity builds upon this idea by moving a pair of points (p,q) to another (p',q') along the coordinate axes instead of the connecting line segment (Murota, 1998). To define $M^\natural$-convexity, we introduce the following notation: the positive support of a point $z \in \mathbb{Z}^n$ is denoted as $supp^+(z) = \{i : z_i > 0\}$, and the negative support of z is denoted as $supp^-(z) = \{i : z_i < 0\}$. Additionally, let e_i be the n-dimensional unit vector with the ith position being 1 and $e_0 = (0, \cdots, 0)$. A function g is said to be $M^\natural$-convex if, for any pair of $p, q \in \text{dom}_{\mathbb{Z}} g$ and any $i \in supp^+(p-q)$, there exists $j \in supp^-(p-q) \bigcup \{0\}$ such that $g(p) + g(q) \geq g(p - e_i + e_j) + g(q + e_i - e_j)$. Notably, when g is defined on $\mathbb{Z}^2$, $M^\natural$-convexity is equivalent to multimodularity (Hajek, 1985). Furthermore, $M^\natural$-convex functions are *supermodular* (a function g is supermodular if $-g$ is submodular). In their work, Moriguchi and Murota (2012) proposed the necessary and sufficient conditions for $M^\natural$-convex functions. They provided a comprehensive understanding of this concept and its properties. See Chen and Li (2021) for a comprehensive survey of applications.

2.3.3.3 Convex-extensibility.

In the context of discrete convex analysis, a function $g : \mathbb{Z}^n \to \mathbb{R}$ is considered *convex-extensible* if there exists a function $\bar{g} : \mathbb{R}^n \to \mathbb{R}$ such that $g(p) = \bar{g}(p)$ for $p \in \mathbb{Z}^n$. The function $\bar{g}$ is called a *convex extension* of g (Murota, 2003). This concept enables us to extend a discrete function defined on integer points to a continuous function defined on real-valued points. By establishing such convex extensions, it becomes possible to apply convex analysis techniques and leverage the properties of continuous convex functions to analyze and optimize discrete functions. It was established by Murota that an $L^\natural$-convex function and $M^\natural$-convex function are convex-extensible (Murota, 2003, 2018).

Chapter 3

Problem Formulation

3.1 Deterministic Demand

In the first part of the study, we delve into a deterministic model for quantitative results in closed forms and managerial insights. We will first reproduce results for a single-product problem which is a minor modification of the EOQ with shipping cost incorporated. Then, we look into the two-product inventory control problem where demand exhibits complementarity. In both single-product and two-product problems, we differentiate between the cases where unavailable items are backlogged and lost. In each of the scenarios mentioned above, we solve the optimal order-up-to level(s) and optimal reorder cycle length sequentially.

In order to solve the problem, we conduct the following procedures.

First, we formulate the demand as a fluid process and set rules on the order fulfillment mechanism. We assume that the firm promises a high-quality service and will ship the items required by customers immediately upon the order placement. In the single-product case, the problem is a generalized EOQ model. While in the two-product case, the problem is much more complicated because we consider a bound shipping strategy adopted by the carrier (the firm), i.e., if the order placed by the customer consists of two items, then the order will be bound into a single package and shipped to the customer upon order placement when both products are available, or until inventory replenishment when both products are unavailable. Note that in the case when the order consists of two items while only one of the items is available, the available item is shipped immediately upon order placement and the unavailable item will be shipped upon inventory replenishment.

Second, we establish parameters like demand rate, holding cost (rate), backorder cost (rate), lost sales cost, and shipping cost and develop the cost function that the firm should optimize in a single inventory replenishment cycle. We solve the optimal order-up-to levels (given exogenous reorder cycle length) and the optimal reorder cycle length (when it is endogenous) sequentially.

Finally, by solving the firm's cost minimization problem, we will arrive at critical observations and important conclusions, which show the necessity of building a new model when shipping cost is included. Moreover, by doing numerical studies, we demonstrate the significance of considering shipping costs when making inventory decisions.

3.1.1 The Economic Order Quantity Model for a Single Product

Here we reproduce the classical economic order quantity (EOQ) results according to our problem background where the shipping cost is included in the total cost calculation. We consider an e-commerce firm's problem of planning inventory for a single product where the relevant cost includes ordering cost, inventory holding cost, backorder or lost-sales cost, and shipping cost. Demand arrives continuously and deterministically at a constant rate λ. A linear inventory holding cost is accrued for every unit held in inventory per unit of time at a rate of h. We consider two different settings when demand arrives but there is no available inventory: the demand is backlogged or lost. In the case of backlog, demand that cannot be satisfied immediately upon its arrival is backlogged and will be fulfilled upon inventory replenishment. A linear backlog cost is accrued for every unit backlogged per unit time at a rate of b. In the case of lost sales, demand that cannot be satisfied immediately upon its arrival is lost with a cost p incurred for every unit lost, referred to as the *lost sales cost*. Since we consider an e-commerce background, there is also a shipping and delivery cost s, also referred to as *shipping cost* in short in the following context, for each unit of demand satisfied. Furthermore, a fixed setup cost of S, also referred to as the ordering cost, is incurred for each replenishment order placed by the firm. We assume the lead time, i.e., the time that elapses between the placement and the receipt of an order, is zero. [1] Because shipping is considered a separate factor in our model, we define the lost sales cost p as the loss of profit, customer loyalty, brand reputation, and so on without the shipping cost being subtracted. That is, the actual loss of profit for every unit demand lost is $p - s$ because shipping is no longer needed in this case.

The firm's problem is to find the optimal ordering policy to minimize the total cost per unit of time in an infinite time horizon. It is easy to see that the optimal policy should exhibit periodicity (e.g., see Figures 4.1 and 4.2) because the problem is stationary and for an infinite-time horizon. As illustrated by the figures, a stationary policy can be characterized by a pair of decision variables (T, Y), where T is the cycle time (the time

[1] Note that we make this zero-lead-time assumption for ease of exposition. It can be relaxed in our model (with deterministic demand).

between two successive replenishments) and Y is the order-up-to level. Define $t := \frac{Y}{\lambda}$, which denotes the time it takes to sell all inventory in a cycle if $t \leq T$. Then, an ordering policy can be equivalently specified by variables (T, t), with which the order-up-to level can be calculated by $Y = \lambda t$. Furthermore, let X denote the inventory position at the end of a cycle and Q denote the order quantity.

3.1.2 The Two-Product Model

Consider a scenario where the e-commerce firm is selling two related products that a significant proportion of customers often buy together. We keep using the EOQ framework we introduced in the last section with the following adaptation.

Demand arrives at the firm in the format of customer orders, which can be of three types: type-1, type-2, and type-12. Specifically, a type-i order requests one unit of product i only, $i = 1, 2$, and a type-12 order requests one unit of both products. Different types of orders arrive continuously and deterministically at constant rates λ_1, λ_2, and λ_{12}, respectively. Let $\hat{\lambda}_i := \lambda_i + \lambda_{12}$ denote the effective demand rate of product-i, $i = 1, 2$.

The firm is committed to providing high-quality logistics services and hence adopts an instant shipping policy. That is, the available part of an order is shipped **immediately** and the unavailable part is backlogged until replenishment (or lost in the case where backorder is not accepted). Furthermore, the firm adopts a consolidated/grouped shipping practice that is commonly used in e-commerce: available items in an order are grouped and shipped in a **single** package to save cost without sacrificing the service quality. For example, if a type-12 order arrives and both products are available, then one unit of both products will be shipped to the customers immediately in one package. If a type-12 order arrives but only product 1 is available, then one unit of product 1 will be shipped immediately to the customer and the demand for product 2 would be backlogged and shipped separately to the customer (or lost if backorders are not allowed).

The shipping cost s is charged per package. The firm orders the two products from the same supplier and thus would be able to replenish both via a single order at a fixed setup cost S. We keep the zero-lead time assumption for ease of exposition. A subscript i is added to the notation to denote the cost parameters and decision variables for each product, e.g., h_i, b_i, p_i, Y_i, etc.

3.2 Stochastic Demand

This section describes our general model with stochastic demand and introduces the basic notation. Consider an e-commerce company, referred to as the *firm*, that sells two different products indexed by $i = 1, 2$, to customers over finite discrete time periods indexed by $t = 1, 2, \cdots, T$.

The Demand Process. The firm receives customer demand in the form of orders, which can be of three different types: type-i (where $i = 1, 2$) and type-12. A type-i order requests one unit of product i only (where $i = 1, 2$), while a type-12 order requests one unit of each product. Customers' demands (orders) arrive randomly and sequentially. In period t, let $r_{t,k} \in \{1, 2, 12\}$ denote the type of the kth arriving order and D_t^{Σ} denote the total number of orders. The demand in period t is fully described by the sequence

$$\mathcal{D}_t^{Seq} := \{r_{t,1}, r_{t,2}, \cdots, r_{t,k}, \cdots, r_{t,D_t^{\Sigma}}\}.$$

Let $D_{t,r}$ denote the total number of type-r orders, where $r \in \{1, 2, 12\}$, in period t, and $\vec{D}_t := (D_{t,1}, D_{t,2}, D_{t,12})^T$ represent the vector of joint demand. The demand process has the following properties. First, the vectors $\vec{D}_t$'s are independent and identically distributed over different periods according to a cumulative distribution function $F(\cdot)$ for $\vec{D}_t$ and marginal distribution functions $F_r(\cdot)$ for $D_{t,r}$, where $r = 1, 2, 12$. Second, given that a set of orders arrive in a period, their arrival sequence is uniformly distributed, meaning that any sequence occurs with an equal probability.

For ease of notation, let $\vec{D} = (D_1, D_2, D_{12})$ and $\mathcal{D}^{seq}$ denote generic variables for one-period demand. Additionally, let $\hat{D}_{t,i} := D_{t,i} + D_{t,12}$ for $i = 1, 2$ represent the total demand for product i in period t, and let $\hat{D}_i$ be the corresponding variable for one-period demand of product i.

In the following context, we use i or j (where $i, j \in \{1, 2\}$) to represent the index of the product unless otherwise specified. If both i and j are present simultaneously, they are considered distinct.

Sequence of events. The sequence of events in a period is illustrated in Figure 3.1. At the beginning of period t, the firm reviews the inventory position $\vec{I}_{t-1} := (I_{t-1,1}, I_{t-1,2})^T$ carried from the end of last period. Note that for product i, $I_{t-1,i}^{+}$ represents its left inventory, and $I_{t-1,i}^{-}$ represents the number of backorders if backlogging is allowed. Then, the firm decides the ordering quantities, denoted by $\vec{Q}_t = (Q_{t,1}, Q_{t,2})^T$, of the two products, raising the inventory position to $\vec{Y}_t = \vec{I}_{t-1} + \vec{Q}_t$. There is no replenishment lead time. Backorders, if any, will be satisfied and shipped immediately after replenishment at the beginning of the next period. During the period t, the demand is realized and the customers' orders are fulfilled by express shipping according to the partial fulfillment policy with either backlog or lost sales. At the end of the period, the inventory position drops to $I_{t,i} = Y_{t,i} - \hat{D}_{t,i}$ in the backlog case or $I_{t,i} = (Y_{t,i} - \hat{D}_{t,i})^{+}$ in the lost-sales case. Note that the event in the dashed box in Figure 3.1 applies to the backlog case only.

Costs. Products are nonperishable, and it costs the firm h_i to carry each unit of product i from one period to another. When stockout occurs, we consider two different assumptions separately in our paper: the unsatisfied part of an order is *backlogged* or

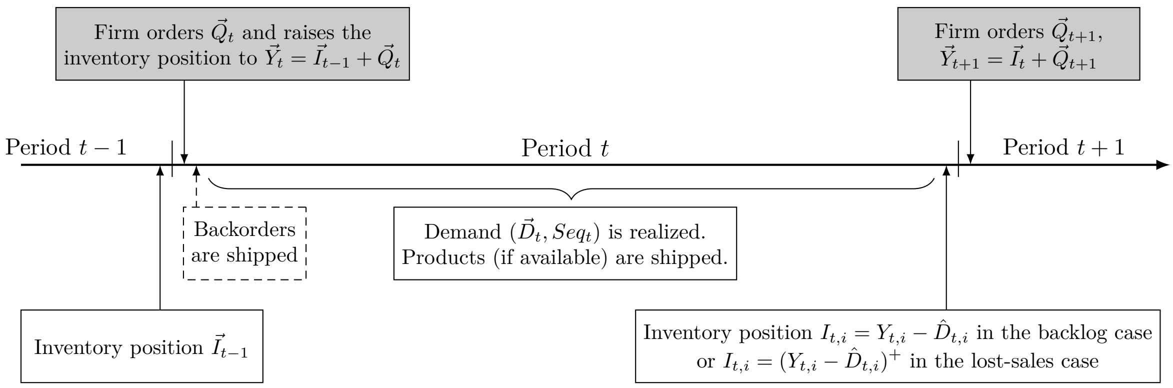


Figure 3.1: Sequence of events

lost. Each unit of demand for product i that is backlogged at the end of a period costs the firm b_i (in the case of backlog). Each unit of lost demand for product i costs the firm p_i (in the case of lost sales). Each package shipped costs s regardless of the number of items it includes.

The firm's problem is to decide the ordering quantity $\vec{Q}_t$ in each period to minimize the expected total cost, which includes inventory holding, backlog or lost sales, and shipping costs. This decision can be equivalently transformed to the beginning inventory position $\vec{Y}_t$.

Part II

Deterministic-Demand Models

Chapter 4

Single-product Problem with Deterministic Demand (The EOQ Model)

In this chapter, we reproduce the results of EOQ model using our notations for ease of comparison between the single-product case and two-product cases.

4.1 The Case of Backlog

If demand that occurs during the stockout period is backlogged, the change in inventory position over time can be illustrated in Figure 4.1. The firm's average cost is represented by

$$\pi_1^b(t, T) = \frac{1}{T}\left[\frac{1}{2}h\lambda t^2 + \frac{1}{2}b\lambda(T-t)^2 + S\right] + s\lambda,$$

for $t \leq T$ (i.e., $Y \leq \lambda T$). Note here that we constrain Y to be no larger than λT and t to be no larger than T because holding positive inventory at the end of a cycle under deterministic demand is suboptimal.. We solve the problem sequentially by optimizing t and T.

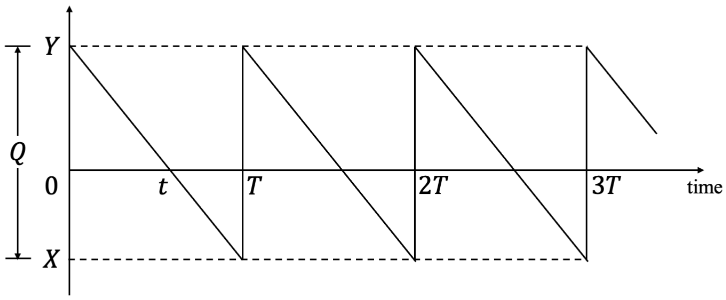


Figure 4.1: Inventory position as a function of time in backlog case

First, given a fixed reorder cycle length T, the optimal soldout time $\tilde{t}^b(T) = \arg\min_{0\le t\le T}\{\pi_1^b(t,T)\}$ and hence the optimal order-up-to level $\tilde{Y}^b(T) = \lambda\tilde{t}^b(T)$ can be obtained by the first-order condition:

$$\tilde{t}^b(T) = \frac{bT}{h+b} \quad \text{and} \quad \tilde{Y}^b(T) = \frac{b\lambda T}{h+b}.$$

As one can see, the expression reveals a Newsvendor-like insight. The optimal beginning inventory, which is proportional to the ratio $\frac{b}{h+b}$, is set to balance the underage and overage costs in a cycle.

Next, we plug $\tilde{t}^b(T)$ into the cost function and solve the optimal cycle time T^b:

$$T^b = \sqrt{\frac{2(h+b)S}{hb\lambda}}, \qquad t^b = \frac{bT^b}{h+b} = \sqrt{\frac{2bS}{h(h+b)\lambda}},$$

$$Y^b = \frac{b\lambda T^b}{h+b} = \sqrt{\frac{2b\lambda S}{h(b+h)}}, \qquad Q^b = \lambda T^b = \sqrt{\frac{2(h+b)S\lambda}{hb}},$$

$$X^b = -\frac{h\lambda T^b}{h+b} = \sqrt{\frac{2h\lambda S}{b(b+h)}}.$$

The solution is known as EOQ with backorders, where the trade-off between balancing setup costs and variable costs is similar to the classical model without backorders. The key difference is that the rate of variable costs is now calculated as the average of the inventory holding and backlog costs over a cycle.

Lastly, note that the shipping cost s is irrelevant in this problem, as all customer orders are shipped once and only once–either immediately upon demand arrival or upon replenishment.

4.2 The Case of Lost Sales

If demand that arrives during the stockout period is lost, the evolution of the inventory position over time can be illustrated in Figure 4.2. The firms average cost is given by

$$\pi_1^\ell(t,T) = \frac{1}{T}\left[\frac{1}{2}h\lambda t^2 + s\lambda t + p\lambda(T-t) + S\right],$$

for $t \le T$ (i.e., $Y \le \lambda T$). We solve the problem by optimizing t and T sequentially.

First, given a fixed reorder cycle length T, the optimal soldout time $\tilde{t}^\ell(T) =$

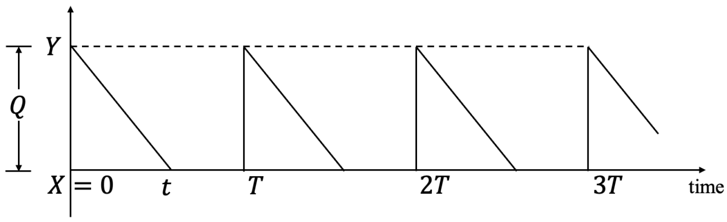


Figure 4.2: Inventory position as a function of time in lost sales case

$\arg\min_{0\le t\le T}\{\pi_1^\ell(t,T)\}$ can be obtained by the first-order condition:

$$\tilde{t}^\ell(T) = \begin{cases} T, & \text{if } T \le \frac{p-s}{h}, \\ \frac{p-s}{h}, & \text{if } T > \frac{p-s}{h}, \end{cases}$$

$$\tilde{Y}^\ell(T) = \min\left\{\lambda T, \frac{(p-s)\lambda}{h}\right\} = \begin{cases} \lambda T, & \text{if } T \le \frac{p-s}{h}, \\ \frac{(p-s)\lambda}{h}, & \text{if } T > \frac{p-s}{h}. \end{cases}$$

That is, it is optimal to stock the exact amount (i.e., let $\tilde{Y}^\ell = \lambda T$) to satisfy all demand in a cycle if T is below a threshold and stock a maximum of $\frac{(p-s)\lambda}{h}$ if T is too large. The reason is as follows. The inventory that is sold right after time t has been carried for time t with inventory cost ht incurred. Further increasing the inventory by one unit means that the firm will carry the additional inventory for time t to reduce one unit of lost sales. Therefore, it is profitable for the firm to do so if and only if the gain $p-s$ can cover the cost ht.

Next, we endogenize the reorder cycle length and solve it optimally. We plug $\tilde{t}^\ell(T)$ into the cost function and obtain that

$$\pi_1^\ell(T) = \begin{cases} \frac{\lambda}{T}\left[\frac{1}{2}hT^2 - (p-s)T + \frac{S}{\lambda}\right] + p\lambda, & \text{if } T \le \frac{p-s}{h}, \\ \frac{\lambda}{T}\left[-\frac{(p-s)^2}{2h} + \frac{S}{\lambda}\right] + p\lambda, & \text{otherwise.} \end{cases}$$

If $\frac{S}{\lambda} > \frac{(p-s)^2}{2h}$, the average cost $\pi_1^\ell(T)$ is decreasing in T and hence the optimal cycle time is $T^\ell = \infty$. If $\frac{S}{\lambda} = \frac{(p-s)^2}{2h}$, the average cost $\pi_1^\ell(T)$ is decreasing in T for $T < \frac{p-s}{h}$ and is 0 for $T \ge \frac{p-s}{h}$. If $\frac{S}{\lambda} < \frac{(p-s)^2}{2h}$, the average cost $\pi_1^\ell(T)$ is decreasing in T for $T < \sqrt{\frac{2S}{\lambda h}}$ and increasing in T for $T \ge \sqrt{\frac{2S}{\lambda h}}$. Therefore, the optimal solution is given by

$$T^\ell = \begin{cases} \sqrt{\frac{2S}{\lambda h}}, & \text{if } S < \frac{\lambda(p-s)^2}{2h}, \\ \infty, & \text{if } S \ge \frac{\lambda(p-s)^2}{2h}, \end{cases}$$

and

$$Y^\ell = \begin{cases} \sqrt{\frac{2S\lambda}{h}}, & \text{if } S < \frac{\lambda(p-s)^2}{2h}, \\ 0, & \text{if } S \ge \frac{\lambda(p-s)^2}{2h}. \end{cases}$$

The case of $S \geq \frac{\lambda(p-s)^2}{2h}$ corresponds to the situation where the business is not profitable and is trivial. In the following, we only consider the case where $S < \frac{\lambda(p-s)^2}{2h}$. In this case, the optimal solution is the classical EOQ. That is, the firm orders $Q^\ell = \sqrt{\frac{2S\lambda}{h}}$ whenever the inventory drops to 0. The reorder cycle time is $\sqrt{\frac{2S}{\lambda h}}$.

Lastly, the shipping cost s is also irrelevant in the problem except that we require $S < \frac{\lambda(p-s)^2}{2h}$. This is because the amount of inventory is planned to satisfy all demand (which is assumed to be known and deterministic) without lost sales.

Chapter 5

Two-Product Problem with Deterministic Demand and Partial Fulfillment

In this chapter, we consider the cases where unfulfilled demand may either be backlogged or lost. We will show that the stockout of one product does not influence the demand or shipping for the other one in an order.

5.1 The Case of Backlog

First, we consider the backlog case. In this situation, if an order requiring both products arrives but only one is available, only the unavailable product will be backlogged and shipped to the customer upon inventory replenishment, potentially resulting in two separate packages shipped. Thus, we refer to this case as "unfulfilled demand backlogged".

Let π_2^{ub} denote the firm's average cost over time, where the subscript "2" means two-product problem and the superscript "ub" means "unfulfilled demand backlogged". Suppose the backorder cost is applied to each product independently. The firm's average cost is given by

$$\pi_2^{ub}(t_1, t_2, T) = \frac{1}{T}\left[\frac{1}{2}h_1\hat{\lambda}_1 t_1^2 + \frac{1}{2}b_1\hat{\lambda}_1(T - t_1)^2 + \frac{1}{2}h_2\hat{\lambda}_2 t_2^2 + \frac{1}{2}b_2\hat{\lambda}_2(T - t_2)^2 \right.$$
$$\left. + s(\lambda_1 + \lambda_2 + \lambda_{12})T + s\lambda_{12}|t_2 - t_1| + S\right],$$

for $t_1, t_2 \leq T$ (i.e., $Y_1 \leq \hat{\lambda}_1 T, Y_2 \leq \hat{\lambda}_2 T$). Note that it is not profitable for the firm to hold the inventory of product i for more than $\hat{\lambda}_i T$ under deterministic demand. We assume that $\frac{b_1}{h_1+b_1} \leq \frac{b_2}{h_2+b_2}$ without loss of generality, and solve t_i and T sequentially to optimality.

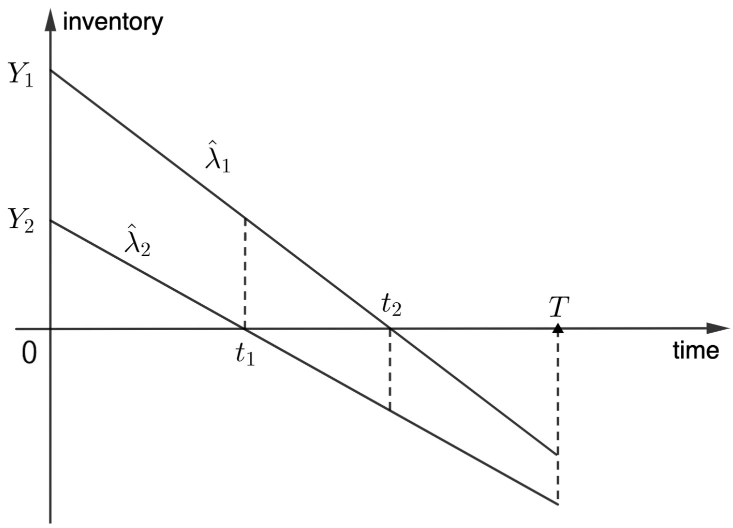


Figure 5.1: Fluid model for the two-product problem with partial fulfillment under backorder setting

From the formula of the average cost function, we know that the average cost is convex in both t_1 and t_2. So, the minimum point can be either solved by using first-order conditions or obtained at the boundary. Hence, we can present the main result as follows. Let

$$\tau^{ub} := \lambda_{12}s \left[\frac{1}{\hat{\lambda}_1(h_1+b_1)} + \frac{1}{\hat{\lambda}_2(h_2+b_2)}\right] \left[\frac{b_2}{h_2+b_2} - \frac{b_1}{h_1+b_1}\right]^{-1},$$

$$\tau_1^{ub}(S) := \sqrt{\frac{2S[\hat{\lambda}_1(h_1+b_1)+\hat{\lambda}_2(h_2+b_2)]}{(\hat{\lambda}_1 b_1+\hat{\lambda}_2 b_2)(\hat{\lambda}_1 h_1+\hat{\lambda}_2 h_2)}},$$

$$\tau_2^{ub}(S) := \sqrt{\frac{2S-(\lambda_{12}s)^2\left(\frac{1}{\hat{\lambda}_1(h_1+b_1)}+\frac{1}{\hat{\lambda}_2(h_2+b_2)}\right)}{\frac{\hat{\lambda}_1 h_1 b_1}{h_1+b_1}+\frac{\hat{\lambda}_2 h_2 b_2}{h_2+b_2}}}.$$

It is proved that $\tau_1^{ub}(S) \le \tau^{ub}$ if and only if $\tau_2^{ub}(S) \le \tau^{ub}$. Thus, let S^{ub}_{sync} be the unique solution of the equation $\tau_1^{ub}(S^{ub}_{sync}) = \tau^{ub}$.

Note that τ^{sb} is independent of the setup cost S because this threshold is involved in the expression of optimal decisions when the cycle length T is exogenous, where S does not play a role because we only consider a single cycle.

Proposition 1. *For a two-product fluid demand system, when the demand for unavailable items is backlogged, we have the following results.*

1. *When reorder cycle length T is exogenous, the optimal initial inventory level is*

$(\bar{Y}_1^{ub}, \bar{Y}_2^{ub}) = (\hat{\lambda}_1 \bar{t}_1^{ub}, \hat{\lambda}_2 \bar{t}_2^{ub})$, *where*

$$(\bar{t}_1^{ub},\ \bar{t}_2^{ub}) = \begin{cases} \left(\frac{\hat{\lambda}_1 b_1 + \hat{\lambda}_2 b_2}{\hat{\lambda}_1(h_1+b_1)+\hat{\lambda}_2(h_2+b_2)} T,\ \frac{\hat{\lambda}_1 b_1 + \hat{\lambda}_2 b_2}{\hat{\lambda}_1(h_1+b_1)+\hat{\lambda}_2(h_2+b_2)} T \right), & \text{if } T \le \tau^{ub}, \\ \left(\frac{b_1 T + \frac{\lambda_{12}}{\hat{\lambda}_1} s}{h_1+b_1},\ \frac{b_2 T - \frac{\lambda_{12}}{\hat{\lambda}_2} s}{h_2+b_2} \right), & \text{otherwise}. \end{cases}$$

2. *When reorder cycle length T is endogenous, the optimal solution depends on the value of S.*
   - *If $S \le S_{sync}^{ub}$ (i.e., $\tau_1^{ub}(S) \le \tau^{ub}$), then $T^{ub*} = \tau_1^{ub}(S)$ and*

   $$(Y_1^{ub*}, Y_2^{ub*}) = \left(\frac{(\hat{\lambda}_1 b_1 + \hat{\lambda}_2 b_2)\hat{\lambda}_1 \tau_1^{b}(S)}{\hat{\lambda}_1(h_1+b_1) + \hat{\lambda}_2(h_2+b_2)},\ \frac{(\hat{\lambda}_1 b_1 + \hat{\lambda}_2 b_2)\hat{\lambda}_2 \tau_1^{ub}(S)}{\hat{\lambda}_1(h_1+b_1) + \hat{\lambda}_2(h_2+b_2)} \right).$$

   Note that $t_1^{ub*} = t_2^{ub*}$ in this case.
   - *If $S > S_{sync}^{ub}$ (i.e., $\tau_2^{ub}(S) > \tau^{ub}$), then $T^{ub*} = \tau_2^{ub}(S)$ and*

   $$(Y_1^{ub*}, Y_2^{ub*}) = \left(\frac{\hat{\lambda}_1 b_1 \tau_2^{ub}(S) + \lambda_{12} s}{h_1 + b_1},\ \frac{\hat{\lambda}_2 b_2 \tau_2^{ub}(S) - \lambda_{12} s}{h_2 + b_2} \right).$$

The solution implies that if T is small, it is optimal for the firm to synchronize the sales of both products, i.e., let $\bar{t}_1^{ub} = \bar{t}_2^{ub}$, so that all orders that demand both products do not split and hence the shipping cost can be minimized. Otherwise, if T is sufficiently large, the firm will choose a $\bar{t}_1^{ub}$ larger ($\bar{t}_2^{ub}$ smaller) than $\frac{b_1 T}{h_1+b_1}$ ($\frac{b_2 T}{h_2+b_2}$), which is the optimal solution obtained by solving a single-product problem. That is, compared with a solution obtained by solving separate single-product models, the optimal multi-product solution sets a higher initial inventory for product 1 and a lower initial inventory for product 2 while it still keeps $\bar{t}_1^{ub} < \bar{t}_2^{ub}$. The threshold τ^{ub} is solved by the boundary condition $\frac{b_1 \tau^{ub} + \frac{\lambda_{12}}{\hat{\lambda}_1} s}{h_1+b_1} = \frac{b_2 \tau^{ub} - \frac{\lambda_{12}}{\hat{\lambda}_2} s}{h_2+b_2}$. For all values of T, the firm's optimal decisions are different from those obtained by solving two single-product problems separately.

5.2 The Case of Lost Sales

In this case, the demand for unavailable products will be lost. The lost sales cost is applied to each product independently. The firm's average cost is given by

$$\pi_2^{u\ell} = \frac{1}{T}\Bigg[\frac{1}{2} h_1 \hat{\lambda}_1 t_1^2 + p_1 \hat{\lambda}_1 (T - t_1) + \frac{1}{2} h_2 \hat{\lambda}_2 t_2^2 + p_2 \hat{\lambda}_2 (T - t_2) \\ + s\lambda_{12} \max\{t_1, t_2\} + s\lambda_1 t_1 + s\lambda_2 t_2 + S\Bigg],$$

for $t_1, t_2 \leq T$ (i.e., $Y_1 \leq \hat{\lambda}_1 T, Y_2 \leq \hat{\lambda}_2 T$). We assume $\frac{p_1-s}{h_1} \leq \frac{p_2-s}{h_2}$ and solve t_i and T sequentially.

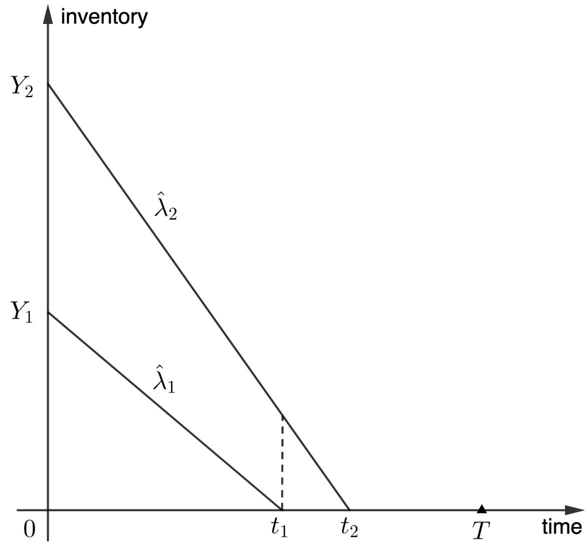


Figure 5.2: Fluid model for the two-product problem with partial fulfillment under lost sales setting

Before we present the main results, we first define thresholds as follows.

$$\tau_0^{u\ell} := \frac{p_1\hat{\lambda}_1 + p_2\hat{\lambda}_2 - s(\lambda_1 + \lambda_2 + \lambda_{12})}{h_1\hat{\lambda}_1 + h_2\hat{\lambda}_2},$$

$$\tau_1^{u\ell} := \frac{p_1 - s\frac{\lambda_1}{\hat{\lambda}_1}}{h_1},$$

$$\tau_2^{u\ell} := \frac{p_2 - s}{h_2},$$

$$\tau^{u\ell}(S) := \sqrt{\frac{2S - \frac{\hat{\lambda}_1(p_1 - s\frac{\lambda_1}{\hat{\lambda}_1})^2}{h_1}}{h_2\hat{\lambda}_2}},$$

$$S_{thold}^{u\ell} := \frac{[p_1\hat{\lambda}_1 + p_2\hat{\lambda}_2 - s(\lambda_1 + \lambda_2 + \lambda_{12})]^2}{2(h_1\hat{\lambda}_1 + h_2\hat{\lambda}_2)}.$$

Proposition 2. *For a two-product fluid demand system, when the demand for unavailable items is lost, we have the following results.*

1. *When reorder cycle length T is exogenous, if the parameters satisfy the condition $\tau_1^{u\ell} < \tau_2^{u\ell}$, then the optimal inventory level is $(Y_1^{u\ell*}, Y_2^{u\ell*}) = (\hat{\lambda}_1 t_1^{u\ell*}, \hat{\lambda}_2 t_2^{u\ell*})$, where*

$$(\bar{t}_1^{u\ell}, \bar{t}_2^{u\ell}) = \begin{cases} (T, T), & \text{if } T \leq \tau_1^{u\ell}, \\ (\tau_1^{\ell}, T), & \text{if } \tau_1^{u\ell} < T \leq \tau_2^{u\ell}, \\ (\tau_1^{\ell}, \tau_2^{\ell}), & \text{otherwise.} \end{cases}$$

If the parameters satisfy condition $\tau_1^{u\ell} \geq \tau_2^{u\ell}$, *then*

$$(\bar{t}_1^{u\ell}, \bar{t}_2^{u\ell}) = \begin{cases} (T, T), & \text{if } T \leq \tau_0^{u\ell}, \\ (\tau_0^{u\ell}, \tau_0^{u\ell}), & \text{otherwise.} \end{cases}$$

2. *When reorder cycle length* T *is endogenous, if the parameters satisfy the condition* $\tau_1^{u\ell} < \tau_2^{u\ell}$, *then the optimal reorder cycle length is*

$$T^{u\ell*} = \begin{cases} \tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}, & \text{if } \tau^{u\ell}(S) \leq \tau_1^{u\ell}, \\ \tau^{u\ell}(S), & \text{if } \tau_1^{u\ell} < \tau^{u\ell}(S) \leq \tau_2^{u\ell}, \\ \infty, & \text{otherwise.} \end{cases}$$

Accordingly, the optimal inventory level is

$$(Y_1^{u\ell*}, Y_2^{u\ell*}) = \begin{cases} (\hat{\lambda}_1\tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}, \hat{\lambda}_2\tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}), & \text{if } \tau^{u\ell}(S) \leq \tau_1^{u\ell}, \\ (\hat{\lambda}_1\tau_1^{u\ell}, \hat{\lambda}_2\tau^{u\ell}(S)), & \text{if } \tau_1^{u\ell} < \tau^{u\ell}(S) \leq \tau_2^{u\ell}, \\ (\hat{\lambda}_1\tau_1^{u\ell}, \hat{\lambda}_2\tau_2^{u\ell}), & \text{otherwise.} \end{cases}$$

if the parameters satisfy the condition $\tau_1^{u\ell} \geq \tau_2^{u\ell}$, *then the optimal reorder cycle length is*

$$T^{u\ell*} = \begin{cases} \tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}, & \text{if } S \leq S_{thold}^{u\ell}, \\ \infty, & \text{otherwise.} \end{cases}$$

Accordingly, the optimal inventory level is

$$(Y_1^{u\ell*}, Y_2^{u\ell*}) = \begin{cases} (\hat{\lambda}_1\tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}, \hat{\lambda}_2\tau_0^{u\ell}\sqrt{\frac{S}{S_{thold}^{u\ell}}}), & \text{if } S \leq S_{thold}^{u\ell}, \\ (\hat{\lambda}_1\tau_0^{u\ell}, \hat{\lambda}_2\tau_0^{u\ell}), & \text{otherwise.} \end{cases}$$

In the case where $\tau_1^{u\ell} \geq \tau_2^{u\ell}$, the results indicate that it is advantageous for the firm to synchronize the sales of both products, resulting in $\bar{t}_1^{u\ell} = \bar{t}_2^{u\ell}$. On the other hand, when $\tau_1^{u\ell} < \tau_2^{u\ell}$, the optimal strategy depends on the value of T. If T is relatively small, it is still beneficial for the firm to synchronize the sales of both products. However, if T is large, a different approach is recommended. In this case, the firm should set $\bar{t}_1^{u\ell}$ to a value greater than the one obtained by solving a single-product problem, specifically $\frac{p_1-s}{h_1}$. Whereas for product 2, $\bar{t}_2^{u\ell}$ should be chosen to be no more than $\frac{p_2-s}{h_2}$. It's worth noting that in this scenario, the optimal solution deviates from what would be obtained by solving two separate single-product models. Consequently, regardless of the value of T, the optimal decisions for the firm differ from those obtained by solving two single-product problems independently.

Chapter 6

Two-Product Problem with Deterministic Demand and Whole-Order Fulfillment

After considering the two situations where the stockout of a product does not affect the demand or shipping of the other product in an order, we now turn to the cases where the whole order would be either lost or backlogged if any item in the order is out of stock.

6.1 The Case of Backlog

First, we consider the backlog case. In this situation, if an order requiring both products arrives but only one is available, the whole order will be backlogged until both products become available. In this case, all the backlogged demand will be fulfilled once inventory is replenished at the beginning of the next cycle. In brief, each order is shipped in a single package including all items requested by the order. Thus, we refer to this case as "whole-order backlogged".

It is obvious that the shipping cost is irrelevant to the problem solution because each order is shipped once and only once. Following the previous setting, we assume that the backlog cost is applied to each product independently, i.e., $b_{12}^{(i)} = b_i$ and $b_{12} = b_1 + b_2$. We may relax this assumption in our extension.

Let $t_i = \frac{Y_i}{\hat{\lambda}_i}$ be the firm's decisions. Let $t_{12} := t_1 + \frac{\hat{\lambda}_2}{\lambda_2}(t_2 - t_1)$ denote the time after which both products are out of stock. Figure 6.1 illustrates an example of the inventory revolution in a cycle with $t_1 < t_2$. Both products are available and all types of orders can be satisfied during $[0, t_1)$. From time t_1 to t_{12}, product 1 is out of stock, and product 2 is still in stock. Thus, all orders that demand product 1 or both products are backlogged. The orders that demand product 2 only can be satisfied immediately. After time t_{12}, both products are out of stock and hence all orders are backlogged.

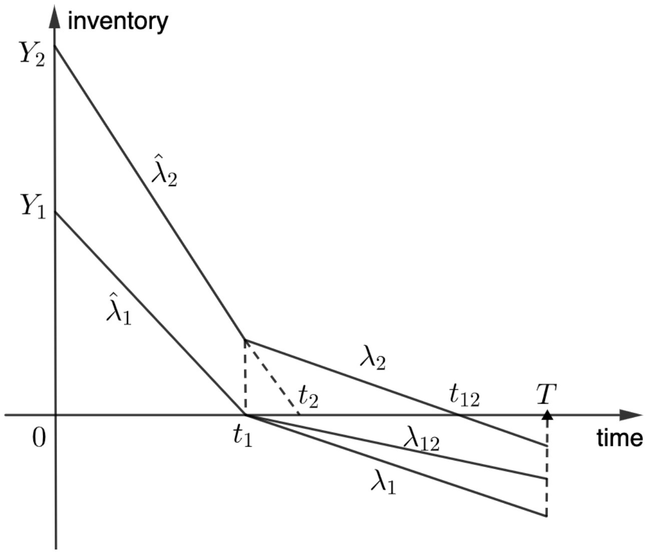


Figure 6.1: Fluid model for the two-product problem with whole-order backlogged

Let π_2^{wb} denote the firm's average cost over time, where the subscript "2" means two-product problem and the superscript "wb" means "whole-order backlogged". For $t_1 < t_2$, the average cost can be expressed by

$$\begin{aligned}\pi_2^{wb} &= \frac{1}{T}\bigg[\frac{1}{2}h_1\hat{\lambda}_1 t_1^2 + \frac{1}{2}b_1\hat{\lambda}_1(T-t_1)^2 + \frac{1}{2}b_2\lambda_{12}(T-t_1)^2 + \frac{1}{2}h_2\hat{\lambda}_2 t_2^2 \\ &\qquad + \frac{1}{2}h_2\hat{\lambda}_2(t_2-t_1)(t_{12}-t_2) + \frac{1}{2}b_2\lambda_2(T-t_{12})^2 + s(\lambda_1+\lambda_2+\lambda_{12})T + S\bigg] \\ &= \frac{1}{T}\bigg[\frac{1}{2}h_1\hat{\lambda}_1 t_1^2 + \frac{1}{2}b_1\hat{\lambda}_1(T-t_1)^2 + \frac{1}{2}b_2\lambda_{12}(T-t_1)^2 + \frac{1}{2}h_2\hat{\lambda}_2 t_2^2 \\ &\qquad + \frac{1}{2}h_2\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}(t_2-t_1)^2 + \frac{1}{2}b_2\lambda_2(T-\frac{\hat{\lambda}_2}{\lambda_2}t_2+\frac{\lambda_{12}}{\lambda_2}t_1)^2 + s(\lambda_1+\lambda_2+\lambda_{12})T + S\bigg].\end{aligned}$$

Without loss of generality, we assume that $\frac{b_1}{h_1+b_1} \le \frac{b_2}{h_2+b_2}$. Under this condition, we expect that $t_1^* \le t_2^*$, which we will prove. To obtain the solution, we take partial derivatives.

$$\begin{aligned}\frac{\partial \pi_2^{wb}}{\partial t_1} &= \frac{1}{T}\bigg\{\bigg[(h_1+b_1)\hat{\lambda}_1 + (h_2+b_2)\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}\bigg]t_1 - \bigg[b_1\hat{\lambda}_1 T + (h_2+b_2)\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}t_2\bigg]\bigg\}, \\ \frac{\partial \pi_2^{wb}}{\partial t_2} &= \frac{\hat{\lambda}_2}{T}\bigg\{(h_2+b_2)\frac{\hat{\lambda}_2}{\lambda_2}t_2 - \bigg[b_2 T + (h_2+b_2)\frac{\lambda_{12}}{\lambda_2}t_1\bigg]\bigg\}.\end{aligned}$$

Therefore, by solving the stationary point by the FOC we obtain

$$\begin{aligned}\bar{t}_1^{wb} &= \frac{b_1\hat{\lambda}_1 + b_2\lambda_{12}}{(h_1+b_1)\hat{\lambda}_1 + (h_2+b_2)\lambda_{12}}T, \\ \bar{t}_2^{wb} &= \frac{b_2 + (h_2+b_2)\frac{\lambda_{12}}{\lambda_2}\frac{\bar{t}_1^{wb}}{T}}{(h_2+b_2)\frac{\hat{\lambda}_2}{\lambda_2}}T.\end{aligned}$$

We can solve the optimal solution by the above equations. The expressions of the equations also reveal the properties of the solution. If the $\bar{t}_i^{wb}$'s on the right-hand side of the expression satisfies $\frac{b_1}{h_1+b_1} < \frac{\bar{t}_2^{wb}}{T}$ and $\frac{\bar{t}_1^{wb}}{T} < \frac{b_2}{h_2+b_2}$, then the $\bar{t}_i^{wb}$'s calculated by the equations (i.e., on the left-hand side) will satisfy $\frac{b_1}{h_1+b_1} < \frac{\bar{t}_1^{wb}}{T}$ and $\frac{\bar{t}_2^{wb}}{T} < \frac{b_2}{h_2+b_2}$, respectively. Further analysis shows that the above equations yield a unique optimal solution $(\bar{t}_1^{wb}, \bar{t}_2^{wb})$ that satisfies $\frac{b_1}{h_1+b_1} < \frac{\bar{t}_1^{wb}}{T} \leq \frac{\bar{t}_2^{wb}}{T} < \frac{b_2}{h_2+b_2}$ if $\frac{b_1}{h_1+b_1} < \frac{b_2}{h_2+b_2}$. (Note that this proof is not difficult but has not been completed.) This suggests that the firm solving a two-product problem would keep a higher inventory for product 1 and a lower inventory for product 2 compared with the solution obtained by solving separate single-product problems.

After solving the problem with exogenous T, we examine the following case where T is endogenous. Consider the optimal cost as a function of T:

$$\pi_2^{wb} = C + \frac{1}{2}T\left(\frac{(b_1\hat{\lambda}_1 + b_2\lambda_{12})(h_1\hat{\lambda}_1 + h_2\lambda_{12})}{(b_1+h_1)\hat{\lambda}_1 + (b_2+h_2)\lambda_{12}} + \lambda_2\frac{b_2h_2}{b_2+h_2}\right) + \frac{S}{T}.$$

Solving T optimally, we obtain that

$$T^{wb*} = \sqrt{\frac{2S}{\frac{(b_1\hat{\lambda}_1+b_2\lambda_{12})(h_1\hat{\lambda}_1+h_2\lambda_{12})}{(b_1+h_1)\hat{\lambda}_1+(b_2+h_2)\lambda_{12}} + \lambda_2\frac{b_2h_2}{b_2+h_2}}}$$

and $t_1^{wb*} < t_2^{wb*}$.

Proposition 3. *For a two-product fluid demand system, when the whole order is backlogged when there is an unavailable product, without loss of generality, we assume that* $\frac{b_1}{h_1+b_1} \leq \frac{b_2}{h_2+b_2}$*, we have the following results.*

1. *When reorder cycle length T is exogenous, the optimal initial inventory level is* $(\bar{Y}_1^{wb}, \bar{Y}_2^{wb}) = \left(\frac{(b_1\hat{\lambda}_1+b_2\lambda_{12})\hat{\lambda}_1}{(h_1+b_1)\hat{\lambda}_1+(h_2+b_2)\lambda_{12}}T, \frac{b_2\lambda_2+(h_2+b_2)\lambda_{12}\frac{b_1\hat{\lambda}_1+b_2\lambda_{12}}{(h_1+b_1)\hat{\lambda}_1+(h_2+b_2)\lambda_{12}}}{h_2+b_2}T\right)$.
2. *When reorder cycle length T is endogenous, the optimal reorder cycle length is* $T^{wb*} = \sqrt{\frac{2S}{\frac{(b_1\hat{\lambda}_1+b_2\lambda_{12})(h_1\hat{\lambda}_1+h_2\lambda_{12})}{(b_1+h_1)\hat{\lambda}_1+(b_2+h_2)\lambda_{12}}+\lambda_2\frac{b_2h_2}{b_2+h_2}}}$.

From this proposition, we can see that $\bar{Y}_1^{wb} \leq \frac{b_1\hat{\lambda}_1}{h_1+b_1}T$ and $\bar{Y}_2^{wb} \geq \frac{b_2\hat{\lambda}_2}{h_2+b_2}T$, where the equalities hold if and only if $\frac{b_1}{h_1+b_1} = \frac{b_2}{h_2+b_2}$. That is, compared with the single-product case, the firm would intentionally understock the product with a smaller critical ratio, and overstock the product with a higher critical ratio. Moreover, the optimal solution is independent of the shipping cost, as all the orders will finally be shipped to the customers.

6.2 The Case of Lost Sales

In this situation, if any product that a customer demands is not available, the customer will leave the firm without the purchase of any product, i.e., the whole order from this potential customer is lost. We refer to this case as "whole-order lost". Note here that the lost sales may not be observed by the firm and thus we don't make such an assumption when we consider the demand learning problem. In this setting, the effective demand rate of a product (which is the summation of the demand from customers who want this product only and the demand from customers who want both products) depends on the availability of the other product.

Let $t_i = \frac{Y_i}{\hat{\lambda}_i}$ be the firm's decisions. Let $t_{12} := t_1 + \frac{\hat{\lambda}_2}{\lambda_2}(t_2 - t_1)$ denote the time after which both products are out of stock. Figure 6.2 illustrates an example of the inventory revolution in a cycle for $t_{12} < T$. Both products are available and all types of orders can be satisfied during $[0, t_1)$. From time t_1 to t_{12}, product 1 is out of stock, and product 2 is still in stock. Therefore, all customers who need product 1, including those who demand product 1 only and those who demand both products, are lost. Customers who demand product 2 only can still be satisfied. After time t_2, both products are out of stock and hence all demand is lost.

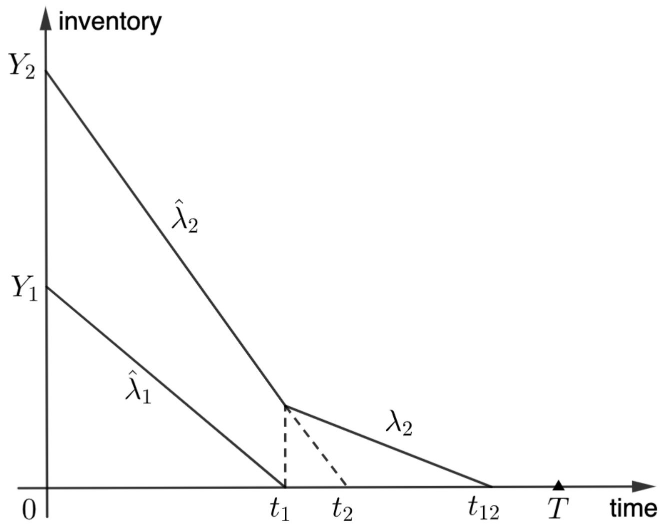


Figure 6.2: Fluid model for the two-product problem with whole-order lost

Let $\pi_2^{w\ell}$ denote the firm's average cost over time, where the subscript "2" means two-product problem and the superscript "$w\ell$" means "whole-order lost". For $t_1 < t_2 < t_{12} <$

T, the average cost can be expressed by

$$\begin{aligned}\pi_2^{w\ell} &= \frac{1}{T}\Bigg[\frac{1}{2}h_1\hat{\lambda}_1 t_1^2 + [p_1\lambda_1 + (p_1+p_2)\lambda_{12}](T-t_1) + \frac{1}{2}h_2\hat{\lambda}_2 t_2^2 + \frac{1}{2}h_2\hat{\lambda}_2(t_2-t_1)(t_{12}-t_2) \\ &\qquad + p_2\lambda_2(T-t_2) + s(\hat{\lambda}_1 t_1 + \lambda_2 t_{12}) + S\Bigg] \\ &= \frac{1}{T}\Bigg[\frac{1}{2}h_1\hat{\lambda}_1 t_1^2 + [p_1\lambda_1 + (p_1+p_2)\lambda_{12}](T-t_1) + \frac{1}{2}h_2\hat{\lambda}_2 t_2^2 + \frac{1}{2}h_2\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}(t_2-t_1)^2 \\ &\qquad + p_2\lambda_2(T - \frac{\hat{\lambda}_2}{\lambda_2}t_2 + \frac{\lambda_{12}}{\lambda_2}t_1) + s(\lambda_1 t_1 + \hat{\lambda}_2 t_2) + S\Bigg].\end{aligned}$$

Without loss of generality, we assume that $\frac{p_1-s}{h_1} \le \frac{p_2-s}{h_2}$. Under this condition, we expect that $\bar{t}_1^{w\ell} \le \bar{t}_2^{w\ell}$, which we will prove. To obtain the solution, we take partial derivatives.

$$\begin{aligned}\frac{\partial \pi_2^{w\ell}}{\partial t_1} &= \frac{1}{T}\left[\left((h_1\hat{\lambda}_1 + h_2\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}\right)t_1 - \left((p_1-s)\hat{\lambda}_1 + s\lambda_{12} + h_2\hat{\lambda}_2\frac{\lambda_{12}}{\lambda_2}t_2\right)\right], \\ \frac{\partial \pi_2^{w\ell}}{\partial t_2} &= \frac{\hat{\lambda}_2}{T}\left[h_2\frac{\hat{\lambda}_2}{\lambda_2}t_2 - \left((p_2-s) + h_2\frac{\lambda_{12}}{\lambda_2}t_1\right)\right].\end{aligned}$$

Therefore, solving the FOC we obtain the following stationary point

$$\begin{aligned}\hat{t}_1 &= \frac{(p_1-s)\hat{\lambda}_1 + p_2\lambda_{12}}{h_1\hat{\lambda}_1 + h_2\lambda_{12}}, \\ \hat{t}_2 &= \frac{(p_2-s) + h_2\frac{\lambda_{12}}{\lambda_2}\hat{t}_1}{h_2\frac{\hat{\lambda}_2}{\lambda_2}} = \hat{t}_1 + \frac{p_2 - s - h_2\hat{t}_1}{h_2\frac{\hat{\lambda}_2}{\lambda_2}}.\end{aligned}$$

We can solve the stationary point by the above equations. Then, the optimal decision when T is sufficiently large is given by $(\bar{t}_1^{w\ell}, \bar{t}_2^{w\ell}) = (\hat{t}_1, \hat{t}_2)$. The expressions of the equations also reveal the properties of the solution. If the $\hat{t}_i$'s on the right-hand side of the expression satisfy $\frac{p_1-s}{h_1} < \hat{t}_2$ and $\hat{t}_1 < \frac{p_2-s}{h_2}$, then the $\hat{t}_i$'s calculated by the equations (i.e., on the left-hand side) will satisfy $\frac{p_1-s}{h_1} < \hat{t}_1$ and $\hat{t}_2 < \frac{p_2-s}{h_2}$, respectively. Further analysis shows that the above equations yield a unique optimal solution $(\hat{t}_1, \hat{t}_2)$ that satisfies $\frac{p_1-s}{h_1} < \hat{t}_1 \le \hat{t}_2 < \frac{p_2-s}{h_2}$ if $\frac{p_1-s}{h_1} < \frac{p_2-s}{h_2}$. This suggests that the firm solving a two-product problem would keep a higher inventory for product 1 and a lower inventory for product 2 compared with the solution obtained by solving separate single-product problems.

Note that the above solution is only for a sufficiently large T and is not complete. Let $\hat{t}_{12} := \hat{t}_1 + \frac{\hat{\lambda}_2}{\lambda_2}(\hat{t}_2 - \hat{t}_1)$. In the following, we discuss the solution for a general T.

- If $T \ge \hat{t}_{12} = \frac{p_2-s}{h_2}$, then $(\bar{t}_1^{w\ell}, \bar{t}_2^{w\ell}) = (\hat{t}_1, \hat{t}_2)$. With endogenous T, the cost function is $p_1\hat{\lambda}_1 + p_2\hat{\lambda}_2 + [S - \frac{[(p_1-s)\hat{\lambda}_1 + p_2\lambda_{12}]^2}{2(h_1\hat{\lambda}_1 + h_2\lambda_{12})} - \frac{\lambda_2(p_2-s)^2}{2h_2}]\frac{1}{T}$, which is monotone in T.
- If $\frac{(p_1-s)\hat{\lambda}_1 + p_2\lambda_{12}}{h_1\hat{\lambda}_1 + h_2\lambda_{12}} = \hat{t}_1 \le T < \hat{t}_{12} = \frac{p_2-s}{h_2}$, then $\bar{t}_1^{w\ell} = \hat{t}_1 = \frac{(p_1-s)\hat{\lambda}_1 + p_2\lambda_{12}}{h_1\hat{\lambda}_1 + h_2\lambda_{12}}$ and $\bar{t}_2^{w\ell} = \frac{\lambda_2 T + \lambda_{12}\bar{t}_1^{w\ell}}{\hat{\lambda}_2}$ such that $\bar{t}_{12}^{w\ell} = T$. With endogenous T, the cost function is $C' + \frac{1}{2}h_2\lambda_2 T +$

$\frac{S-\frac{1}{2}\frac{[(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}]^2}{h_1\hat{\lambda}_1+h_2\lambda_{12}}}{T}$, where C' is a constant. Hence, $T^{w\ell*}=\sqrt{\frac{2S-\frac{[(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}]^2}{h_1\hat{\lambda}_1+h_2\lambda_{12}}}{h_2\lambda_2}}$.

- If $T\le\hat{t}_1=\frac{(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}}{h_1\hat{\lambda}_1+h_2\lambda_{12}}$, then $\bar{t}_1^{w\ell}=\bar{t}_2^{w\ell}=T$. With endogenous T, the cost function is $C''+\frac{1}{2}(h_1\hat{\lambda}_1+h_2\hat{\lambda}_2)T+\frac{S}{T}$, where C'' is a constant. Hence, $T^{w\ell*}=\sqrt{\frac{2S}{h_1\hat{\lambda}_1+h_2\hat{\lambda}_2}}$.

By letting $\mu^{w\ell}(S)=\sqrt{\frac{2S-\frac{[(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}]^2}{h_1\hat{\lambda}_1+h_2\lambda_{12}}}{h_2\lambda_2}}$, $\mu_1^{w\ell}=\frac{(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}}{h_1\hat{\lambda}_1+h_2\lambda_{12}}$, and $\mu_2^{w\ell}=\frac{p_2-s}{h_2}$, we have discussions as follows.

- If $\mu^{w\ell}(S)\le\mu_1^{w\ell}$, then $T^{w\ell*}=\sqrt{\frac{2S}{h_1\hat{\lambda}_1+h_2\hat{\lambda}_2}}, t_1^{w\ell*}=t_2^{w\ell*}=T^{w\ell*}$.
- If $\mu^{w\ell}(S)\in(\mu_1^{w\ell},\mu_2^{w\ell}]$, then $T^{w\ell*}=\mu^{w\ell}(S), t_1^{w\ell*}=\hat{t}_1<t_2^{w\ell*}=T^{w\ell*}$.
- If $\mu^{w\ell}(S)>\mu_2^{w\ell}$, then $T^{w\ell*}=\infty$.

To summarize, we conclude the aforementioned results in Proposition 4.

Proposition 4. *For a two-product fluid demand system, when the whole order is lost when there is an unavailable product, without loss of generality, we assume that* $\frac{p_1-s}{h_1}\le\frac{p_2-s}{h_2}$*, we have the following results.*

1. *When reorder cycle length T is exogenous, the optimal initial inventory level is*

$$(\bar{Y}_1^{w\ell},\bar{Y}_2^{w\ell})=\begin{cases}(\frac{[(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}]\hat{\lambda}_1}{h_1\hat{\lambda}_1+h_2\lambda_{12}},\frac{[(p_2-s)+h_2\frac{\lambda_{12}}{\lambda_2}\frac{(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}}{h_1\hat{\lambda}_1+h_2\lambda_{12}}]\hat{\lambda}_2}{h_2\frac{\hat{\lambda}_2}{\lambda_2}}) & \text{if } T\ge\mu_2^{w\ell},\\ (\frac{[(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}]\hat{\lambda}_1}{h_1\hat{\lambda}_1+h_2\lambda_{12}},\lambda_2T+\lambda_{12}\frac{(p_1-s)\hat{\lambda}_1+p_2\lambda_{12}}{h_1\hat{\lambda}_1+h_2\lambda_{12}}) & \text{if } \mu_1^{w\ell}\le T<\mu_2^{w\ell},\\ (\hat{\lambda}_1T,\hat{\lambda}_2T) & \text{if } T\le\mu_1^{w\ell}.\end{cases}$$

2. *When reorder cycle length T is endogenous, the optimal reorder cycle length is*

$$T^{w\ell*}=\begin{cases}\sqrt{\frac{2S}{h_1\hat{\lambda}_1+h_2\hat{\lambda}_2}} & \text{if } \mu^{w\ell}(S)\le\mu_1^{w\ell},\\ \mu^{w\ell}(S) & \text{if } \mu_1^{w\ell}<\mu^{w\ell}(S)\le\mu_2^{w\ell},\\ \infty & \text{if } \mu^{w\ell}(S)>\mu_2^{w\ell}.\end{cases}$$

The proposition indicates that, when T is small enough ($T\le\mu_1^{w\ell}$), the firm should synchronize the sale of both products and be able to fulfill all the demand; otherwise, the firm should overstock product 1 and understock product 2.

Chapter 7

Conclusions

Compared to the stochastic model, all deterministic models offer closed-form solutions. We approach the deterministic problems by initially treating the reorder cycle lengths as exogenous and subsequently as endogenous.

In the single-product problem, the solution for the case of backlog corresponds to the classic EOQ model and is independent of shipping costs. In the case of lost sales, however, the optimal decision is influenced by shipping costs when the reorder cycle length is exogenous. Once the reorder cycle length is endogenized, shipping costs no longer affect the optimal solutions, except for differentiating between various ranges of setup costs.

For the two-product problem, the solution diverges as we treat the sales of the two products separately. The complexities introduced by fulfillment processes give rise to solutions' contingency on shipping costs in these two-product scenarios, highlighting the importance of considering this cost component in inventory decision-making.

These findings underscore the critical role that fulfillment dynamics and shipping costs play in effective inventory planning. Our study paves the avenue for further exploration of adaptive strategies that incorporate real-time demand data within the whole-order fulfillment policy.

By incorporating the fulfillment dynamics into both stochastic and deterministic models, we provide a more comprehensive framework for decision-making in multi-product inventory control. These methods encourage e-commerce companies to holistically consider back-end inventory decisions and front-end order fulfillment, revealing significant potential for cost reduction in practice.

Part III

Stochastic-Demand Models under Partial Fulfillment

Chapter 8

Partial Fulfillment with Known Demand Information

This chapter addresses the problem under the assumption that the firm has complete knowledge of the demand distribution. Since there is no lead time, solving the multi-period problem is equivalent to tackling a single-period problem, provided there is an appropriate initial inventory (e.g., zero). Below, we will solve the problem separately for the cases of backlogged and lost sales.

In what follows, we make the following assumption to guarantee the viability of the business under the partial fulfillment and instant shipping policy.

Assumption 1. *We assume that* $s \leq b_i + h_i$ *in the case of backlog and* $s \leq p_i + h_i$ *in the case of lost sales, where* $i = 1, 2$.

The assumption requires the shipping cost to be lower than the total inventory holding and shortage cost. It is observed in practice that the logistics cost has declined considerably as e-commerce develops and many firms (e.g., Amazon and JD) choose premium logistics services as their one-order winner. For example, Amazon and JD have streamlined their operations and reduced logistics costs to such an extent that they can use a partial fulfillment policy by default for customers in major cities. According to F. Curtis Barry and Company, backorders can incur high costs ranging from \$15 to \$20 per fulfillment. In contrast, the shipping cost for standard-sized goods on Amazon generally does not exceed \$15. This comparison suggests that Assumption 1 naturally holds for a wide range of products.

Under this assumption, the use of a partial fulfillment policy by the firm is justified and the optimal inventory decision is non-trivial (e.g., not zero).

Assumption 1 will play a crucial role in proving the convex-extensibility of the conditional cost function in both the cases of backlogged and lost sales.

8.1 The Case of Backlog

We start with the case where the demand for unavailable items is backlogged. Let $\pi^b(\vec{Y})$ be the expected single-period cost when the firm sets a beginning inventory $\vec{Y}$, which is formulated as

$$\pi^b(\vec{Y}) := \mathbb{E}\left\{\sum_{i=1}^{2}\left[h_i(Y_i - \hat{D}_i)^+ + b_i(\hat{D}_i - Y_i)^+\right] + s\text{NS}(\vec{Y}, \mathcal{D}^{Seq})\right\},$$

where $\text{NS}(\vec{Y}, \mathcal{D}^{Seq})$ is the number of shipments needed for the orders that arrive in the period, including those shipped immediately after order placement and those shipped later after being backlogged.

To prove that the optimal policy is of a base-stock type, it is sufficient to prove the unimodality or convexity of the cost function. It is obvious that the first two components, i.e., the inventory holding and backlog costs, are convex in $\vec{Y}$. The challenge lies in the derivation and analysis of the last component, the number of shipments needed. Under the partial shipment policy, it depends not only on the amount of orders received but also on the sequence they arrive. By a complete examination of all possible demand realizations, we, unfortunately, found that $\text{NS}(\vec{Y}, \mathcal{D}^{seq})$ is not convex in all sample paths (a counter-example can be found in Example 1). Therefore, we need to look into its expectation, which is more complex than examining sample paths separately.

Example 1 (A sample path on which the cost is not convex-extensible)**.** *Suppose the realized demand sequence is* $\mathcal{D}^{seq} = \{1, 12, 2, 1, 12, 1\}$. *For different initial inventory levels, we have:*

Initial inventory level	*(3, 1)*	*(4, 1)*	*(5, 1)*
Backorder cost	$2b_1 + 2b_2$	$b_1 + 2b_2$	$2b_2$
Holding cost	*0*	*0*	*0*
Shipping cost	$6s$	$7s$	$7s$

Then we can observe that:

$$\left[\pi^b(3,1|\mathcal{D}^{seq}) - \pi^b(4,1|\mathcal{D}^{seq})\right] - \left[\pi^b(4,1|\mathcal{D}^{seq}) - \pi^b(5,1|\mathcal{D}^{seq})\right] = -s < 0$$

This contradicts the definition of convexity.

Let $\overline{\text{NS}}(\vec{Y}|\vec{D})$ denote the expected number of shipments conditional on the number of orders received, $\vec{D}$, i.e.,

$$\overline{\text{NS}}(\vec{Y}|\vec{D}) := \mathbb{E}_{Seq}\left[\text{NS}(\vec{Y}, \mathcal{D}^{Seq})\right],$$

where the expectation, $\mathbb{E}_{Seq}$, is taken over all possible arrival sequences. To derive the properties of this conditional expected number of shipments, we first introduce a lemma that is crucial in our analysis.

Under our modeling assumptions, conditional on a realization of (D_1, D_2, D_{12}) in a period, if we treat orders as differentiable (as is in practice where each order is assigned a unique order ID), any permutation of the arrival sequence of these $D_1 + D_2 + D_{12}$ orders happens with an equal probability $1/(D_1 + D_2 + D_{12})!$. Given this property, we prove the following lemma.

Lemma 1. *Conditional on a realization of $\vec{D} = (d_1, d_2, d_{12})^T$ in a period, we have the following results.*

(i) Knowing that a specific type-12 order arrives in this period, the probability that it arrives at the kth place is equal to $\frac{1}{d_1+d_2+d_{12}}$.

(ii) Knowing that a specific type-12 order arrives in this period, the probability that it is the kth that requests product i is equal to $\frac{1}{d_i+d_{12}}$.

(iii) The probability that the kth order that arrives is of type-r is equal to $\frac{d_r}{d_1+d_2+d_{12}}$, $r = 1, 2, 12$.

(iv) For any sequence $d^{Seq} = \{r_1, r_2, \cdots, r_{d_1+d_2+d_{12}}\}$ where

$$\sum_{k=1}^{d_1+d_2+d_{12}} \mathbb{I}\{r_k = r\} = d_r \text{ for } r \in \{1, 2, 12\},$$

the probability that $\mathcal{D}^{Seq} = d^{Seq}$ is equal to $\frac{d_1!d_2!d_{12}!}{(d_1+d_2+d_{12})!}$.

Given a beginning inventory position $\vec{Y}$ and a demand realization $\mathcal{D}^{seq}$ of a period, we use $\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq})$ and $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq})$ to denote the index (in the arrival sequence in this period) and type of the order that depletes the inventory of product i (i.e., product i runs out of stock for the first time after this order arrives). Specifically, let

$$\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) := \min_k \left\{ Y_i - \sum_{k'=1}^{k} \Big(\mathbb{I}\{r_{k'} = i\} + \mathbb{I}\{r_{k'} = 12\} \Big) \leq 0 \right\},$$

$$\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) := r_{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq})}.$$

Furthermore, let $\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty$ and $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 0$ if product i never goes out of stock in the period and $\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = 0$ and $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 0$ if $Y_i \leq 0$.

By Lemma 1 and the above notation, we are able to rewrite $\overline{\text{NS}}(\vec{Y}|\vec{D})$ as follows.

$$\begin{aligned}\overline{\text{NS}}(\vec{Y}|\vec{D}) =& \mathbb{I}\left\{\hat{D}_1 < Y_1, \hat{D}_2 < Y_2\right\}(D_1 + D_2 + D_{12}) \\ &+ \mathbb{I}\left\{\hat{D}_1 \geq Y_1, \hat{D}_2 < Y_2\right\}\left[D_1 + D_2 + D_{12} + \left(1 - \frac{Y_1}{\hat{D}_1}\right)D_{12}\right] \\ &+ \mathbb{I}\left\{\hat{D}_1 < Y_1, \hat{D}_2 \geq Y_2\right\}\left[D_1 + D_2 + D_{12} + \left(1 - \frac{Y_2}{\hat{D}_2}\right)D_{12}\right] \\ &+ \mathbb{I}\left\{\hat{D}_1 \geq Y_1, \hat{D}_2 \geq Y_2\right\}\left[D_1 + D_2 + D_{12} + \mathbb{E}_{Seq}\left(\sum_{k=\mathcal{M}_1\wedge\mathcal{M}_2+1}^{\mathcal{M}_1\vee\mathcal{M}_2} \mathbb{I}\left\{r_k = 12\right\}\right)\right]\end{aligned}$$

When neither product stocks out in a period, all orders can be fulfilled immediately via one shipment, implying $\overline{\text{NS}}(\vec{Y}|\vec{D}) = D_1 + D_2 + D_{12}$. When one product (e.g., product-1) goes out of stock and another (e.g., product-2) does not during a period, some type-12 orders may be satisfied first, partially, by on-hand inventory and then by future replenishment, requiring two shipments each. By Lemma 1, a specific type-12 order arrives as the kth order that requests product-2 with a probability $\frac{1}{D_2+D_{12}}$, implying a total probability $1 - \frac{Y_1}{\hat{D}_1}$ that it requires two separate shipments. Therefore, $\overline{\text{NS}}(\vec{Y}|\vec{D}) = D_1 + D_2 + D_{12} + \left(1 - \frac{Y_1}{\hat{D}_1}\right) D_{12}$ in this case. For all the above three cases, the shipment number is convex in $\vec{Y}$.

The most complicated case is the one when both products go out of stock during the period. In this case, the number of shipments depends on when the two stockout events occur, which are not independent. Here, we tackle the problem by grouping non-convex sample paths with strictly convex ones and hope to show that their summation is convex. After a complicated analysis involving grouping and rearranging events, and taking conditional expectation multiple times, we prove that the expected number of shipments is $L^\natural$-convex in $\vec{Y}$ in the region $\hat{D}_1 \geq Y_1, \hat{D}_2 \geq Y_2$ and the expected cost function is convex-extensible in the whole feasible region U.

First, we show the convexity of the conditionally expected number of shipments $\overline{\text{NS}}(\vec{Y}|\vec{D})$.

Proposition 5. *Conditional on a realization of $\vec{D}_t := (D_1, D_2, D_{12})$ in a period, the expected number of shipments $\overline{NS}(\vec{Y}|\vec{D})$ is $L^\natural$-convex in $\vec{Y}$ in the region $\{\vec{Y} : \hat{D}_1 \geq Y_1, \hat{D}_2 \geq Y_2\} \cap U$.*

Under Assumption 1, the conditionally expected cost function is convex-extensible and submodular in $\vec{Y}$.

Next, we extend the result to the total cost function.

Theorem 1. *Under Assumption 1, the expected cost function $\pi^b(\vec{Y})$ is convex-extensible and submodular in $\vec{Y} \in U$.*

This theorem follows directly from the proposition by recognizing that convex-extensibility and submodularity are preserved when taking expectations.

By the theorem, the cost function $\pi^b(\vec{Y})$ is unimodal and any local minimum is also a global minimum. Computationally, the minimization problem $\min_{\vec{Y}}\{\pi^b(\vec{Y})\}$ can be solved efficiently by a subgradient method. Let $\vec{Y}^*$ be the optimal solution to the problem if its solution is unique. If there are multiple solutions, these solutions are contained in a convex set where all integer points inside the set are minimal. In this case, we define $\vec{Y}^*$ to be the one that is selected according to the following preference: the one with the smallest Y_1 and then the one with the smallest Y_2. Then, it is straightforward to conclude the optimal policy for the firm by the above theorem.

The Optimal Policy with Known Demand Information. A base-stock policy with order-up-to level $\vec{Y}^*$ is optimal for minimizing the firm's expected cost in a multi-period problem.

8.1.1 Effect of Shipping Cost

In the following, we reveal the effect of the shipping cost on the optimal inventory decisions by comparing the inventory policies derived with and without considering shipping cost, assuming the shipping cost is zero in the latter case.

To study the impact of the inclusion of the shipping cost on the optimal inventory policy, we now analyze how the expected cost function changes with regard to the parameter s. To this end, we include s as a variable and write the expected cost function as $\pi^b(\vec{Y}, s)$ in this subsection. Furthermore, because we have proven that the expected cost function is convex-extensible, we select a smooth and convex extension of expected total holding and backorder cost to the continuous domain and denote it as $\bar{\pi}^b(\vec{Y}, 0)$. We choose a smooth extension of $\overline{\mathrm{NS}}(\vec{Y})$, denoted as $\bar{\overline{\mathrm{NS}}}(\vec{Y})$, such that $\bar{\pi}^b(\vec{Y}, s) := \bar{\pi}^b(\vec{Y}, 0) + s\bar{\overline{\mathrm{NS}}}(\vec{Y})$ is convex.

A *myopic* firm that disregards the impact of inventory on the fulfillment process and cost when making inventory decisions solves a problem $\min_{\vec{Y}}\{\bar{\pi}^b(\vec{Y}, 0)\}$, where $\bar{\pi}^b(\vec{Y}, 0)$ is the sum of the inventory holding and backlog costs. Its optimal decision $\vec{Y}^o := (Y_1^o, Y_2^o)$ satisfies the first-order condition

$$\frac{\partial \bar{\pi}^b(\vec{Y}^o, 0)}{\partial Y_i} = (b_i + h_i)F_{\hat{D}_i}(Y_i^o) - b_i = 0. \tag{8.1}$$

A *strategic* firm that accounts for fulfillment dynamics in inventory decision-making optimizes $\bar{\pi}^b(\vec{Y}, s) := \bar{\pi}^b(\vec{Y}, 0) + s\bar{\overline{\mathrm{NS}}}(\vec{Y})$, which includes an additional shipping-cost term,

$s\overline{\overline{\mathrm{NS}}}(\vec{Y})$, in its objective. Its optimal decision $\vec{Y}^*$ satisfies the first-order condition

$$\frac{\partial \bar{\pi}^b(\vec{Y}^*, s)}{\partial Y_i} = (b_i + h_i)F_{\hat{D}_i}(Y_i^*) - b_i + s\left.\frac{\partial \overline{\overline{\mathrm{NS}}}(\vec{Y})}{\partial Y_i}\right|_{\vec{Y}=\vec{Y}^*} = 0. \tag{8.2}$$

By comparing Equations (8.2) and (8.1) and analyzing based on the convexity property, we find that the myopic firm overstocks product i, i.e., $Y_i^* < Y_i^o$, if $\left.\frac{\partial \overline{\overline{\mathrm{NS}}}(\vec{Y})}{\partial Y_i}\right|_{\vec{Y}^*} > 0$, and understocks product i, i.e., $Y_i^* > Y_i^o$ if $\left.\frac{\partial \overline{\overline{\mathrm{NS}}}(\vec{Y})}{\partial Y_i}\right|_{\vec{Y}^*} < 0$ for $i = 1, 2$.

The sign of $\left.\frac{\partial \overline{\overline{\mathrm{NS}}}(\vec{Y})}{\partial Y_i}\right|_{\vec{Y}^o}$ depends on the value of the myopic decision $\vec{Y}^o$ relative to the shipping cost function, which is affected by the demand process. To illustrate how the demand process (specifically demand correlation) influences the optimal decision, we plot the contours of the functions $\bar{\pi}^b(\vec{Y}, 0)$, $s\overline{\overline{\mathrm{NS}}}(\vec{Y})$, $\bar{\pi}^b(\vec{Y}, s)$ in Figure 8.1 for scenarios of different demand correlations. The three pictures in the left column of the figures are identical and show the total inventory holding and backlog cost, with its minimum point marked by a yellow diamond. This is the optimal inventory level, $\vec{Y}^o$, set by a myopic firm. We mark the optimal decision of the strategic firm by a green triangle in the three pictures of the right column. The three pictures in the right column are for the total cost including shipping considered the strategic firm, with its minimum point $\vec{Y}^*$, marked by a green triangle. The three pictures in the middle column show the shipping cost for scenarios where the demand for the two products is substitutable (each order requests one and only one product), or independent (each order may include a product with a certain probability independent of whether it includes another product), or complementary (e.g., every order that requests product 1 also demands product 2). As one can see, the shipping cost in the scenario of substitutable demand is independent of the inventory level set by the firm (i.e., showing a flat contour) because every order requires one and only one shipment. With the possibility of multiple products being requested by a single order (in scenarios of independent or complementary demand), the shipping cost exhibits a saddle pattern with peaks at the upper-left and lower-right, saddle around the middle, and minimums at the lower-left and upper-right. The lower-left to upper-right diagonal (a relatively flat trail) represents the region of decisions Y_1 and Y_2 such that the chance of aligning the depletion of the inventory of both products is high, with a slope of approximately $\frac{\hat{D}_2}{\hat{D}_1}$. The lower-left point represents the case where all orders are backlogged so that only one shipment is needed for each when the inventory is replenished. The upper-right point represents the case where inventory of both products is high enough, so there is always inventory when an order arrives and backlog rarely happens. Therefore, the minimum point of the inventory holding and backlog cost and that of the shipping cost are different whenever there is a chance an order demands both products. By also marking the myopic optimal $\vec{Y}^o$ in the contour map of the shipping cost, we can see that the value/sign of $\frac{\partial \bar{\overline{\text{NS}}}(\vec{Y})}{\partial Y_i}\Big|_{\vec{Y}^o}$ depends on where $\vec{Y}^o$ is located. If the firm accounts for the fulfillment process and shipping cost in its decision-making process, the optimal decision should move (from the point of the yellow diamond) toward the dark blue area (to the point of the green triangle). As one can see, the ridge in the contour map is steeper and the decision change (the difference between the yellow diamond and the green triangle) is larger as the demand correlation gets stronger (by comparing sub-figures (e) and (h)). Roughly speaking, the larger the myopic optimal $\vec{Y}^o$ deviates away from the diagonal flat trail (in dark blue color in the shipping cost contour map), the greater the difference between $\vec{Y}^o$ and $\vec{Y}^*$ and the benefit of accounting for shipping process in inventory planning.

8.2 The Case of Lost Sales

In the case where the demand for unavailable items is lost, the expected single-period cost $\pi^\ell(\vec{Y})$ when the firm sets a beginning inventory $\vec{Y}$ is formulated as

$$\pi^\ell(\vec{Y}) := \mathbb{E}\left\{\sum_{i=1}^{2}\left[h_i(Y_i - \hat{D}_i)^+ + p_i(\hat{D}_i - Y_i)^+\right] + s\text{NS}(\vec{Y}, \mathcal{D}^{Seq})\right\},$$

where $\text{NS}(\vec{Y}, \mathcal{D}^{Seq})$ is the number of shipments needed for the orders that arrive in the period, including those that are fully available and those that are partially available.

To derive the base-stock policy as the optimal policy, the challenge also lies in the derivation and analysis of the number of shipments needed. We still use $\overline{\text{NS}}(\vec{Y}|\vec{D})$ to denote the expected number of shipments conditional on the number of orders received, $\vec{D}$, i.e.,

$$\overline{\text{NS}}(\vec{Y}|\vec{D}) := \mathbb{E}_{Seq}\left[\text{NS}(\vec{Y}, \mathcal{D}^{Seq})\right],$$

By using our modeling assumptions and Lemma 1, taking advantage of the notations of $\mathcal{M}_i(\vec{Y}, \mathcal{D}^{Seq})$ and $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{Seq})$, we can rewrite $\overline{\text{NS}}(\vec{Y}|\vec{D})$ as

$$\begin{aligned}\overline{\text{NS}}(\vec{Y}|\vec{D}) =& \mathbb{I}\left\{\hat{D}_1 < Y_1, \hat{D}_2 < Y_2\right\}(D_1 + D_2 + D_{12}) \\ &+ \mathbb{I}\left\{\hat{D}_1 \geq Y_1, \hat{D}_2 < Y_2\right\}\left[\frac{D_1}{\hat{D}_1}Y_1 + D_2 + D_{12}\right] \\ &+ \mathbb{I}\left\{\hat{D}_1 < Y_1, \hat{D}_2 \geq Y_2\right\}\left[D_1 + \frac{D_2}{\hat{D}_2}Y_2 + D_{12}\right] \\ &+ \mathbb{I}\left\{\hat{D}_1 \geq Y_1, \hat{D}_2 \geq Y_2\right\}\left[Y_1 + Y_2 - \mathbb{E}_{Seq}\left(\sum_{k=1}^{\mathcal{M}_1 \wedge \mathcal{M}_2} \mathbb{I}\{r_k = 12\}\right)\right]\end{aligned}$$

When neither product stocks out in a period, all orders can be fulfilled immediately via one shipment, implying $\overline{\text{NS}}(\vec{Y}|\vec{D}) = D_1 + D_2 + D_{12}$. When one product (e.g., product-

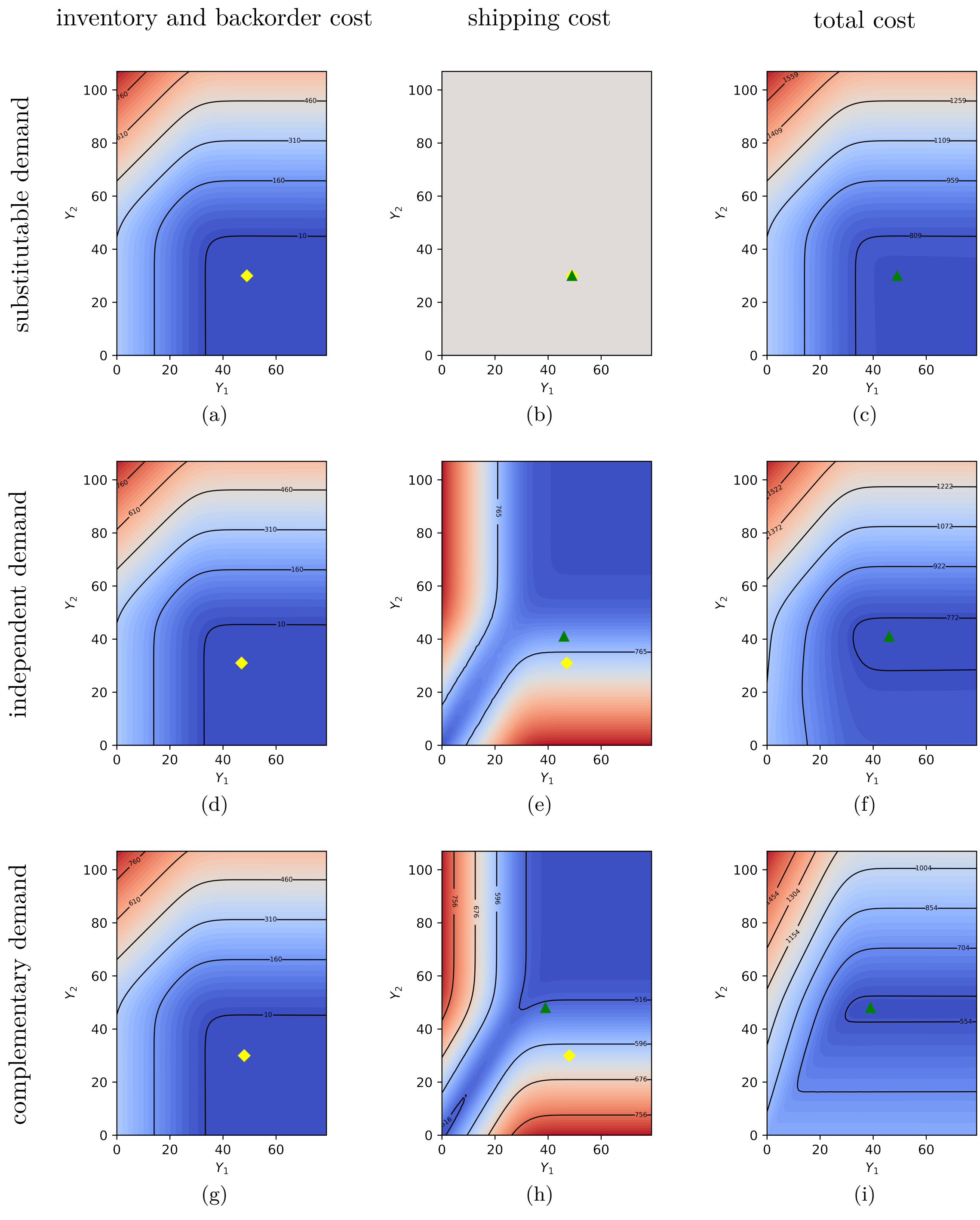


Figure 8.1: Contour map of costs with different demand correlation

1) goes out of stock and another (e.g., product-2) does not during a period, all the orders requiring product-2 will be shipped immediately, leading to $D_2 + D_{12}$ shipments, although some type-12 order may be fulfilled partially. By Lemma 1, a specific type-12 order arrives as the kth order that requests product-1 with a probability $\frac{1}{D_1+D_{12}}$, implying a total probability $\frac{D_1}{\hat{D}_1}$ that an order requesting product-1 is of type-1. Therefore, $\overline{\mathrm{NS}}(\vec{Y}|\vec{D}) = \frac{D_1}{\hat{D}_1}Y_1 + D_2 + D_{12}$. For all the above three cases, the number of shipments is convex in $\vec{Y}$.

When both products go out of stock during the period, the number of shipments depends on when the two stockout events occur. We use a similar technique as in the case of backlog to deal with the expected number of shipments and finally prove that $\overline{\mathrm{NS}}(\vec{Y}|\vec{D})$ is also $L^\natural$-convex in $\vec{Y}$ in the region $\hat{D}_1 \geq Y_1, \hat{D}_2 \geq Y_2$ and the expected cost function is convex-extensible in the whole feasible region U.

Hence, we obtain the same results as in Proposition 5 and Theorem 1.

Given the convexity results, the optimal policy with known demand information can be derived accordingly.

Chapter 9

Partial Fulfillment with Learning

We then consider the problem where the firm does not know the demand distribution a priori and can only observe the realized (or censored) demand over time. In this setting, the firm needs to incorporate the newly acquired information into the current decision-making process so as to learn demand information in an online manner. However, if the firm tends to explore the "true" demand and set the order-up-to levels high at the first stages, it may risk losing profit. On the other hand, if the firm aims to gain more profit, it may lose the opportunity of knowing the true demand distribution and hence risk losing profit in the long run. This is the so-called *exploration-exploitation tradeoff*, requiring a balance between learning the demand patterns and making optimal inventory decisions. Our problem differs from existing inventory learning problems in that we incorporate a discontinuous and dynamic fulfillment process, and the problem is situated in a high-dimensional space. These factors complicate the process of revealing unknown demand information from observed demand and necessitate new methods for developing the algorithm.

As demonstrated in Chapter 8, the optimal policy for our problem with known demand information is of a base-stock type. Therefore, our objective is to develop an online learning algorithm that adjusts the base-stock level in each period as more demand and sales data are collected, ultimately identifying the optimal base-stock level over time.

Having established the convex extensibility of the cost function, can our problem be addressed simply by existing online convex optimization methods? Note that the traditional online convex optimization approaches are designed for convex objective functions where the decision point x_t is chosen from a continuous, compact set. However, our decisions take discrete integer values, which poses the first issue. Moreover, the learning process depends not only on the demand quantities $\vec{D}$ but also on the demand correlation and fulfillment processes Seq. How do we leverage this complex information in our learning? This constitutes the second issue.

To address the first issue, we need to extend the discrete cost function to a continuous domain. How should we define a continuous extension that consistently preserves convexity when the demand information is unknown, and consequently, the shape of the discrete cost function is also unknown? Does this imply that we also need to learn the shape for the continuation operation? To address these issues, we devise a piecewise linear continuation. By utilizing the structural properties we developed in Section 8, we can determine the shape of the continuation. However, after extending the function, another problem arises: the decision point in the online convex optimization algorithm is a continuous variable. How do we obtain an implementable integer-valued decision? This requires us to separate the decision *update* and *implementation* processes, maintaining a sequence of continuous variables for the former and a sequence of integer variables for the latter. While this adaptation is intuitive, it presents several challenges that need to be resolved: We must ensure that (1) the outcome from the implemented decision provides sufficient information for the next update at the continuous point, and (2) the regret is minimized, meaning that the implemented decision aligns with the optimal decision in the long term.

To address the second issue, we make the following observations: The online subgradient descent (OSD) method can described as follows. Suppose a decision maker needs to find the minimizer of an unknown objective function $f_e(x) := \mathbb{E}[f(x, \epsilon)]$ based on observations of the function value $f(x_t, \epsilon_t)$ and the (sub)gradient $\nabla_{x_t} f(x_t, \epsilon_t)$ in each period. If the observed subgradient $\nabla_{x_t} f(x_t, \epsilon_t)$ is an unbiased estimator of $\nabla_{x_t} f_e(x_t)$, the OSD method suggests updating the decision x_t using the following formula:

$$x_{t+1} = x_t - \alpha_t \nabla_{x_t} f(x_t, \epsilon_t),$$

which means we update the decisions iteratively in the opposite direction of the objective function's (sub)derivative, using a step-size factor α_t at each step. Although the observed $\nabla_{x_t} f(x_t, \epsilon_t)$ may not align with $\nabla_{x_t} f_e(x_t)$, which points in the direction that minimizes the expected objective function, it provides an indication of the direction in expectation if it is an unbiased estimator. According to the law of large numbers, this update mechanism increases the likelihood of identifying the correct adjustment direction as the number of iterations grows. Therefore, it is essential that the observed subgradient be an unbiased estimator of the true subgradient at the decision point, allowing the update mechanism to guide the decision sequence closer to the optimal solution (with unsummable diminishing step sizes) with a high probability. What is the form of the unbiased estimator of the subgradient?

In next subsection, we elucidate how we address the aforementioned issues and develop our algorithm. In this part of our study where demand information needs to be learned, we assume that the total demand is positive and bounded.

Assumption 2 (Bounded Demand)**.** *We assume known demand upper and lower bounds* *as follows.*

1. *(Upper bound) There exists a known upper bound β_i for the total demand of product i in each period, i.e., $\hat{D}_{t,i} \leq \beta_i$ for $i = 1, 2$.*

2. *(Lower bound) The total demand for product i in each period is positive, i.e., $\hat{D}_{t,i} \geq 1$ for $i = 1, 2$.*

The condition is mild and commonly holds in practice as demand is typically finite in real-world businesses. Moreover, it is assumed in most of existing literature to the best of our knowledge.

Under the above assumption, the optimal beginning inventory of a single-period problem must lie in the region $U := \{(Y_1, Y_2) : Y_i \leq \beta_i, i = 1, 2\} \bigcap \mathbb{N}^2$. Thus, in the following, we define the set of admissible policies of our problem to be the set of all base-stock policies with order-up-to level $(Y_1, Y_2) \in U$.

9.1 The Learning Algorithm

The online learning Parallel Implementation and Optimization (PIO) Algorithm 1 operates as follows. We maintain two series of variables, $\vec{Y}_t$ and $\vec{Z}_t$. The series $\vec{Z}_t$'s are defined on the extended continuous domain $\bar{U}$ and are used to keep track of the decision points instructed by a subgradient approach that requires continuity. The series $\vec{Y}_t$'s are implementable integer decision variables that are (randomly) chosen based on the values of $\vec{Z}_t$'s. These two series of variables are interconnected and updated iteratively through two subroutines, as illustrated in Figure 9.1. First, the realized demand and cost information from implementing a decision $\vec{Y}_t$ is used to reveal the subgradient (the opposite direction of improvement) at $\vec{Z}_t$. The subgradient at $\vec{Z}_t$ and $\vec{Z}_t$ itself are subsequently used to calculate $\vec{Z}_{t+1}$. The calculation of the subgradient at $\vec{Z}_t$ is accomplished by employing an Order-based Subgradient Calculation (O-SGD) subroutine, i.e., Subroutine 1, which is devised based on the structural properties (convexity and submodularity) obtained in Chapter 8. Second, we employ a Triangular-based Probabilistic Rounding (TriPR) subroutine, i.e., Subroutine 2, to obtain a target order-up-to level $\rho(\vec{Z}_{t+1})$. The target order-up-to level, together with the carry-over inventory $\vec{X}_{t+1}$ from period t, is used to determine the order-up-to level for implementation. Note that when the carry-over inventory from the previous period surpasses the target order-up-to level in the current period, the firm should refrain from placing an order, and therefore the actual implemented order-up-to level should be set as the carry-over inventory.

As can be seen, the two subroutines play crucial roles in ensuring the optimality and implementability of the algorithm. In what follows, we introduce their definitions and explain the insights.

Algorithm 1: PIO Algorithm

(1) **Input:** Randomly choose $\vec{Y}_1$ from the feasible set U, set $\vec{Z}_1 = \vec{Y}_1$.

(2) **for** $t = 1, 2, \cdots, T$ **do**

(3) Observe the realized demand sequence $\mathcal{D}_t^{seq}$, calculate $\mathcal{M}_{t,i}(\vec{Y}^{sgd}, \mathcal{D}^{seq})$ and $\mathcal{T}_{t,i}(\vec{Y}^{sgd}, \mathcal{D}^{seq})$ accordingly.

(4) Calculate $\Delta = (Z_{t,1} - \lfloor Z_{t,1} \rfloor) - (Z_{t,2} - \lfloor Z_{t,2} \rfloor)$, set

$$\vec{Y}^{sgd} = \begin{cases} (\lfloor Z_{t,1} \rfloor, \lceil Z_{t,2} \rceil) & \text{if } \Delta < 0, \\ (\lceil Z_{t,1} \rceil, \lfloor Z_{t,2} \rfloor) & \text{otherwise.} \end{cases}$$

and $\vec{v}(\Delta) = ((-1)^{\mathbb{I}\{\Delta \geq 0\}}, (-1)^{\mathbb{I}\{\Delta < 0\}})$.

(5) Calculate $\vec{Z}_{t+1} = \mathbf{P}_{\bar{U}}(\vec{Z}_t - \alpha_t \vec{H}^{\vec{v}(\Delta)}(\vec{Y}^{sgd}, \mathcal{D}_t^{seq}))$ where $\bar{U} = [0, \beta_1] \times [0, \beta_2]$ is the feasible region, $\alpha_t = \frac{\gamma}{\sqrt{t}}$ is the step size, and $\vec{H}^{\vec{v}(\Delta)}(\vec{Y}^{sgd}, \mathcal{D}^{seq}))$ is a subgradient given by Subroutine 1.

(6) Calculate the carry-over inventory from period t to period $t+1$ as $\vec{I}_t = (Y_{t,1} - \hat{D}_{t,1}, Y_{t,2} - \hat{D}_{t,2})^+$.

(7) Obtain $\vec{Y}_{t+1}$ from $\rho(\vec{Z}_{t+1}) \vee \vec{I}_t$, where $\rho(\vec{Z}_{t+1})$ is a random function defined in Subroutine 2.

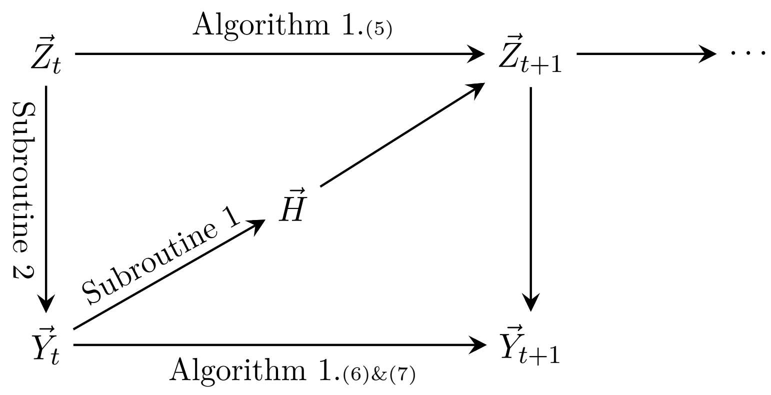


Figure 9.1: A road map for the algorithm

9.1.1 Calculating Subgradient and Updating $\vec{Z}_t$

Note that subgradient descent methods only apply to convex functions defined on a continuous domain, as their asymptotic optimality generally requires diminishing step sizes. Therefore, in the algorithm, we maintain a sequence of continuous variables $\vec{Z}_t$'s for online optimization. Two major challenges arise in the development of this continuation approach.

First, how should we extend the discrete cost function to a fractional point $\vec{Z}_t$? Since our objective is to minimize the expected cost, the extension must be defined for the expected cost function (which is unknown and cannot be perfectly revealed even with abundant data) rather than case by case for each sample path (which can be observed). This means that the definition of the extension should rely on the shape of the expected cost function. Specifically, suppose we use a convex hull extension; in this case, the value of the extension function $\hat{\pi}^b(\cdot)$ at a fractional point $\vec{Z}$ is defined by its function value on the down-facing surface of the convex hull of the graph of π^b, i.e., $\mathrm{Hull}(\{\Gamma_{\pi^b}(\vec{Y}) : \vec{Y} \in U\})$. Then, the graph of the extended cost function is piecewise linear, with each linear segment forming a triangle. This implies that $\Gamma_{\hat{\pi}^b}(\vec{Z})$ can be uniquely defined as a convex combination of the three neighboring corners of the convex extension graph. However, which three corners among the four nearest corners of the unit square partition of $\mathbb{R}^2$ where $\vec{Z}$ resides should be used to define the function value at $\vec{Z}$? It depends on the local shape of the graph of the extended cost function. For example, the graph of a function illustrated in Figure 9.2.(a) results in linear segmentation as shown in Figure 9.2.(c) when projected onto the first two dimensions. In this case, for every unit-sized integer square, the upper-left and lower-right triangles are used to determine the linear continuation for fractional points inside them. Conversely, the graph of a function illustrated in Figure 9.2.(b) leads to linear segmentation depicted in Figure 9.2.(d). Here, linear continuation is generated based on the upper-left and lower-right triangles in some unit-sized integer squares (e.g., $CDFE$), while in others (e.g., $ABDC$) it is based on the upper-right and lower-left triangles, showing inconsistency across the domain.

This situation seemingly creates a dilemma in our continuation approach: we attempt to solve the online learning and optimization problem by applying a continuation extension; however, an explicit continuation extension requires knowledge of the shape of the cost function. Such a dilemma is resolved through an in-depth analysis of the structural properties of our problem. We have proven that the cost function is convex-extensible and submodular. These properties ensure that the convex hull extension $\hat{\pi}^b$ always results in a linear segmentation illustrated in Figure 9.2.(c), which is consistent across the domain and does not rely on the unknown function values that need to be learned. This allows us to provide an explicit definition for the continuity extension of the cost function at

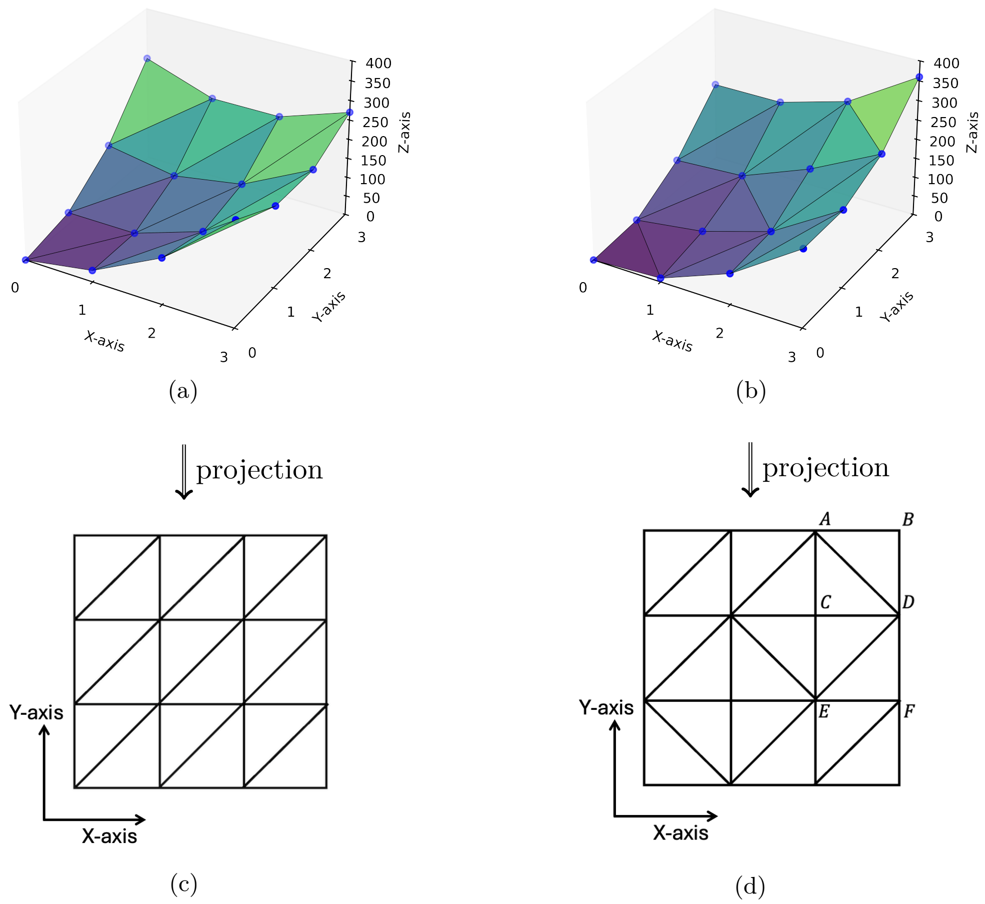


Figure 9.2: Two different patterns of the cost function

any fractional point $\vec{Z}$ by

$$\hat{\pi}^b(\vec{Z}) := \begin{cases} [1-(Z_1-\lfloor Z_1\rfloor)-(\lceil Z_2\rceil-Z_2)]\pi^b(\lfloor Z_1\rfloor,\lceil Z_2\rceil) \\ \quad +(Z_1-\lfloor Z_1\rfloor)\pi^b(\lceil Z_1\rceil,\lceil Z_2\rceil) \\ \quad +(\lceil Z_2\rceil-Z_2)\pi^b(\lfloor Z_1\rfloor,\lfloor Z_2\rfloor), & \text{if } (Z_1-\lfloor Z_1\rfloor)\leq(Z_2-\lfloor Z_2\rfloor), \\ [1-(Z_2-\lfloor Z_2\rfloor)-(\lceil Z_1\rceil-Z_1)]\pi^b(\lceil Z_1\rceil,\lfloor Z_2\rfloor) \\ \quad +(Z_2-\lfloor Z_2\rfloor)\pi^b(\lceil Z_1\rceil,\lceil Z_2\rceil) \\ \quad +(\lceil Z_1\rceil-Z_1)\pi^b(\lfloor Z_1\rfloor,\lfloor Z_2\rfloor), & \text{otherwise.} \end{cases}$$

This formulation does not require learning the unknown expected cost function. As can be seen, the structural properties derived in Chapter 8 play a crucial role in developing the continuation approach and cannot be obtained without an in-depth analysis of the operational characteristics of the problem. For ease of reference, we define the following two concepts to describe the linear segmentation where a point $\vec{Z}$ lies in under our continuity extension. First, the *neighborhood square* of a point $\vec{Z}$ is defined as the unit-sized integer square where $\vec{Z}$ resides; choose the one on the left and/or above of the point if there are multiple (which happens when Z_1 and/or Z_2 are integer). Second, we define the *segmentation triangle* of a point $\vec{Z}$ as the linear segmentation triangle that contains $\vec{Z}$, which is either the upper-left or lower-right triangle of the neighborhood square of $\vec{Z}$; choose the one on the left and/or above of the point if there are multiple (which happens when $\vec{Z}$ is on the boundary or at the vertex).

The second challenge lies in the revelation of the accurate subgradient for decision update, which is complicated by the decomposition of the optimization and implementation processes. Specifically, we need to estimate the subgradient at the point $\vec{Z}_t$, which has a fractional value, to update the optimal points in the online optimization process. However, the inventory decision is implemented and the cost is realized at integer-valued $\vec{Y}_t$. Thus, the critical issue here is how we can transform the information collected at an implemented (integer) point ($\vec{Y}_t$) to the information used for the update at a fractional point ($\vec{Z}_t$). The method we used to address this issue is based on the convex-submodular continuation defined above. Under our continuity extension, the (sub)gradient at a fractional point $\vec{Z}$ can be prescribed by the plane where segmentation triangle of the point resides. Therefore, we propose the following Subroutine 1 to calculate the subgradient at $\vec{Z}$ by a subgradient at the right-angled vertex of its segmentation triangle (which is either the upper-left or the lower-right vertex of its neighborhood square).

Subroutine 1: O-SGD Subroutine

Denote the left and right differential along the axis-i at an integer point $\vec{Y}$ by $H_i^-(\vec{Y}, \mathcal{D}^{seq})$ and $H_i^+(\vec{Y}, \mathcal{D}^{seq})$, which can be calculated by

$$\begin{aligned} H_i^-(\vec{Y}, \mathcal{D}^{seq}) = & \, h_i - (h_i + b_i)\,\mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty\} \\ & - s\,\mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) \le \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12\} \\ & + s\,\mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) > \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12\}, \\ H_i^+(\vec{Y}, \mathcal{D}^{seq}) = & \, H_i^-(\vec{Y} + \vec{e}_i, \mathcal{D}^{seq}). \end{aligned}$$

Define

$$\begin{aligned} \vec{H}^{(-1,+1)}(\vec{Y}, \mathcal{D}^{seq}) &:= \Big(H_1^-(\vec{Y}, \mathcal{D}^{seq}),\ H_2^+(\vec{Y}, \mathcal{D}^{seq})\Big), \\ \vec{H}^{(+1,-1)}(\vec{Y}, \mathcal{D}^{seq}) &:= \Big(H_1^+(\vec{Y}, \mathcal{D}^{seq}),\ H_2^-(\vec{Y}, \mathcal{D}^{seq})\Big), \end{aligned}$$

which are the subgradients at $\vec{Y}$ prescribed by the segmentation triangles to its northwest and southeast, respectively.

Explanation of Subroutine 1. By the segmentation pattern, only the subgradients toward the northwest and the southeast direction at an integer point are relevant for the calculation of gradients at fractional points. Therefore, we only define $\vec{H}^{(-1,+1)}(\vec{Y}, \mathcal{D}^{seq})$ and $H_i^+(\vec{Y}, \mathcal{D}^{seq})$ in Subroutine 1.

The function $H_i^-(\vec{Y}, \mathcal{D}^{seq})$ denotes the left-differential of product i at a given point $\vec{Y}$ for a sample path $\mathcal{D}^{seq}$, which is the cost difference if the inventory of product i were **reduced** by 1. It depends on whether and how stockout events happen in the demand realization in a period.

1. <u>If Y_i product i is sold out during the period $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty)$:</u>

   Reducing a unit of product i will cause a more backorder of product i, incurring an additional cost of b_i. This potential cost difference is captured by a component $-b_i \mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty\}$.

   If the order that consumes the last unit of product i is of type-i, reducing one unit of inventory of product i influences the shipping time (immediately or after backlog) of this order but does not change the number of shipments needed, i.e., not influencing the shipping cost. The number of shipments is influenced only when the last unit of product i is consumed by a type-12 order, i.e., $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12$. Specifically, the number of shipments needed to fulfill this order may increase or decrease as follows.

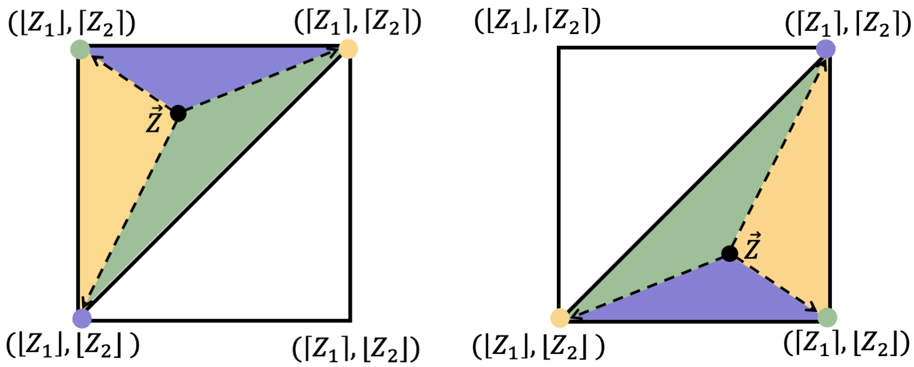
   (a) When product j still has a nonnegative inventory position upon the stock-out of product i $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) \leq \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty)$, this order is originally shipped immediately by a single package. The inventory reduction causes the product i in this order to be backlogged and shipped later, increasing the number of shipments from one to two.

   (b) When product j has already been sold out before the stock-out of product i $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) > \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty)$, the original fulfillment arrangement of this order is that the product i is shipped immediately and the product j is backlogged, requiring two shipments. The inventory reduction makes the product i in this order to be backlogged as well, reducing the number of shipments needed.

   Note that we eliminate the condition $\mathcal{M}(\vec{Y}, \mathcal{D}^{seq}) < \infty$ when we already have $\mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12$ because the latter implies the former.

2. <u>If Y_i product i is not sold out during the period $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty)$:</u>

   Reducing one unit of product i will decrease one unit of unsold inventory, leading to a cost reduction of h_i. This potential cost difference is captured by a component $h_i \mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty\} = h_i[1 - \mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty\}]$. The shipping cost is not affected in this scenario.



(a) $(Z_1 - \lfloor Z_1 \rfloor) \leq (Z_2 - \lfloor Z_2 \rfloor)$ (b) $(Z_1 - \lfloor Z_1 \rfloor) > (Z_2 - \lfloor Z_2 \rfloor)$

Figure 9.3: Illustration of probabilistic rounding

9.1.2 Probabilistic Rounding for an Implementable $\vec{Y}_{t+1}$

This process helps to identify implementable order-up-to levels for each period. The simplest way to convert a fractional number to an integer value is to round it to the nearest integer. However, this method may perform poorly and introduce accumulating errors when the learning algorithm has been implemented for many rounds and the step size of the update is small. The issue with this straightforward and naive method lies in its deterministic nature, which makes it impossible for the algorithm to escape from a suboptimal point if the adjustment (equal to the step size times the subgradient) is smaller than the rounding-off. To avoid such a trap in rounding, we need a mechanism to jump away from a current decision with some chance when performing rounding. However, such jumps should not occur too frequently, as they may hinder convergence to the optimal decision point. Based on these ideas, we devise the following probabilistic rounding subroutine, denoted as $\rho(\vec{Z})$, which is an extension of the probabilistic rounding method used in the literature (Huh et al. 2009). The subroutine involves sampling a decision $\vec{Y}$ from the three vertices of the segmentation triangle at $\vec{Z}$. As illustrated in Figure 9.3, the probability of rounding to each vertex, labeled in different colors, of the segmentation triangle is proportional to the area of the corresponding small triangle (in a position opposite to the rounded-to vertex) of its color. This operation ensures that $\vec{Y}$ is an unbiased estimate of $\vec{Z}$.

In brief, a fractional point $\vec{Z}$ can be uniquely expressed as a convex combination of the three vertices of its segmentation triangle. The subroutine rounds $\vec{Z}$ to each of the three vertices with a probability equal to its coefficient in the expression of the convex combination. Such a treatment guarantees that the expectation of the coordinates of the three potential sampled (integer) points is equal to the original (fractional) point.

Subroutine 2: TriPR Subroutine

Given a $\vec{Z}$, we select a decision $Y = \rho(\vec{Z})$, where

$$\rho(\vec{Z}) := \begin{cases} (\lfloor Z_1 \rfloor, \lfloor Z_2 \rfloor) & \text{with prob. } \lceil Z_2 \rceil - Z_2, \\ (\lceil Z_1 \rceil, \lceil Z_2 \rceil) & \text{with prob. } Z_1 - \lfloor Z_1 \rfloor, \\ (\lfloor Z_1 \rfloor, \lceil Z_2 \rceil) & \text{with prob. } 1 - (\lceil Z_2 \rceil - Z_2) - (Z_1 - \lfloor Z_1 \rfloor). \end{cases}$$

if $(Z_1 - \lfloor Z_1 \rfloor) \leq (Z_2 - \lfloor Z_2 \rfloor)$, and

$$\rho(\vec{Z}) := \begin{cases} (\lfloor Z_1 \rfloor, \lfloor Z_2 \rfloor) & \text{with prob. } \lceil Z_1 \rceil - Z_1, \\ (\lceil Z_1 \rceil, \lceil Z_2 \rceil) & \text{with prob. } Z_2 - \lfloor Z_2 \rfloor, \\ (\lceil Z_1 \rceil, \lfloor Z_2 \rfloor) & \text{with prob. } 1 - (\lceil Z_1 \rceil - Z_1) - (Z_2 - \lfloor Z_2 \rfloor), \end{cases}$$

if $(Z_1 - \lfloor Z_1 \rfloor) > (Z_2 - \lfloor Z_2 \rfloor)$.

By utilizing the piecewise-linearity property of our continuity extension, we are able to prove that the expected cost obtained at $\vec{Y}$ is an unbiased estimate of the cost at $\vec{Z}$. It is noteworthy that our probabilistic rounding is different from simply applying the existing single-dimensional rounding method to each dimension independently. It is built on the two-dimensional structural properties (convexity and submodularity) and our continuity extension with piecewise-linear triangular segments, and thus is named Triangular-based Probabilistic Rounding (TriPR).

9.2 Performance of the Algorithm

In this section, we first prove that our PIO algorithm for the backlog case is asymptotically optimal in the long run, yielding a regret no worse than $O(\sqrt{T})$. We further prove that any admissible policy cannot achieve a regret better than $\Omega(\sqrt{T})$ for all cases. Therefore, our algorithm attains the best achievable performance $\Theta(\sqrt{T})$ in the limit.

9.2.1 Regret Upper Bound

Firstly, we establish the upper bound of the regret in Theorem 2. This result is consistent with the regret upper bound presented in Huh and Rusmevichientong (2009) and Shi et al. (2016) for classic inventory control problems without consideration of shipping cost and demand correlation, thereby providing a validation of the efficiency and effectiveness of our algorithm.

Theorem 2 (Asymptotic Optimality)**.** *The PIO algorithm can achieve a sublinear regret* *bound in the case of backlog. Specifically,*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi^b(\vec{Y}_t) - \pi^b(\vec{Z}^*))\right] \leq C^b\sqrt{T},$$

where C^b is a constant independent of T.

The framework of the proof is as follows. We use the term *optimal* order-up-to level to refer to the clairvoyant optimal solution $\vec{Z}^*$, which equals to $\vec{Y}^*$ under our convex-continuity extension. Note that our problem is dynamic in the way that the system state depends on the firm's historical decisions and demand realization, i.e., the leftover inventory affects the feasible region of the decision in a new period. Thus, losses may not only come from a suboptimal choice of order-up-to level due to insufficient exploration but also result from the constraint imposed by stochastic system dynamics. Therefore, to establish the upper bound of the regret, we bridge the *actually implemented* order-up-to level $\vec{Y}_t = \rho(\vec{Z}_t) \vee \vec{X}_t$, which is a truncation function, and the optimal order-up-to level $\vec{Z}^*$ by using a "mediator" $\vec{Y}_t^\rho = \rho(\vec{Z}_t)$, referred to as the *target* order-up-to level sampled by the TriPR subroutine. In this case, the total loss of the algorithm can be decomposed into

$$\begin{aligned}
\mathbb{E}\left[\sum_{t=1}^{T}(\pi^b(\vec{Y}_t) - \pi^b(\vec{Z}^*))\right] = \; & \mathbb{E}[\sum_{t=1}^{T}(\pi^b(\vec{Z}_t) - \pi^b(\vec{Z}^*))] && \text{(Exploration loss)} \\
& + \mathbb{E}[\sum_{t=1}^{T}(\pi^b(\vec{Y}_t^\rho) - \pi^b(\vec{Z}_t))] && \text{(Rounding error)} \\
& + \mathbb{E}[\sum_{t=1}^{T}(\pi^b(\vec{Y}_t) - \pi^b(\vec{Y}_t^\rho))], && \text{(Truncation loss)}
\end{aligned}$$

where the exploration loss is caused by the mismatch between the true optimal $\vec{Z}^*$ and the "learned optimal" $\vec{Z}_t$, the rounding error is introduced by probabilistic rounding in each iteration, and the truncation loss is present when the excessive leftover inventory prevents us from using the target order-up-to level. Firstly, our designs of the continuity extension and probabilistic rounding mechanism help to eliminate the rounding error. In the following, we bound the exploration and truncation losses by Lemmas 2 and 3 respectively.

Lemma 2 (Bound on Exploration Loss)**.** *It holds that*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi^b(\vec{Z}_t) - \pi^b(\vec{Z}^*))\right] \leq \left(\gamma_0 + \frac{1}{\gamma_0}\right) diam(U)\bar{B}\sqrt{T},$$

where $\bar{B} = \max\{b_1 + s, b_2 + s, h_1, h_2\}$.

The proof follows the idea used by Theorem 1 in Zinkevich (2003) for a differentiable function with a single-dimensional variable. We adapt the approach for multi-dimensional discrete decisions using the subgradient concept. The main idea of the proof can be illustrated by the inequality

$$\begin{aligned}\mathbb{E}[\pi^b(\vec{Z}_t) - \pi^b(\vec{Z}^*)] \leq & \mathbb{E}\left[\left\langle \vec{H}(\vec{Z}_t), \vec{Z}_t - \vec{Z}^* \right\rangle\right] \\ \leq & \frac{\mathbb{E}||\vec{Z}_t - \vec{Z}^*||^2}{2\alpha_t} - \frac{\mathbb{E}||\vec{Z}_{t+1} - \vec{Z}^*||^2}{2\alpha_t} + \frac{\alpha_t}{2}\mathbb{E}||\vec{H}(\vec{Z}_t)||,\end{aligned}$$

where the first inequality holds because of convexity and the second follows from the decision update formula. The summation of the first two terms over all t's leads to a loss of $C' diam(U)^2\sqrt{T}$ since decision error is bounded by $diam(U)$ and the step size α_t shrinks in the order of $\sqrt{T}$. The summation of the third term over all t's leads to a loss of $C'\bar{B}^2\sqrt{T}$ where $\sqrt{T}$ comes from the summation of the step sizes α_t's. The parameter γ in the PIO algorithm, influencing the coefficients C' and C'', is then chosen to balance the squared terms $C' diam(U)^2$ and $C''\bar{B}^2$, yielding a factor $diam(U)\bar{B}$ for $\sqrt{T}$. Detailed proof can be found in Appendix B.1.

Lemma 3 (Bound on Truncation Loss)**.** *It holds that*

$$\mathbb{E}[\sum_{t=1}^{T}(\pi^b(\vec{Y}_t) - \pi^b(\vec{Y}_t^\rho))] \leq C_1\sqrt{T},$$

where C_1 is a constant which only relies on h_1, h_2, s, γ and the demand distribution.

The main idea of the proof is as follows. By the convexity and Lipschitz-continuity, we can bound the cost difference by the difference between $Y_{t,i}$ and $Y_{t,i}^\rho$. Note that $Y_{t+1} = Y_{t+1}^\rho \vee I_t$ and $Y_{t+1} \neq Y_{t+1}^\rho$ only when $Y_{t+1,i}^\rho < Y_{t,i} - \hat{D}_{t,i}$ for some i, i.e., the demand is sufficiently small that the carry-over inventory is larger than the target order-up-to level. Therefore, we can express the difference as

$$\begin{aligned}\vec{Y}_{t+1} - \vec{Y}_{t+1}^\rho &= (\vec{Y}_t - \vec{\hat{D}}_t - \vec{Y}_{t+1}^\rho)^+ \\ &= \left(\vec{Y}_t - \vec{\hat{D}}_t - \rho(\vec{Z}_t - \alpha_t H(\vec{Z}_t))\right)^+ \\ &\leq \left(\vec{Y}_t - \vec{\hat{D}}_t - \rho(\vec{Z}_t) + \alpha_t H(\vec{Z}_t) + \vec{1}\right)^+ \\ &= \left(\vec{Y}_t - \vec{Y}_t^\rho + \alpha_t H(\vec{Z}_t) - (\vec{\hat{D}}_t - \vec{1})\right)^+ \\ &\leq \left(\vec{Y}_t - \vec{Y}_t^\rho + \alpha_t \vec{h} - (\vec{\hat{D}}_t - \vec{1})\right)^+.\end{aligned}$$

This inequality suggests a recursive relationship resembling a Lindley equation for a single-server FIFO queue where we view $(\vec{Y}_t - \vec{Y}_t^\rho)$ as the waiting time, $\alpha_t\vec{h}$ as the service time, and $(\vec{\hat{D}}_t - \vec{1})$ as the interarrival time. To establish an upper bound, we construct a related $GI/G/1$ queue (a Lindley process) where we can bound the expected waiting time, $Y_{t,i} - Y^{\rho}_{t,i}$, of the tth "customer" by the mean duration of a busy period of the process. In B.2, we provide a detailed mathematical formulation of the process and proof. The step size in each iteration of our PIO algorithm plays a crucial role in the proof as it determines the magnitude of change in the target inventory level. Note that similar construction is also used by Huh and Rusmevichientong (2009) and Shi et al. (2016). Notably, we establish an upper bound by a new proof without the need for additional assumptions except for the finite demand expectation that we have assumed for our model.

9.2.2 Regret Lower Bound

In this section, we establish the result that the regret of any admissible policy is at least in the order of $\Omega(\sqrt{T})$, which implies that our PIO algorithm yields the best achievable regret in an asymptotic sense.

Note that in the beginning stage of the problem, making errors is unavoidable due to the limited information the firm has about demand distribution. If the firm were to *explore* the optimal order-up-to level and *then commit* to it (ETC), the regret could grow linearly over time if the estimations are not accurate enough to reveal the true optimal decision, leading the firm to commit to a suboptimal decision. Therefore, the firm should refrain from adopting an ETC policy. Instead, by perpetually updating the order-up-to level, the firm can mitigate the risk of making constant errors. Noting that this approach does not eliminate the possibility of making incorrect decisions but controls the size of error over time. To ensure that this cost difference is sublinear, it is crucial to implement appropriate updating procedures.

Theorem 3 (Regret Lower Bound)**.** *The regret for any admissible policy addressing our problem in the case of backlog is bounded from below by* $\Omega(\sqrt{T})$*.*

Because demand is unknown, the firm must sufficiently explore to differentiate among the possible demand distributions to identify the true optimal order-up-to level. Information theory and statistical inequalities can be used to estimate the exploration cost needed in this process. For example, Keskin and Zeevi (2014) utilize the Van Tree inequality and the concept of Fisher information, and Besbes and Zeevi (2011); Broder and Rusmevichientong (2012) and Besbes and Muharremoglu (2013) utilize Kullback-Leiber divergence to establish the order of worst-case regret.

To prove the lower bound for the regret of our problem, we construct a special case where the exploration cost is shown to be in the order of $\Omega(\sqrt{T})$. The intuition behind our construction is that if two demand distributions F_a and F_b, are "similar" but have different order-up-to levels, the algorithm will struggle to differentiate between them and inevitably make errors. Using an information-theoretical argument, we bound the regret from below by a transformation of the "distance" between these two distributions, which is captured by the *Kullback-Leibler divergence*. As a result, we establish that a regret of $\sqrt{T}$ is inevitable. It is important to note that this regret is not caused by demand censoring (because the demand is observable in the backlogged case). Instead, it arises due to the difficulty in distinguishing between the two "similar" distributions and consequently finding the optimal order-up-to inventory level. Readers are referred to C.1 for detailed construction and proof. It is worth mentioning that our lower bound of $\Omega(\sqrt{T})$ is consistent with the results obtained in existing inventory studies with learning without accounting for the fulfillment process. The lower bound in Besbes and Muharremoglu (2013) can be further reduced to $\Omega(\log T)$ under a stronger assumption of minimal separation around optimal quantity.

9.3 The Case of Lost-Sales

In the lost-sales case, demand is censored and hence the firm cannot differentiate whether the inventory is just enough or there are lost sales. Furthermore, in our problem, the demand censoring issue also hinders the firm from learning the potential savings/spending in shipping caused by a change of inventory.

To address this issue, existing nonparametric studies with unknown discrete demand distributions commonly assume that a lost-sales indicator function is known (Huh and Rusmevichientong, 2009; Besbes and Muharremoglu, 2013; Lyu et al., 2024). We follow this practice and make the assumption that the firm has knowledge of the indicator functions

$$\mathbb{I}\left\{\mathcal{M}_i(\lceil \vec{Z}_t \rceil, \mathcal{D}^{seq}) < \infty\right\},$$

and

$$\mathbb{I}\left\{\mathcal{M}_i(\lceil \vec{Z}_t \rceil, \mathcal{D}^{seq}) \leq \mathcal{M}_j(\lceil \vec{Z}_t \rceil, \mathcal{D}^{seq}), \mathcal{T}_i(\lceil \vec{Z}_t \rceil, \mathcal{D}^{seq}) = 12\right\}.$$

As noted in Besbes and Muharremoglu (2013), active exploration is necessary in the case of discrete demand, while a lost-sales indicator function enables "free experimentation." Since our problem is multi-dimensional and involves complex fulfillment dynamics, we require indicator functions that are slightly stronger than those in the existing literature for single-product problems.

9.3.1 The Learning Algorithm

The main algorithm is the same as Algorithm 1, while the order-based subgradient calculation (Subroutine 1) is replaced by Subroutine 3. The differences are in the expression of the left-subderivative when the inventory of a product is reduced by 1. Noting that unfilled demand is lost, there is no need to ship a type-12 order twice. Moreover, losing part of a type-12 order will result in a loss of p_i, but further losing the remaining part of the order will incur an additional profit (and goodwill) loss of $p_j - s$. Therefore, a lost-sales indicator with type information is needed to instruct the true cost difference.

Subroutine 3: O-SGD Subroutine in the Case of Lost Sales

Define

$$\vec{H}^{(-1,+1)}(\vec{Y}, \mathcal{D}^{seq}) := \Big(H_1^-(\vec{Y}, \mathcal{D}^{seq}),\ H_2^-(\vec{Y} + \vec{e}_2, \mathcal{D}^{seq})\Big),$$
$$\vec{H}^{(+1,-1)}(\vec{Y}, \mathcal{D}^{seq}) := \Big(H_1^-(\vec{Y} + \vec{e}_1, \mathcal{D}^{seq}),\ H_2^-(\vec{Y}, \mathcal{D}^{seq})\Big).$$

Calculate $(H_1^-(\vec{Y}, \mathcal{D}^{seq}), H_2^-(\vec{Y}, \mathcal{D}^{seq}))$ as follows:

$$\begin{aligned} H_i^-(\vec{Y}, \mathcal{D}^{seq}) =& h_i - (h_i + p_i - s)\, \mathbb{I}\{\mathcal{M}_1(\vec{Y}, \mathcal{D}^{seq}) < \infty\} \\ & - s\, \mathbb{I}\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) \leq \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12\}, \\ H_i^+(\vec{Y}, \mathcal{D}^{seq}) =& H_i^-(\vec{Y} + \vec{e}_i, \mathcal{D}^{seq}). \end{aligned}$$

Explanation of Subroutine 3. The function $\vec{H}_i^-(\vec{Y}, \mathcal{D}^{seq})$ is the left-subdifferential of product i at a given point $\vec{Y}$ for a sample path $\mathcal{D}^{seq}$, which is the cost difference if the inventory of product i were reduced by 1. It, again, depends on whether and how stockout events happen in the demand realization in a period.

1. <u>If Y_i product i is sold out during the period ($\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty$):</u>
   Reducing a unit of product i will cause one more lost-sales cost of product i, incurring an additional cost of p_i. This potential cost difference is captured by a component $-p_i \mathbb{I}\left\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty\right\}$.
   If the order that consumes the last unit of product i is of type-i, reducing one unit of inventory of product i reduces the shipping cost by one unit as. If, on the other hand, the order that consumes the last unit of i is of type-12, the number of shipments may or may not change as follows.

   (a) When product j still has a nonnegative inventory upon the stock-out of product i ($\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) \leq \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty$), this order is originally shipped to the customer. The inventory reduction will not affect the number of shipments because product j has to be shipped anyway.

   (b) When product j has already been sold out before the stock-out of product i ($\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) > \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty$), the inventory reduction will save a unit of shipment because product i in this order cannot be fulfilled.

2. If Y_i product i is not sold out during the period $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty)$:
   Reducing one unit of product i will decrease one unit of unsold inventory, leading to a cost saving of h_i. This potential cost difference is captured by a component $h_i \mathbb{I}\left\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty\right\} = h_i[1 - \mathbb{I}\left\{\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty\right\}]$. The lost-sales cost and shipping cost is not affected in this scenario.

Next, we introduce the derivation of the true left-subderivative $\vec{H}^{(-1,-1)}(\vec{Y}, \mathcal{D}^{seq})$. The ith component of this left-subderivative, $\vec{H}_i^-(\vec{Y}, \mathcal{D}^{seq})$, represents the marginal reduction in cost when the order-up-to level of product i, Y_i, decreases by 1 unit $(i = 1, 2)$. To express this marginal decrease in terms of accessible observations, we now consider the following cases:

1. When all Y_i units of product i can be sold out within the period $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) < \infty)$, reducing one unit of product i from the initial inventory level corresponds to a unit of lost sales cost. Meanwhile, if at the depletion time of product i, product j has non-negative inventory and the incoming order requiring the last unit of product i is of type-12 $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) \le \mathcal{M}_j(\vec{Y}, \mathcal{D}^{seq}), \mathcal{T}_i(\vec{Y}, \mathcal{D}^{seq}) = 12)$, although the demand for product i is lost, the demand for product j can still be satisfied, leading to unchanged shipping cost; otherwise, the shipping cost decreases by one unit.

2. If product i has a positive inventory position at the end of the period when the initial order-up-to level is Y_i $(\mathcal{M}_i(\vec{Y}, \mathcal{D}^{seq}) = \infty)$, deducting one unit from the order-up-to level of product i will result in a unit of decrease in holding cost. Meanwhile, the shipping cost remains unchanged.

By considering the different cases described above, we now conclude the expression of the "left" subgradient in the algorithm using the available observations.

9.3.2 Algorithm Performance

9.3.2.1 Regret Upper Bound.

The regret upper bound is in the same order, $O(\sqrt{T})$, as that in the case of backlog.

Theorem 4 (Asymptotic Optimality)**.** *The PIO algorithm can achieve a sublinear regret bound in the case of lost sales. Specifically,*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi^\ell(\vec{Y}_t) - \pi^\ell(\vec{Z}^*))\right] \le C^\ell \sqrt{T},$$

where C^ℓ is a constant independent of T.

9.3.2.2 Regret Lower Bound.

Unlike the case of backlog, in the case of lost sales, we do not have access to fully revealed demand information in each period. However, we can prove that even with uncensored demand in each period, the worst-case regret of any policy addressing our problem is lower bounded by $\Omega(\sqrt{T})$. The theorem is stated as follows.

Theorem 5 (Regret Lower Bound)**.** *The worst-case regret for all policies addressing our problem in the case of lost sales is bounded from below by* $\Omega(\sqrt{T})$ *even under uncensored demand.*

The proof follows a similar approach to that in the case of backlog but requires a different construction of demand distributions. Readers are referred to C.2 for details.

Chapter 10

Numerical Experiment

In this chapter, we conduct numerical experiments to show the effectiveness of the PIO algorithm and present a numerical comparative study to demonstrate the value of accounting for the fulfillment process in inventory planning.

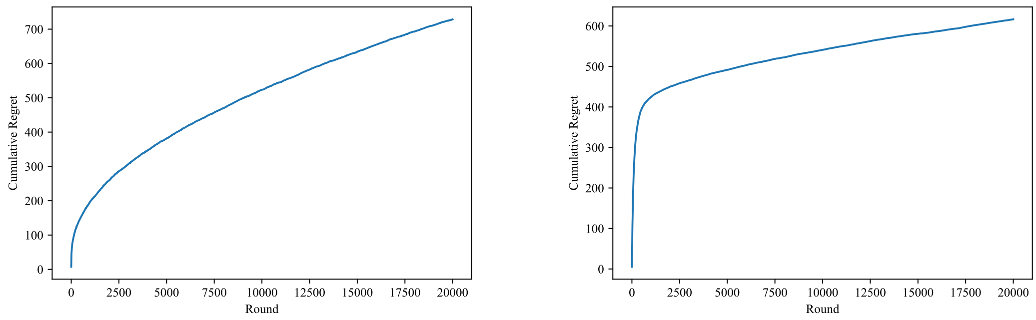
For ease of exposition, let $U[D^{\min}, D^{\max})$ denote a discrete uniform distribution wherein each of the whole numbers in $\{D^{\min}, D^{\min}+1, \ldots, D^{\max}-2, D^{\max}-1\}$ are equally likely to be realized.

10.1 Performance of the PIO Algorithm

In this section, we numerically show the converging behavior and the regret of our PIO algorithm under the setting of backlog or lost-sales cases. Specifically, we set the demand for type-1, type-2, and type-12 orders to be $D_1 \sim U[1, 50)$, $D_2 \sim U[1, 10)$, and $D_{12} \sim U[10, 80)$. In the case of backlog, the cost parameters are set to be $h_1 = 0.2$, $b_1 = 0.8$, $h_2 = 0.6$, $b_2 = 0.4$, and $s = 0.8$; in the case of lost sales, $h_1 = 0.2$, $p_1 = 0.8$, $h_2 = 0.6$, $p_2 = 0.4$, and $s = 0.8$, reflecting the same critical ratios as the case of backlog. The initialization of $\vec{Z}_0$ is set to be at the single-product optimal solution, i.e., $Z_{0,1} = \inf\{x \in \mathbb{N} : F_{\hat{D}_1}(x) \geq CR_1 = 0.8\} = 103$ and $Z_{0,2} = \inf\{x \in \mathbb{N} : F_{\hat{D}_2}(x) \geq CR_2 = 0.4\} = 36$. The coefficient of step size $\gamma = 10$, and we run for $T = 20,000$ periods.

Figure 10.1 shows the cumulative regret of the PIO algorithm. As can be seen, the regret is sublinear, indicating the effectiveness and efficiency of our algorithm.

Figure 10.2 exhibits the converging behavior of $\vec{Z}$ and the corresponding implemented order-up-to level $\vec{Y}$. It can be seen that $\vec{Z}$ converges very fast to the clairvoyant optimal solution which is marked as the black horizontal line shown in the figure.



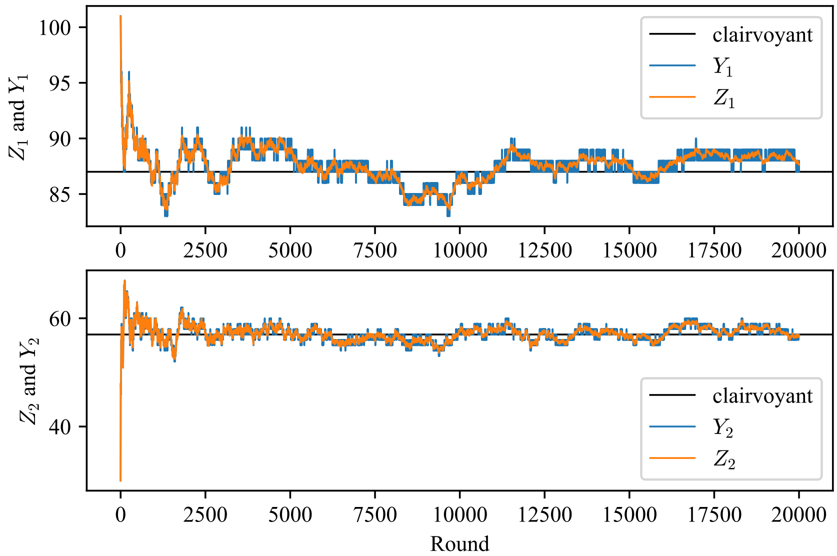
(a) The case of backlog (b) The case of lost sales

Figure 10.1: Cumulative regret of the PIO algorithm

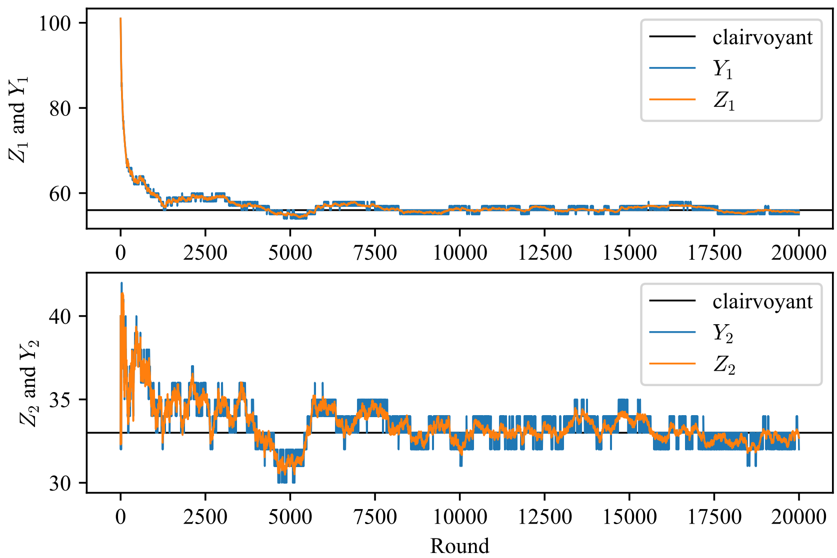
10.2 The Value of Accounting for the Fulfillment Process

In this section, we present a numerical comparative study of a firm's cost when its inventory decision accounts for or does not account for the fulfillment process. We consider two firms. The first firm, referred to as a myopic firm as defined in Chapter 8, plans the inventory of the two products separately by applying the single-product online learning algorithm, denoted by $\tilde{\phi}$, proposed by Huh and Rusmevichientong (2009). The second firm, referred to as a strategic firm, realizes the impact of inventory decisions on the fulfillment process and hence applies our PIO algorithm, denoted by ϕ, to jointly optimize the inventory of the two products.

The setting of our numerical study is as follows. We set the demand for type-1, type-2, and type-12 orders to be $D_1 \sim U[1, 10)$, $D_2 \sim U[1, 10)$, and $D_{12} \sim U[10, 80)$. This corresponds to a scenario where the demand for the two products is highly correlated and a significant proportion of customers tend to purchase them together. We run numerical experiments for different combinations of the values of s, h_i, and b_i or p_i as follows. The unit shipping cost s takes a value in $\{0.8, 0.6, 0.4, 0.2\}$. Noting that the single-product problem is a newsvendor problem with the optimal quantity determined by the critical ratio, we vary the inventory holding and backlog or lost-sales costs to have scenarios where the critical ratios of the two products are both low, or both high, or low and high. Here we regard a critical ratio above 0.5 as high and one below 0.5 as low. Our studies include scenarios where the difference of the critical ratios $CR_2 - CR_1$ ranges between 0.2 and 0.7. The combinations of the parameters are presented in Table 10.1 and 10.2. For each combination of the parameters, we run for $T = 10,000$ periods and evaluate the performance (the total cost and the shipping cost) by taking an average of 100 sample paths.



(a) The case of backlog



(b) The case of lost sales

Figure 10.2: Converging performance of the PIO algorithm

We report the relative cost difference, which is the ratio of the cost difference between the two firms to the cost of the strategic firm, in the table. It is evident that the relative cost difference can reach as high as 17.26% in the case of backlog and 20.06% in the case of lost sales. This finding highlights the importance of accounting for the shipping costs in the inventory planning process in e-commerce, where shipping costs take up a significant proportion of total expenses. It indicates that our extended modeling framework with fulfillment dynamics incorporated has great potential in helping e-commerce firms achieve better cost management and improve overall performance.

Table 10.1: Relative cost difference of the myopic and strategic firms (the backlog case)

Critical ratio (CR)	h_1	b_1	h_2	b_2	$CR_2 - CR_1$	s	Relative cost difference (%)	
							Shipping	Total
Low-High	0.9	0.1	0.2	0.8	0.7	0.8	39.13	17.26
						0.6	33.62	12.56
						0.4	21.03	7.88
						0.2	10.91	3.37
	0.8	0.2	0.2	0.8	0.6	0.8	30.55	13.93
						0.6	28.21	10.39
						0.4	18.21	6.34
						0.2	9.47	2.60
	0.6	0.4	0.2	0.8	0.4	0.8	16.18	8.16
						0.6	15.52	6.43
						0.4	12.81	4.03
						0.2	6.53	1.56
	0.6	0.4	0.4	0.6	0.2	0.8	9.43	5.79
						0.6	9.41	5.04
						0.4	9.10	3.82
						0.2	7.96	1.91
Low-Low	0.8	0.2	0.6	0.4	0.2	0.8	13.35	8.96
						0.6	13.24	7.92
						0.4	13.12	6.36
						0.2	12.22	3.55
High-High	0.4	0.6	0.2	0.8	0.2	0.8	5.58	3.29
						0.6	5.43	2.77
						0.4	5.07	2.01
						0.2	3.69	0.89

Table 10.2: Relative cost difference of the myopic and strategic firms (the lost-sales case)

Critical ratio (CR)	h_1	p_1	h_2	p_2	$CR_2 - CR_1$	s	Relative cost difference (%)	
							Shipping	Total
Low-High	0.9	0.1	0.2	0.8	0.7	0.8	172.06	10.61
						0.6	18.41	3.72
						0.4	5.89	1.51
						0.2	1.89	0.45
	0.8	0.2	0.2	0.8	0.6	0.8	90.31	9.18
						0.6	17.44	3.30
						0.4	5.52	1.31
						0.2	1.76	0.36
	0.6	0.4	0.2	0.8	0.4	0.8	39.31	6.75
						0.6	16.56	2.92
						0.4	5.21	1.12
						0.2	1.61	0.30
	0.6	0.4	0.4	0.6	0.2	0.8	93.99	14.37
						0.6	38.04	8.05
						0.4	18.64	3.92
						0.2	7.06	1.08
Low-Low	0.8	0.2	0.6	0.4	0.2	0.8	17.58	4.58
						0.6	150.22	20.06
						0.4	61.32	10.45
						0.2	20.84	2.98
High-High	0.4	0.6	0.2	0.8	0.2	0.8	16.60	4.11
						0.6	9.28	2.35
						0.4	4.71	1.07
						0.2	1.55	0.28

Chapter 11

Extension: Without Lost-Sales Indicator Function(s)

In the single-product lost-sales case, the indicator function was necessitated in most existing literature like Huh and Rusmevichientong (2009), Besbes and Muharremoglu (2013), and Lyu et al. (2024). As stated in Lugosi et al. (2024), without such an indicator function, a naive implementation of a biased subgradient will lead to linear regret even in the constant demand case. Works by Huh et al. (2011) and Lyu et al. (2024) make use of the Kaplan-Meier estimator, a well-known tool in statistics, to deal with the discrete demand case. However, the algorithm they developed doesn't make use of the inherent convexity nature of the problem.

In this study, we first provide an algorithm that pertains to the subgradient method by devising cycles within which the decision is not updated (only *exploit*) while between two consecutive ones a specified decision is made (*explore*). Hence, our algorithm is named Cycle-based Parallel Implementation and Optimization (Cyc-PIO). The regret of Cyc-PIO is $O(T^{2/3})$, significantly improving the linear regret result demonstrated in Lugosi et al. (2024). Furthermore, we design an Estimator-based PIO (Est-PIO) algorithm which has $O(\sqrt{T})$ regret under a mild assumption that there is only a single optimal decision. Even under this assumption, we can prove that the best-achievable regret is $\Omega(\sqrt{T})$ with the same construction as that in Chapter 9.

Note that without the lost-sales indicator function, our problem can also be regarded as a bandit convex optimization (BCO) problem, where in each round only the loss on the decision point is revealed. Previous works mostly deal with strongly convex loss functions or smooth convex settings (See Table 11.1). Our method works in a piecewise-linear setting (non-smooth and non-strongly-convex) with an inventory management background. It proves that our algorithm has a regret of $O(T^{2/3})$.

In this chapter, we use $\pi(\cdot)$ to denote the expected cost in the single-product problem. Next, we sequentially state our $O(T^{2/3})$-regret Cyc-PIO algorithm and $O(\sqrt{T})$-regret Est-

Setting	Conv.	Lin.	Smooth	Str.-Conv.	Str.-Conv.&Smooth	Piecewise-Lin.
BCO	$\tilde{O}(T^{3/4})$	$\tilde{O}(\sqrt{T})$	$\tilde{O}(T^{2/3})$	$\tilde{O}(T^{2/3})$	$\tilde{O}(\sqrt{T})$	$O(T^{2/3})$ & $O(\sqrt{T})$

Table 11.1: BCO performance

PIO algorithm and their performances.

11.1 Cyc-PIO Algorithm: $O(T^{2/3})$ Regret

We will start with the commonly studied single-product case to provide some heuristics.

11.1.1 Single-Product Case

We first provide a high-level intuition of our algorithm. First, we devise a cycle, and within each cycle, if the true subgradient information is not revealed (because of demand censoring), we continue using the current decision point to do probabilistic rounding. Once the true subgradient information is fully observed, we update the decision point and proceed to the next cycle. However, if we solely rely on probabilistic rounding to update the decision point, the updating process could be slow due to the closeness to the largest integer that is smaller than the decision point. Therefore, we force the cycle to end and implement an order-up-to level that can reveal the subgradient if the length of the cycle reaches a threshold. Note that as the decision point approaches the optimal, the update can be less frequent, enabling exploitation. So, the cycle length is set to be increasing with the index of cycles. Specifically, our single-product Cyc-PIO Algorithm works in the following way.

Cyc-PIO Algorithm: Single-Product

Step 0 (Initialization) Randonly select a Z_1 from the admissible region, and set $Y_1 = \lceil Z_1 \rceil$. The cycle index $k = 1$, and a counter $\psi = 0$.

Step 1 Calculate $\delta_k = Z_k - \lfloor Z_k \rfloor$. If $\delta_k = 0$, implement Z_k, and proceed to Step 4; otherwise, let $n_k = \lceil \sqrt{k} \rceil$.

Step 2 For $\psi < n_k - 1$, do probabilistic rounding on Z_k, and implement the rounded-to decision. If $\rho(Z_k) = \lceil Z_k \rceil$, proceed to Step 4; otherwise $\psi = \psi + 1$, repeat Step 2.

Step 3 If $\psi = n_k - 1$, implement $\lceil Z_k \rceil$, proceed to the next step.

Step 4 Update $Z_{k+1} = Z_k - \alpha_k H(Z_k)$, where $\alpha_k = \frac{\gamma}{\sqrt{k}}$,

$$H(Z_k) = \begin{cases} h & \text{if } d < \lceil Z_k \rceil \\ -p & \text{if } d \geq \lceil Z_k \rceil \end{cases}$$

d is the demand realization in the $(\psi+1)$th round in the kth cycle. Update $k = k+1$, and set $\psi = 0$.

Next, we establish the regret upper bound for our single-product Cyc-PIO algorithm.

Theorem 6 (Regret Upper Bound). *The Cyc-PIO algorithm can achieve a sublinear regret bound. Specifically,*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi(Y_t) - \pi(Z^*))\right] \leq C_1^{cyc} T^{2/3},$$

where C_1^{cyc} is a constant independent of T.

Here, we provide some intuition of the proof. From the PIO algorithm (without cycle), we know that $\pi(Z_k)$ converges to $\pi(Z^*)$ with a difference in the order of $O(\frac{1}{\sqrt{k}})$ in expectation, implying that Z_k converges to Z^* with a difference in the order of $O(\frac{1}{\sqrt{k}})$ in expectation. Therefore, when k is large enough and $Z_k > Z^*$, the probabilistic rounding would hit the optimal point frequently in the cycle, enabling powerful exploitation. When, conversely, k is large enough and $Z_k < Z^*$, the update would be instant (i.e., the cycle length is short), allowing for active exploration. When the forced cycle length is taken to be $\sqrt{k}$, the expected cycle length is in the order of $\Omega(\sqrt{k})$, which helps to bound the number of cycles that $0 < Z_k - Z^* < \frac{1}{2}$. As when $0 < Z^* - Z_k < \frac{1}{2}$, the update is frequent, we hence know that the loss caused by these k's are upper bounded by the PIO regret upper bound. Next, we examine those k's such that $|Z_k - Z^*| \geq \frac{1}{2}$. The probability for $|Z_k - Z^*| \geq \frac{1}{2}$ is in the order of $O(\frac{1}{\sqrt{k}})$. When the forced cycle length is taken to be $\sqrt{k}$, the expected loss in the kth cycle given $|Z_k - Z^*| \geq \frac{1}{2}$ is upper bounded by $O(\sqrt{k})$. Moreover, the total number of rounds such that $|Z_k - Z^*| \geq \frac{1}{2}$ is in the order of $O(\sqrt{K})$. Above all, we can obtain the stated upper bound for the total loss. Interested readers can refer to Appendix D.1.1 for detailed proof. The cumulative regret can be found in Figure 11.1.

11.1.2 Two-Product Case with Fulfillment Dynamics

By adopting a similar idea of the single-product algorithm, we devise the following Cyc-PIO algorithm for our two-product problem with fulfillment dynamics.

Cyc-PIO Algorithm: Two-Product

Step 0 (Initialization) Randomly select a $\vec{Z}_1$ from the admissible region, and set $\vec{Y}_1 = \lceil \vec{Z}_1 \rceil$. The cycle index $k = 1$, and a counter $\psi = 0$.

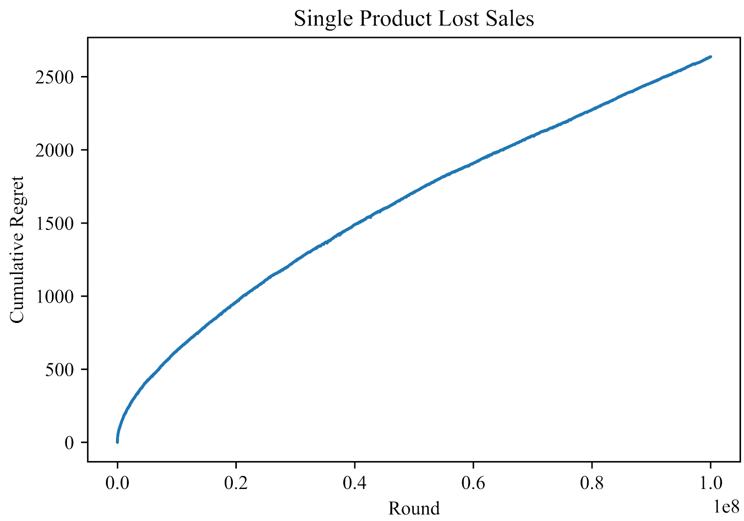


Figure 11.1: Cumulative regret of single-product Cyc-PIO algorithm

Step 1 If $\theta_k = \max\{Z_{k,1} - \lfloor Z_{k,1} \rfloor, Z_{k,2} - \lfloor Z_{k,2} \rfloor\} = 0$, implement Z_k, and proceed to Step 4; otherwise, let $n_k = \sqrt{k}$.

Step 2 For $\psi < n_k - 1$, do probabilistic rounding on $\vec{Z}_k$, and implement the rounded-to decision. If $\rho(\vec{Z}_k) = \lceil \vec{Z}_k \rceil$, proceed to Step 4; otherwise $\psi = \psi + 1$, repeat Step 2.

Step 3 If $\psi = n_k - 1$, implement $\lceil \vec{Z}_k \rceil$, proceed to the next step.

Step 4 Update $\vec{Z}_{k+1} = \vec{Z}_k - \alpha_k H(\vec{Z}_k)$ according to the O-SGD subroutine, where $\alpha_k = \frac{\gamma}{\sqrt{k}}$. Update $k = k + 1$, and set $\psi = 0$.

The performance of our Cyc-PIO exhibits efficiency and effectiveness. Specifically,

Theorem 7 (Regret Upper Bound). *The Cyc-PIO algorithm can achieve a sublinear regret bound. Specifically,*

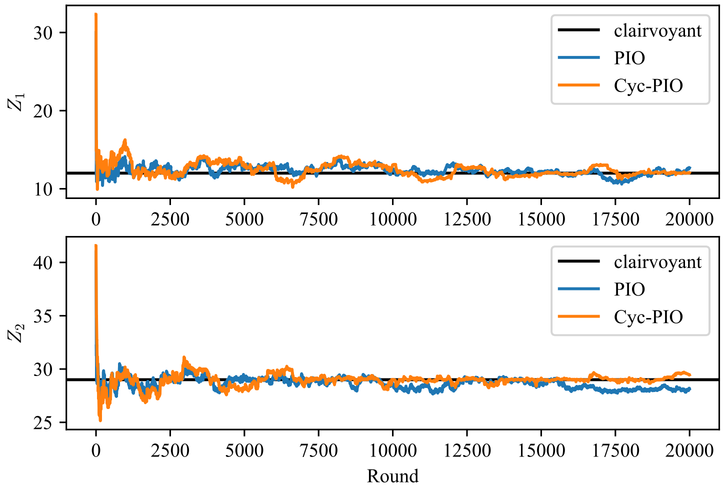
$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi^{\ell}(\vec{Y}_t) - \pi^{\ell}(\vec{Z}^*))\right] \leq C_2^{cyc} T^{2/3},$$

where C_2^{cyc} is a constant independent of T.

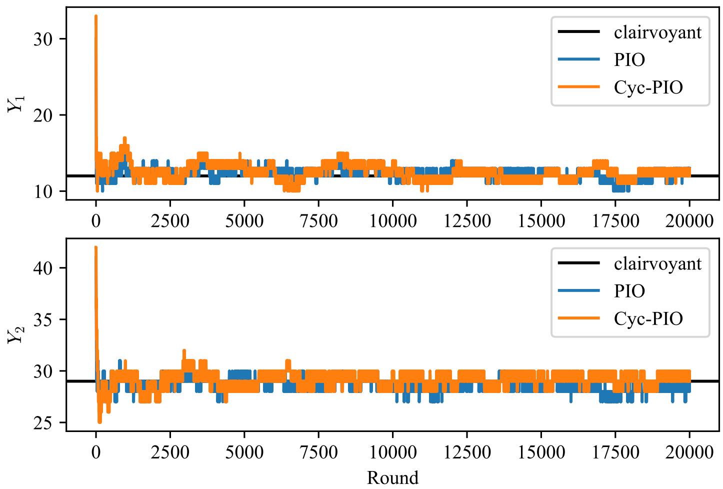
In the two-product case, considering the complexity of fulfillment dynamics, the proof is not a trivial generalization of the single-product case. Interested readers may refer to Appendix D.1.2 for detailed proof.

11.1.2.1 Numerical experiments

In Figure 11.2, we numerically show the comparison between our Cyc-PIO and PIO algorithms. For the PIO algorithm, we assume that there is a clairvoyant who tells the firm about the two indicator functions involved in the algorithm. The demand for type-1, type-2, and type-12 orders follow $U[1, 10)$, $U[1, 15)$, and $U[5, 30)$, respectively. The cost parameters are given by $h_1 = 0.8$, $p_1 = 0.2$, $h_2 = 0.2$, $p_2 = 0.8$, and $s = 0.4$. The coefficient of step size $\gamma = 10$. In the evolution of decision variables, we run for $T = 20,000$ rounds. Note that this is not a fair comparison as the Cyc-PIO algorithm lacks information about the lost-sales indicator functions. We present the result to demonstrate the convergence to the optimal of our Cyc-PIO algorithm at a relatively fast speed.



(a) Evolution of $\vec{Z}$



(b) Evolution of $\vec{Y}$

Figure 11.2: Evolution of decision variables over time in the two-product problem

11.2 Est-PIO Algorithm: $O(\sqrt{T})$ Regret

In this section, we make a mild assumption that the firm only has a single optimal decision. We begin with the single-product problem to obtain some heuristics.

11.2.1 Single-Product Case

To get started, we provide a high-level intuition of our algorithm. Note that in the PIO algorithm with censored demand (i.e., lost sales), we cannot differentiate whether the inventory is just enough or is not sufficient. Hence, the lost-sales indicator function may not be fully revealed by available information. However, if we can obtain an unbiased estimator for the expected value of the lost-sales indicator function, we can still reach the optimal. Unfortunately, due to the correlation among implemented order-up-to levels in different periods, we cannot easily formulate such an unbiased estimator. For example, the following estimator

$$\frac{\sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0, D_t < Y_0\right\}}{\sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0\right\}}$$

is unbiased for $\mathbb{E}[\mathbb{I}\left\{D < Y_0\right\}] = \mathbb{P}(D < Y_0)$ for any given integer Y_0 if $\{Y_k\}_{k=1}^{K}$ is mutually independent and independent of $\{D_k\}_{k=1}^{K}$. However, $\mathbb{I}\left\{Y_k \geq Y_0\right\}$ is $\sigma(D_1, D_2, \cdots, D_{k-1})$-measurable, meaning that $\{Y_k\}_{k=1}^{K}$ is dependent on $\{D_k\}_{k=1}^{K}$. To see why it is biased, let's check a special case when $K = 2$ and

$$D_k = \begin{cases} 0, & \text{with prob. } \frac{1}{2}, \\ 2Y_0, & \text{with prob. } \frac{1}{2}. \end{cases}$$

Suppose $Y_1 = Y_0 + \xi$ where ξ is a small positive integer, and the step size is large such that $Y_2 < Y_0$ if $D_1 = 0$. Then

$$\begin{aligned} \mathbb{E}\left[\frac{\sum_{k=1}^{2} \mathbb{I}\left\{Y_k \geq Y_0, D_t < Y_0\right\}}{\sum_{k=1}^{2} \mathbb{I}\left\{Y_k \geq Y_0\right\}}\right] =& \mathbb{P}(D_1 = 0, D_2 = 0) \cdot \frac{\mathbb{I}\left\{Y_1 \geq Y_0\right\}}{\mathbb{I}\left\{Y_1 \geq Y_0\right\}} \\ &+ \mathbb{P}(D_1 = 0, D_2 = 2Y_0) \cdot \frac{\mathbb{I}\left\{Y_1 \geq Y_0\right\}}{\mathbb{I}\left\{Y_1 \geq Y_0\right\}} \\ &+ \mathbb{P}(D_1 = 2Y_0, D_2 = 0) \cdot \frac{\mathbb{I}\left\{Y_2 \geq Y_0\right\}}{\mathbb{I}\left\{Y_1 \geq Y_0\right\} + \mathbb{I}\left\{Y_2 \geq Y_0\right\}} \\ &+ \mathbb{P}(D_1 = 2Y_0, D_2 = 2Y_0) \cdot 0 \\ =& \frac{5}{8} \neq \frac{1}{2} = \mathbb{E}[D_k < Y_0]. \end{aligned}$$

Nevertheless, from Huh et al. (2011) Lemma 1, we know that

$$\frac{\sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0, D_t < Y_0\right\}}{\sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0\right\}} \xrightarrow{K\to\infty} \mathbb{P}(D < Y_0) \quad a.s.$$

if $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0\right\} = \infty) = 1$. By the strong law of large numbers, this convergence happens almost surely. Further by using Kolmogorov's law of iterated logarithm, we know that

$$\begin{cases} \limsup \frac{S_{n(K)} - n(K)\mathbb{P}(D<Y_0)}{\sqrt{2\sigma^2 n(K) \log\log n(K)}} = 1 & a.s. \\ \liminf \frac{S_{n(K)} - n(K)\mathbb{P}(D<Y_0)}{\sqrt{2\sigma^2 n(K) \log\log n(K)}} = -1 & a.s. \end{cases}$$

if $\mathbb{P}(\lim_{K\to\infty} n(K) = \infty) = 1$, where

$$n(K) = \sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0\right\},$$

$$S_{n(K)} = \sum_{k=1}^{K} \mathbb{I}\left\{Y_k \geq Y_0, D_t < Y_0\right\},$$

and

$$\sigma = \sqrt{\operatorname{var}(\mathbb{I}\left\{D < Y_0\right\})} \leq \frac{1}{2}.$$

This indicates that we can estimate $\mathbb{P}(D < Y_0)$ to sufficient accuracy if there are enough k's such that $Y_k \geq Y_0$.

Furthermore, to ensure that there is ample exploration, we assign $\frac{c}{\sqrt{t}}$ more probability for the decision point in iteration t to be rounded to its ceiling integer.

Next, we propose our learning algorithm.

Est-PIO Algorithm: Single-Product

Step 0 (Initialization) Choose $Z_1 = \beta$, which is the upper bound of the admissible region, and implement $Y_1 = \beta$.

Step 1 Calculate

$$\hat{\theta}_t(Z_t) = \frac{\sum_{k=1}^{t} \mathbb{I}\left\{Y_k \geq \lceil Z_t \rceil, D_k < \lceil Z_t \rceil\right\}}{\sum_{k=1}^{t} \mathbb{I}\left\{Y_k \geq \lceil Z_t \rceil\right\}}$$

Step 2 Calculate $Z_{t+1} = Z_t - \alpha_t H(Z_t)$, where $\alpha_t = \frac{\gamma}{\sqrt{t}}$ and $H(Z_t) = -p + (h+p)\hat{\theta}_t(Z_t)$.

Step 3 Do probabilistic rounding on Z_{t+1} with $\frac{c}{\sqrt{t+1}}$ more probability being rounded to $\lceil Z_{t+1} \rceil$ to obtain a target order-up-to level Y_{t+1}, i.e.,

$$Y_{t+1} = \begin{cases} \lceil Z_{t+1} \rceil, & \text{with prob. } \min\{Z_{t+1} - \lfloor Z_{t+1} \rfloor + \frac{c}{\sqrt{t+1}}, 1\}, \\ \lfloor Z_{t+1} \rfloor, & \text{with prob. } \max\{1 - (Z_{t+1} - \lfloor Z_{t+1} \rfloor + \frac{c}{\sqrt{t+1}}), 0\}. \end{cases}$$

Calculate carry-over inventory $I_t = (Y_t - D_t)^+$. Implement $Y_{t+1} \vee I_t$ and update $t = t + 1$.

Note that we do not need to track all the history demands in each iteration. Maintaining two lists for each integer number m in the admissible region to record $\mathbb{I}\{Y_t \geq m\}$ and $\mathbb{I}\{Y_t \geq m, D_t < m\}$ in iteration t helps to sufficiently reduce the computational time.

Next, we establish the asymptotic optimality result for our Est-PIO algorithm.

Theorem 8 (Regret Upper Bound)**.** *The Est-PIO algorithm can achieve a sublinear regret bound. Specifically,*

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi(Y_t) - \pi(Z^*))\right] \leq C_1^{est}\sqrt{T},$$

where C_1^{est} *is a constant independent of* T*.*

Here, we provide some intuition of the proof. First, by Glivenko-Cantelli theorem or Kolmogorov's law of iterated logarithm, we know that the convergence of $\frac{\sum_{k=1}^{K}\mathbb{I}\{Y_k \geq Y_0, D_t < Y_0\}}{\sum_{k=1}^{K}\mathbb{I}\{Y_k \geq Y_0\}}$ is uniform when $\sum_{k=1}^{K}\mathbb{I}\{Y_k \geq Y_0\}$ diverges almost surely. Then, we check the number of steps it requires to be rounded to a ceiling integer and prove its finite expectation. Therefore, we know that with high probability, a certain range of integers, $\{1, 2, \cdots, Z^*, Z^*+1\}$, will have their estimated left-subderivative updated in the correct direction (i.e., with the correct sign). Moreover, when the step size is small enough, Z will not leave the region bounded by these integers. Therefore, the regret will be bounded according to the step size. Interested readers are recommended to read Appendix D.2.1 for reference.

Note that in the case of backlog, we can still use the Est-PIO algorithm with some modifications. It is worth noting that the demand is not censored anymore, therefore we can construct an unbiased estimator for the left-subderivative at Y_0 as

$$-b + (b+h)\frac{\sum_{k=1}^{t}\mathbb{I}\{D_k < Y_0\}}{t},$$

where $\frac{\sum_{k=1}^{t}\mathbb{I}\{D_k < Y_0\}}{t}$ is the empirical demand distribution. Moreover, there is no need for more probability to be rounded to the ceiling in each iteration. That is, we can simply use the general probabilistic rounding to decide the actual implemented order-up-to level.

11.2.1.1 Numerical Experiments

In Figure 11.3, we numerically show the comparison between our Est-PIO and PIO algorithms in both cases of lost sales and backlog. Particularly, in both the cases of lost sales and backlog, we assume that demand follows a discrete $U[1, 50)$ distribution. The unit inventory holding cost is $h = 0.4$, the unit lost-sales cost $p = 0.6$ or the unit backlog cost $b = 0.6$. The coefficient of step size $\gamma = 10$. The constant $c = 500$ in the lost-sales case. In the evolution of decision variables, we run for $T = 20,000$ rounds. For the regret comparison, we run for $T = 20,000$ rounds over 500 sample paths. It is worth noting that we assume that when PIO algorithm permits known indicator functions, there is $\frac{c}{\sqrt{t}}$ more probability to be rounded to $\lceil Z_t \rceil$ in the lost-sales case to ensure a fair comparison. It is clear that Est-PIO outperforms PIO in both the cases of lost sales and backlog as its convergence is more stable, benefiting from the slow update of the estimated left-subderivative.

11.2.2 Two-Product Case with Fulfillment Dynamics

Similarly, for our two-product problem, we can also construct estimators for the expected value of the two indicator functions. As the denominator of the estimators goes to infinity, the estimator tends to the probability of the two soldout events by the strong law of large numbers (Huh et al. 2011, Lemma 1).

Here, we directly give the Est-PIO for our two-product problem with fulfillment dynamics.

Est-PIO Algorithm: Two-Product

Step 0 (Initialization) Choose $\vec{Z}_1 = \vec{\beta}$, which is the upper bound of the admissible region, and implement $\vec{Y}_1 = \vec{\beta}$.

Step 1 Calculate

$$\hat{\theta}_{t,i}(\vec{Z}_t) = \frac{\sum_{k=1}^{t} \mathbb{I}\left\{\vec{Y}_k \geq \lceil \vec{Z}_t \rceil, \hat{D}_{k,i} \geq \lceil \vec{Z}_t \rceil\right\}}{\sum_{k=1}^{t} \mathbb{I}\left\{\vec{Y}_k \geq \lceil \vec{Z}_t \rceil\right\}}$$

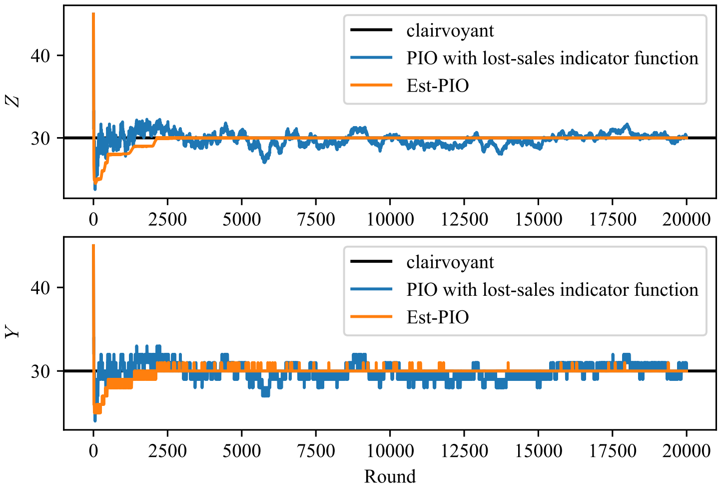
and

$$\hat{\zeta}_{t,i}(\vec{Z}_t) = \frac{\sum_{k=1}^{t} \mathbb{I}\left\{\vec{Y}_k \geq \lceil \vec{Z}_t \rceil, \mathcal{M}_{k,i}(\lceil \vec{Z}_t \rceil) \leq \mathcal{M}_{k,j}(\lceil \vec{Z}_t \rceil), \mathcal{T}_{k,i}(\lceil \vec{Z}_t \rceil) = 12\right\}}{\sum_{k=1}^{t} \mathbb{I}\left\{\vec{Y}_k \geq \lceil \vec{Z}_t \rceil\right\}}$$

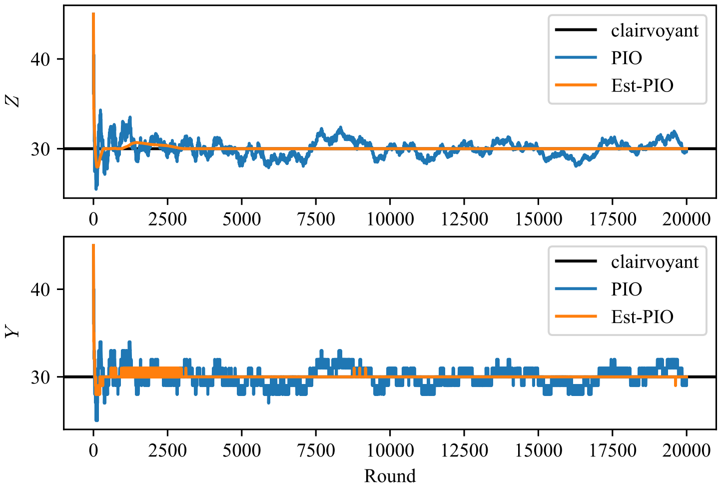
where $\mathcal{M}_{k,i}(\lceil \vec{Z}_t \rceil)$ is the depletion time of product i with initial order-up-to level $\lceil \vec{Z}_t \rceil$ when it is in the kth period, and $\mathcal{T}_{k,i}(\lceil \vec{Z}_t \rceil)$ is the corresponding type of the order in that depletion time.

Step 2 Calculate $\vec{Z}_{t+1} = \vec{Z}_t - \alpha_t \vec{H}(\vec{Z}_t)$, where $\alpha_t = \frac{\gamma}{\sqrt{t}}$ and $H_i(\vec{Z}_t) = h_i - (h_i + p_i - s)\hat{\theta}_{t,i} - s\hat{\zeta}_{t,i}$.

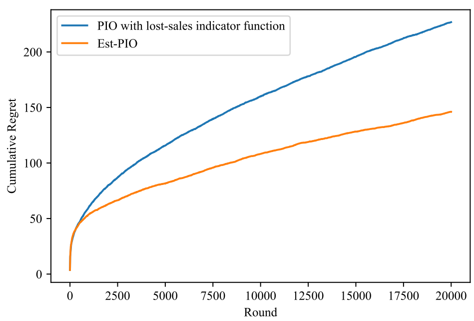
Step 3 Do probabilistic rounding on $\vec{Z}_{t+1}$ with $\frac{c}{\sqrt{t+1}}$ more probability being rounded to



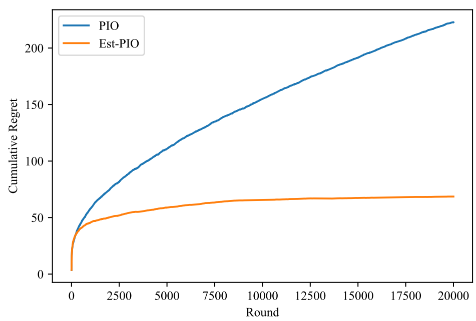
(a) Evolution of decision variables in the case of lost sales



(b) Evolution of decision variables in the case of backlog



(c) Regret in the case of lost sales



(d) Regret in the case of backlog

Figure 11.3: Comparison of Est-PIO and PIO algorithms in single-product problem

$\lceil \vec{Z}_{t+1} \rceil$ to obtain a target order-up-to level $\vec{Y}_{t+1}$, i.e.,

$$\vec{Y}_{t+1} := \begin{cases} (\lfloor Z_{t+1,1} \rfloor, \lfloor Z_{t+1,2} \rfloor) & \text{with prob. } \max\{1 - (Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor + \frac{c}{\sqrt{t+1}}), 0\} \\ & \quad \times \frac{\lceil Z_{t+1,2} \rceil - Z_{t+1,2}}{1-(Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor)}, \\ (\lceil Z_{t+1,1} \rceil, \lceil Z_{t+1,2} \rceil) & \text{with prob. } \min\{Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor + \frac{c}{\sqrt{t+1}}, 1\}, \\ (\lfloor Z_{t+1,1} \rfloor, \lceil Z_{t+1,2} \rceil) & \text{with prob. } \max\{1 - (Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor + \frac{c}{\sqrt{t+1}}), 0\} \\ & \quad \times \frac{1-(\lceil Z_{t+1,2} \rceil - Z_{t+1,2}) - (Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor)}{1-(Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor)}. \end{cases}$$

if $(Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor) \le (Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor)$, and

$$\vec{Y}_{t+1} := \begin{cases} (\lfloor Z_{t+1,1} \rfloor, \lfloor Z_{t+1,2} \rfloor) & \text{with prob. } \max\{1 - (Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor + \frac{c}{\sqrt{t+1}}), 0\} \\ & \quad \times \frac{\lceil Z_{t+1,1} \rceil - Z_{t+1,1}}{1-(Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor)}, \\ (\lceil Z_{t+1,1} \rceil, \lceil Z_{t+1,2} \rceil) & \text{with prob. } \min\{Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor + \frac{c}{\sqrt{t+1}}, 1\}, \\ (\lceil Z_{t+1,1} \rceil, \lfloor Z_{t+1,2} \rfloor) & \text{with prob. } \max\{1 - (Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor + \frac{c}{\sqrt{t+1}}), 0\} \\ & \quad \times \frac{1-(\lceil Z_{t+1,1} \rceil - Z_{t+1,1}) - (Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor)}{1-(Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor)}, \end{cases}$$

if $(Z_{t+1,1} - \lfloor Z_{t+1,1} \rfloor) > (Z_{t+1,2} - \lfloor Z_{t+1,2} \rfloor)$. Calculate carry-over inventory $\vec{I}_t = (\vec{Y}_t - \hat{\vec{D}}_t)^+$. Implement $\vec{Y}_{t+1} \vee \vec{I}_t$ and update $t = t + 1$.

Next, we present the result that our Est-PIO for the two-product problem is efficient and effective. Interested readers can refer to Appendix D.2.2 for proof.

Theorem 9 (Regret Upper Bound)**.** *The Est-PIO algorithm can achieve a sublinear regret bound. Specifically,*

$$\mathbb{E}\left[\sum_{t=1}^{T} (\pi^{\ell}(\vec{Y}_t) - \pi^{\ell}(\vec{Z}^*))\right] \le C_2^{est}\sqrt{T},$$

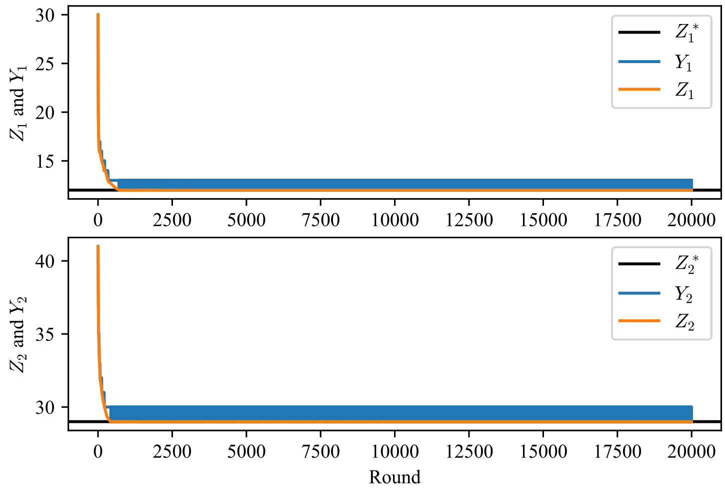
where C_2^{est} is a constant independent of T.

Note that in the case of backlog, we can also use the Est-PIO algorithm. Like in the single-product case, we can obtain an unbiased estimator for the expected left subgradient. Because active exploration is not required in the backlog case, the rounding procedure can be the same as that introduced in Chapter 9.

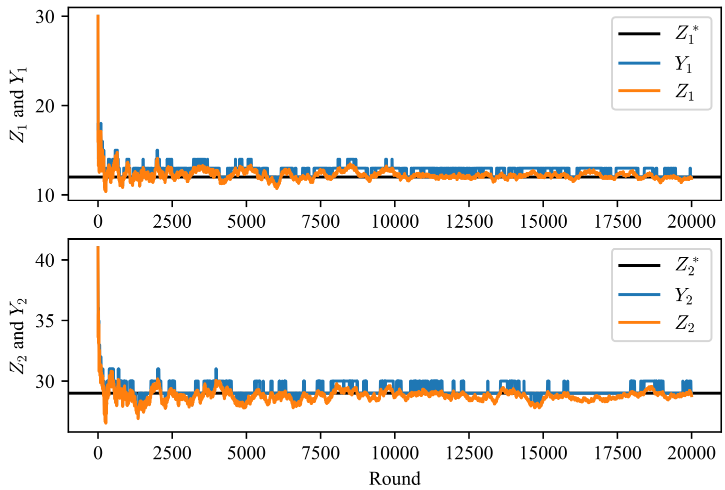
11.2.2.1 Numerical Experiments

The following Figures 11.4 and 11.5 show the comparison of Est-PIO and PIO algorithms in the cases of lost sales and backlog, respectively. The demand for type-1, type-2, and type-12 orders follow $U[1, 10)$, $U[1, 15)$, and $U[5, 30)$, respectively. The cost parameters are given by $h_1 = 0.8$, $b_1 = 0.2$, $h_2 = 0.2$, $b_2 = 0.8$, and $s = 0.4$, or $h_1 = 0.8$, $p_1 = 0.2$, $h_2 = 0.2$, $p_2 = 0.8$, and $s = 0.4$. The coefficient of step size $\gamma = 10$. The constant $c = 500$

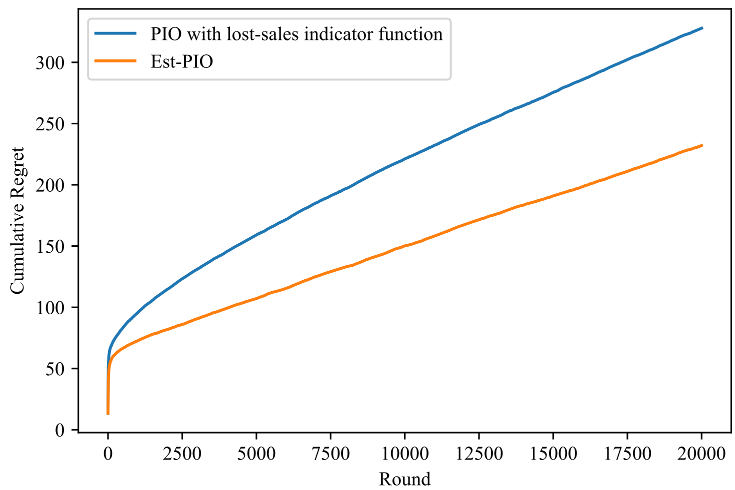
in the lost-sales case. In the evolution of decision variables, we run for $T = 20,000$ rounds. For the regret comparison, we run for $T = 20,000$ rounds over 500 sample paths. To guarantee a fair comparison, we assume that when PIO algorithm permits known indicator functions in the case of lost sales, there is $\frac{c}{\sqrt{t}}$ more probability to be rounded to $\lceil \vec{Z}_t \rceil$ in the tth iteration, as in the design of Est-PIO algorithm.



(a) Est-PIO algorithm

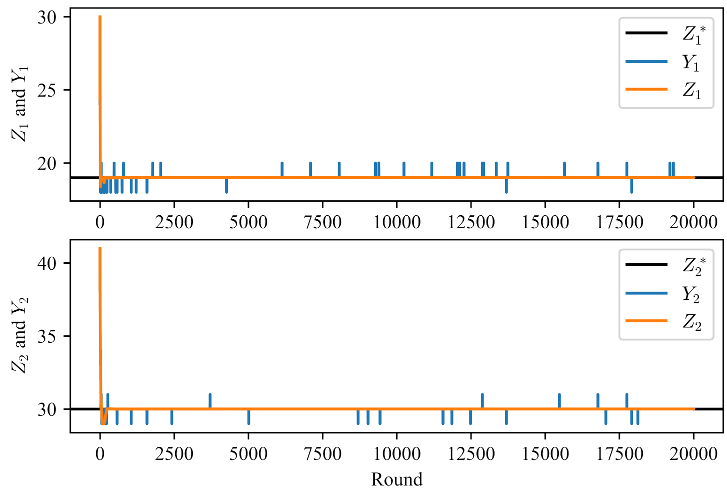


(b) PIO algorithm

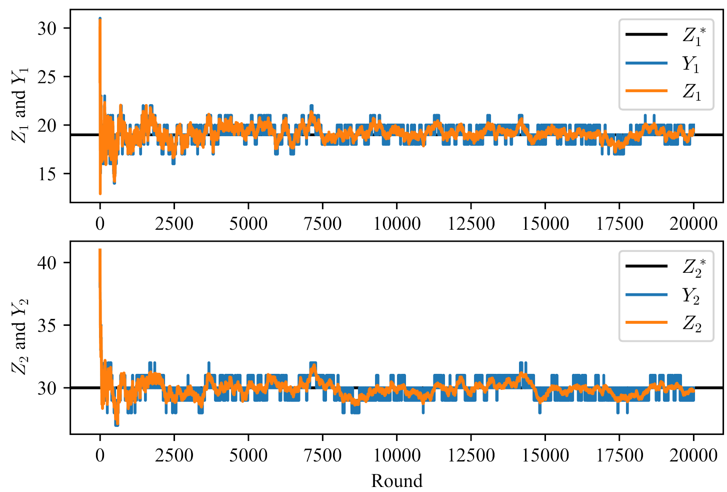


(c) Regret

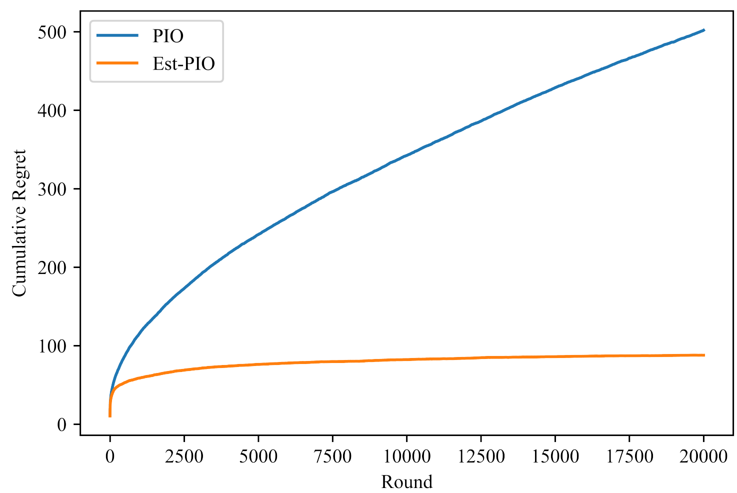
Figure 11.4: Comparison of Est-PIO and PIO algorithms in two-product problem with lost sales



(a) Est-PIO algorithm



(b) PIO algorithm



(c) Regret

Figure 11.5: Comparison of Est-PIO and PIO algorithms in two-product problem with backlog

Chapter 12

Discussions and Conclusions

This paper has investigated the multi-product inventory control problem with demand complementarity, partial fulfillment, and shipping costs in the context of e-commerce, in a zero lead time setting. Our research has made significant contributions to the field by:

- Developing a novel model that captures the demand complementarity and partial fulfillment, while incorporating shipping costs.
- Proving the convex-extensibility and submodularity of the cost function under full information, enabling the development of online convex optimization algorithms.
- Adapting the online subgradient algorithm to our specific setting and proving its asymptotically optimal performance with a tight regret bound $\Omega(\sqrt{T})$, ensuring real-time decision-making and alignment with practical discrete demand scenarios.
- Highlighting the importance of considering shipping costs in inventory control through numerical experiments, demonstrating significant cost savings potential.
- Extending to the scenario without lost-sales indicator functions in the case of lost-sales.

Our research opens up exciting avenues for future exploration. Incorporating positive lead time and fixed setup costs into the model can further enhance its practicality and provide insights into real-world supply chain factors. Furthermore, in the future, we will delve into designing online learning algorithms for other fulfillment dynamics, such as whole-order fulfillment.

By adopting the methodologies and strategies proposed in this paper, e-commerce companies can significantly improve their inventory control practices, reduce costs, and gain a competitive edge in the rapidly evolving e-commerce landscape. We encourage e-commerce companies to leverage our findings to optimize their inventory management processes and achieve greater operational efficiency.

Bibliography

Akbari, M., B. Gharesifard, and T. Linder (2015). Distributed online convex optimization on time-varying directed graphs. *IEEE Transactions on Control of Network Systems 4*(3), 417–428.

Arrow, K. J., T. Harris, and J. Marschak (1951). Optimal inventory policy. *Econometrica: Journal of the Econometric Society*, 250–272.

Bai, T., M. Wu, and S. X. Zhu (2019). Pricing and ordering by a loss averse newsvendor with reference dependence. *Transportation Research Part E: Logistics and Transportation Review 131*, 343–365.

Besbes, O. and A. Muharremoglu (2013). On implications of demand censoring in the newsvendor problem. *Management Science 59*(6), 1407–1424.

Besbes, O. and A. Zeevi (2011). On the minimax complexity of pricing in a changing environment. *Operations research 59*(1), 66–79.

Blackburn, J. D. and R. A. Millen (1984). Simultaneous lot-sizing and capacity planning in multi-stage assembly processes. *European Journal of Operational Research 16*(1), 84–93.

Boyd, S., L. Xiao, and A. Mutapcic (2003). Subgradient methods. *lecture notes of EE392o, Stanford University, Autumn Quarter 2004*, 2004–2005.

Braden, D. J. and M. Freimer (1991). Informational dynamics of censored observations. *Management Science 37*(11), 1390–1404.

Broder, J. and P. Rusmevichientong (2012). Dynamic pricing under a general parametric choice model. *Operations Research 60*(4), 965–980.

Browne, S. and P. Zipkin (1991). Inventory models with continuous, stochastic demands. *The Annals of Applied Probability*, 419–435.

Bu, J., X. Gong, and X. Chao (2023). Asymptotic optimality of base-stock policies for perishable inventory systems. *Management Science 69*(2), 846–864.

Cao, X. and T. Başar (2020). Decentralized online convex optimization with event-triggered communications. *IEEE Transactions on Signal Processing 69*, 284–299.

Chen, X. (2017). $L^\natural$-convexity and its applications in operations. *Frontiers of Engineering Management 4*(3), 283–294.

Chen, X. and M. Li (2021). $M^\natural$-convexity and its applications in operations. *Operations Research 69*(5), 1396–1408.

Dvoretzky, A., J. Kiefer, and J. Wolfowitz (1953). On the optimal character of the (s, S) policy in inventory theory. *Econometrica: Journal of the Econometric Society*, 586–596.

Edgeworth, F. Y. (1888). The mathematical theory of banking. *Journal of the Royal Statistical Society 51*(1), 113–127.

Eisenhut, P. (1975). A dynamic lot sizing algorithm with capacity constraints. *AIIE transactions 7*(2), 170–176.

Erdem, A. S., M. M. Fadiloglu, and S. Özekici (2006). An EOQ model with multiple suppliers and random capacity. *Naval Research Logistics (NRL) 53*(1), 101–114.

Federgruen, A. and Y.-S. Zheng (1992). An efficient algorithm for computing an optimal (r, Q) policy in continuous review stochastic inventory systems. *Operations Research 40*(4), 808–813.

Federgruen, A. and P. Zipkin (1986). An inventory model with limited production capacity and uncertain demands ii. the discounted-cost criterion. *Mathematics of Operations Research 11*(2), 208–215.

Feng, Q., S. P. Sethi, H. Yan, and H. Zhang (2006). Are base-stock policies optimal in inventory problems with multiple delivery modes? *Operations Research 54*(4), 801–807.

Flaxman, A. D., A. T. Kalai, and H. B. McMahan (2004). Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*.

Gallego, G. and I. Moon (1993). The distribution free newsboy problem: review and extensions. *Journal of the Operational Research Society 44*(8), 825–834.

Galliher, H., P. M. Morse, and M. Simond (1959). Dynamics of two classes of continuous-review inventory systems. *Operations Research 7*(3), 362–384.

Gao, X. and H. Zhang (2022a). An efficient learning framework for multiproduct inventory systems with customer choices. *Production and Operations Management 31*(6), 2492–2516.

Gao, X. and H. Zhang (2022b). Inventory control with censored demand. In *The Elements of Joint Learning and Optimization in Operations Management*, pp. 273–303. Springer.

Godfrey, G. A. and 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.

Goldenshluger, A. and A. Zeevi (2009). Woodroofes one-armed bandit problem revisited.

Gordon, G. J. (1999). Regret bounds for prediction problems. In *Proceedings of the twelfth annual conference on Computational learning theory*, pp. 29–40.

Hajek, B. (1985). Extremal splittings of point processes. *Mathematics of operations research 10*(4), 543–556.

Haksever, C. and J. Moussourakis (2005). A model for optimizing multi-product inventory systems with multiple constraints. *International Journal of Production Economics 97*(1), 18–30.

Harris, F. W. (1915). Operations and cost. *Factory management series*, 48–52.

Hausman, W. H., H. L. Lee, and A. X. Zhang (1998). Joint demand fulfillment probability in a multi-item inventory system with independent order-up-to policies. *European Journal of Operational Research 109*(3), 646–659.

Hazan, E. (2022). *Introduction to online convex optimization.* MIT Press.

Hazan, E., A. Agarwal, and S. Kale (2007). Logarithmic regret algorithms for online convex optimization. *Machine Learning 69*(2), 169–192.

Hill, R. and T. Pakkala (2007). Base stock inventory policies for a multi-item demand process. *International Journal of Production Economics 109*(1-2), 137–148.

Huh, W. T., G. Janakiraman, J. A. Muckstadt, and P. Rusmevichientong (2009). 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., R. Levi, P. Rusmevichientong, and J. B. Orlin (2011). Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research 59*(4), 929–941.

Huh, W. T. and P. Rusmevichientong (2009). A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research 34*(1), 103–123.

Karlin, S. and C. R. Carr (1962). Prices and optimal inventory policy. *Studies in Applied Probability and Management Science 4*(1), 159–172.

Kazaz, B. and S. Webster (2015). Price-setting newsvendor problems with uncertain supply and risk aversion. *Operations Research 63*(4), 807–811.

Keskin, N. B. and A. Zeevi (2014). Dynamic pricing with an unknown demand model: Asymptotically optimal semi-myopic policies. *Operations research 62*(5), 1142–1167.

Kouki, C., Z. Jemaï, and S. Minner (2015). A lost sales (r, Q) inventory control model for perishables with fixed lifetime and lead time. *International Journal of Production Economics 168*, 143–157.

Li, J. (1997). An approximation method for the analysis of GI/G/1 queues. *Operations Research 45*(1), 140–144.

Liberatore, M. J. (1979). The EOQ model under stochastic lead time. *Operations research 27*(2), 391–396.

Lu, X., J.-S. Song, and K. Zhu (2008). Analysis of perishable-inventory systems with censored demand data. *Operations Research 56*(4), 1034–1038.

Lugosi, G., M. G. Markakis, and G. Neu (2024). On the hardness of learning from censored and nonstationary demand. *INFORMS Journal on Optimization 6*(2), 63–83.

Lyu, C., H. Zhang, and L. Xin (2024). Ucb-type learning algorithms with kaplan–meier estimator for lost-sales inventory models with lead times. *Operations Research*.

Marshall, K. T. (1968). Some inequalities in queuing. *Operations Research 16*(3), 651–668.

Mills, E. S. (1959). Uncertainty and price theory. *The Quarterly Journal of Economics 73*(1), 116–130.

Moriguchi, S. and K. Murota (2012). On discrete hessian matrix and convex extensibility. *Journal of the Operations Research Society of Japan 55*(1), 48–62.

Murota, K. (1998). Discrete convex analysis. *Mathematical Programming 83*, 313–371.

Murota, K. (2003). *Discrete Convex Analysis*. Society for Industrial and Applied Mathematics.

Murota, K. (2009). Recent developments in discrete convex analysis. *Research Trends in Combinatorial Optimization: Bonn 2008*, 219–260.

Murota, K. (2018). Main features of discrete convex analysis.

Murota, K. (2022). Discrete convex analysis: A tool for economics and game theory. *arXiv preprint arXiv:2212.03598*.

Murota, K. and A. Tamura (2023). Recent progress on integrally convex functions. *Japan Journal of Industrial and Applied Mathematics*, 1–55.

Orabona, F. (2019). A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*.

Orabona, F., L. Jie, and B. Caputo (2012). Multi kernel learning with online-batch optimization. *Journal of Machine Learning Research 13*(2).

Pentico, D. W. and M. J. Drake (2009). The deterministic eoq with partial backordering: a new approach. *European Journal of Operational Research 194*(1), 102–113.

Petruzzi, N. C. and M. Dada (1999). Pricing and the newsvendor problem: A review with extensions. *Operations Research 47*(2), 183–194.

Poormoaied, S. (2022). Inventory decision in a periodic review inventory model with two complementary products. *Annals of Operations Research 315*(2), 1937–1970.

Poormoaied, S. and Z. Atan (2020). A continuous review policy for two complementary products with interrelated demand. *Computers & Industrial Engineering 150*, 106980.

Qin, Y., R. Wang, A. J. Vakharia, Y. Chen, and M. M. Seref (2011). The newsvendor problem: Review and directions for future research. *European Journal of Operational Research 213*(2), 361–374.

Ratliff, N. D., J. A. Bagnell, and M. A. Zinkevich (2007). (approximate) subgradient methods for structured prediction. In *Artificial Intelligence and Statistics*, pp. 380–387. PMLR.

Rice, S. (1962). Single server systemsi. relations between some averages. *Bell System Technical Journal 41*(1), 269–278.

Riordan, J. (1962). Stochastic service systems.

Roundy, R. (1986). A 98%-effective lot-sizing rule for a multi-product, multi-stage production/inventory system. *Mathematics of operations research 11*(4), 699–727.

Scarf, H., K. Arrow, S. Karlin, and P. Suppes (1960). The optimality of (S, s) policies in the dynamic inventory problem. *Optimal pricing, inflation, and the cost of price adjustment*, 49–56.

Scarf, H. E. (1957). *A min-max solution of an inventory problem.* Rand Corporation Santa Monica.

Schulte, B. and A.-L. Sachs (2020). The price-setting newsvendor with poisson demand. *European Journal of Operational Research 283*(1), 125–137.

Shaked, M. and J. Shanthikumar (2007). *Stochastic Orders*. Springer Series in Statistics. Springer New York.

Shalev-Shwartz, S. and Y. Singer (2007). Logarithmic regret algorithms for strongly convex repeated games. *The Hebrew University*.

Shalev-Shwartz, S., Y. Singer, and N. Srebro (2007). Pegasos: Primal estimated sub-gradient solver for svm. In *Proceedings of the 24th international conference on Machine learning*, pp. 807–814.

Shi, C., W. Chen, and I. Duenyas (2016). Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research 64*(2), 362–370.

Simchi-Levi, D., X. Chen, and J. Bramel (2013). *The Logic of Logistics: Theory, Algorithms, and Applications for Logistics Management*. Springer Science & Business Media.

Simchi-Levi, D., X. Chen, J. Bramel, et al. (2005). The logic of logistics. *Theory, algorithms, and applications for logistics and supply chain management*.

Song, J.-S. (1998). On the order fill rate in a multi-item, base-stock inventory system. *Operations research 46*(6), 831–845.

Tang, J., C. Shi, and I. Duenyas (2022). Online learning and matching for multiproduct systems with general upgrading. *Available at SSRN 3978123*.

Tsybakov, A. (2008). *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer New York.

Turken, N., Y. Tan, A. J. Vakharia, L. Wang, R. Wang, and A. Yenipazarli (2012). The multi-product newsvendor problem: Review, extensions, and directions for future research. *Handbook of Newsvendor Problems: Models, Extensions and Applications*, 3–39.

Veinott Jr, A. F. (1965a). Optimal policy for a multi-product, dynamic, nonstationary inventory problem. *Management science 12*(3), 206–222.

Veinott Jr, A. F. (1965b). Optimal policy in a dynamic, single product, nonstationary inventory model with several demand classes. *Operations research 13*(5), 761–778.

Veinott Jr, A. F. and H. M. Wagner (1965). Computing optimal (s, S) inventory policies. *Management Science 11*(5), 525–552.

Whitin, T. M. (1955). Inventory control and price theory. *Management Science 2*(1), 61–68.

Wilson, R. (1934). *A scientific routine for stock control.* Harvard Univ.

Xiao, L. and C. Wang (2023). Multi-location newsvendor problem with random yield: Centralization versus decentralization. *Omega 116*, 102795.

Zhang, H., X. Chao, and C. Shi (2018). Perishable inventory systems: Convexity results for base-stock policies and learning algorithms under censored demand. *Operations Research 66*(5), 1276–1286.

Zheng, Y.-S. and A. Federgruen (1991). Finding optimal (s, S) policies is about as simple as evaluating a single policy. *Operations Research 39*(4), 654–665.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th international conference on machine learning (icml-03)*, pp. 928–936.

Appendix A

Proof for the Convexity of the Expected Cost Function

A.1 The Case of Backlog

Proof. In this section, we prove that the cost is both convex-extensible and submodular for each given triple of joint demand (D_1, D_2, D_{12}). It is well-known that an $L^\natural$-convex discrete function is convex-extensible (Murota, 2009, 2022; Murota and Tamura, 2023). We first prove that the shipping cost or the number of shipments is $L^\natural$-convex for each given triple of joint demand (D_1, D_2, D_{12}) within the region $\{(Y_1, Y_2) : \hat{D}_1 > Y_1, \hat{D}_2 > Y_2\}$, i.e., $\mathrm{NS}(Y_1, Y_2)$ is $L^\natural$-convex. By Theorem 3.5 in Moriguchi and Murota (2012), we know that the number of shipments is $L^\natural$-convex if and only if

$$[\mathrm{NS}(Y_1+1, Y_2+1) - \mathrm{NS}(Y_1, Y_2+1)] - [\mathrm{NS}(Y_1+1, Y_2) - \mathrm{NS}(Y_1, Y_2)] \leq 0 \quad \text{(A.1)}$$

$$[\mathrm{NS}(Y_1+2, Y_2+1) - \mathrm{NS}(Y_1+1, Y_2+1)] - [\mathrm{NS}(Y_1+1, Y_2) - \mathrm{NS}(Y_1, Y_2)] \geq 0 \quad \text{(A.2)}$$

$$[\mathrm{NS}(Y_1+1, Y_2+2) - \mathrm{NS}(Y_1+1, Y_2+1)] - [\mathrm{NS}(Y_1, Y_2+1) - \mathrm{NS}(Y_1, Y_2)] \geq 0 \quad \text{(A.3)}$$

Note that by increasing the number of product 1 by one unit, the shipping cost would only be decreased when the depletion time of product 2 is later than the depletion time of $Y_1 + 1$ product 1, and the order depleting the last product 1 is of type-12 because the type-12 order should be shipped twice originally (without increasing the number of product 1); the shipping cost would only be increased when the depletion time of product 2 is later than the depletion time of $Y_1 + 1$ product 1, and the order depleting the last product 1 is of type-12. This is because the last type-12 order should be shipped twice now while it only needs to be shipped once upon replenishment originally. Hereafter, we obtain that

$$\begin{aligned}\mathrm{NS}(Y_1+1,Y_2)-\mathrm{NS}(Y_1,Y_2) =&\mathbb{E}[\mathbb{I}\{\mathcal{M}_1'>\mathcal{M}_2,\mathcal{T}_1'=12\}-\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}]\\ =&\mathbb{E}[(1-\mathbb{I}\{\mathcal{T}_1'=1\}-\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\})\\ &-\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}]\\ =&\mathbb{E}[1-\mathbb{I}\{\mathcal{T}_1'=1\}-2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}]\end{aligned}$$

Then, we have

$$\begin{aligned}&[\mathrm{NS}(Y_1+1,Y_2+1)-\mathrm{NS}(Y_1,Y_2+1)]-[\mathrm{NS}(Y_1+1,Y_2)-\mathrm{NS}(Y_1,Y_2)]\\ =&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}-\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2',\mathcal{T}_1'=12\}]\\ =&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}-\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}-\mathbb{I}\{\mathcal{M}_2<\mathcal{M}_1'=\mathcal{M}_2',\mathcal{T}_1'=12\}]\\ =&-2\mathbb{P}(\mathcal{M}_1'=\mathcal{M}_2')\\ \le&0\end{aligned}$$

Hence the inequality (A.1) is satisfied.

On the other hand,

$$\begin{aligned}&[\mathrm{NS}(Y_1+2,Y_2)-\mathrm{NS}(Y_1+1,Y_2)]-[\mathrm{NS}(Y_1+1,Y_2)-\mathrm{NS}(Y_1,Y_2)]\\ =&\mathbb{E}[-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1''=12\}-\mathbb{I}\{\mathcal{T}_1''=1\}+2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}+\mathbb{I}\{\mathcal{T}_1'=1\}]\\ =&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2,\mathcal{T}_1'=12\}-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1''=12\}-\mathbb{I}\{\mathcal{T}_1''=1\}+\mathbb{I}\{\mathcal{T}_1'=1\}]\\ =&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=12\}+2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2<\mathcal{M}_1'',\mathcal{T}_1'=12\}\\ &-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1''=12\}-\mathbb{I}\{\mathcal{T}_1''=1\}+\mathbb{I}\{\mathcal{T}_1'=1\}]\\ =&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=12,\mathcal{T}_1''=12\}+2\mathbb{I}\{\mathcal{M}_1''<\mathcal{M}_2,\mathcal{T}_1'=12,\mathcal{T}_1''=1\}\\ &+2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2<\mathcal{M}_1'',\mathcal{T}_1'=12\}-\mathbb{I}\{\mathcal{T}_1''=1\}+\mathbb{I}\{\mathcal{T}_1'=1\}\\ &-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=12,\mathcal{T}_1''=12\}-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=1,\mathcal{T}_1''=12\}]\\ =&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1''<\mathcal{M}_2,\mathcal{T}_1'=12,\mathcal{T}_1''=1\}+2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2<\mathcal{M}_1'',\mathcal{T}_1'=12\}\\ &-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=1,\mathcal{T}_1''=12\}-\mathbb{I}\{\mathcal{T}_1''=1\}+\mathbb{I}\{\mathcal{T}_1'=1\}]\end{aligned}$$

Thereafter,

$$\begin{aligned}&[\mathrm{NS}(Y_1+2,Y_2+1)-\mathrm{NS}(Y_1+1,Y_2+1)]-[\mathrm{NS}(Y_1+1,Y_2)-\mathrm{NS}(Y_1,Y_2)]\\ =&[\mathrm{NS}(Y_1+2,Y_2+1)-\mathrm{NS}(Y_1+1,Y_2+1)]-[\mathrm{NS}(Y_1+2,Y_2)-\mathrm{NS}(Y_1+1,Y_2)]\\ &+[\mathrm{NS}(Y_1+2,Y_2)-\mathrm{NS}(Y_1+1,Y_2)]-[\mathrm{NS}(Y_1+1,Y_2)-\mathrm{NS}(Y_1,Y_2)]\\ =&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1''<\mathcal{M}_2,\mathcal{T}_1'=12,\mathcal{T}_1''=1\}+2\mathbb{I}\{\mathcal{M}_1'\le\mathcal{M}_2<\mathcal{M}_1'',\mathcal{T}_1'=12\}\\ &-2\mathbb{I}\{\mathcal{M}_1''\le\mathcal{M}_2,\mathcal{T}_1'=1,\mathcal{T}_1''=12\}-\mathbb{I}\{\mathcal{M}_1''=\mathcal{M}_2'\}\end{aligned}$$

$$
\begin{aligned}
&- \mathbb{I}\{\mathcal{T}_1'' = 1\} + \mathbb{I}\{\mathcal{T}_1' = 1\}]\\
=&\mathbb{E}[2\mathbb{I}\{\mathcal{M}_1'' < \mathcal{M}_2, \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\} + 2\mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2 < \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&- 2\mathbb{I}\{\mathcal{M}_1'' \leq \mathcal{M}_2, \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\} - 2\mathbb{I}\{\mathcal{M}_1'' = \mathcal{M}_2'\}\\
&- \mathbb{I}\{\mathcal{T}_1' = 1, \mathcal{T}_1'' = 1\} - \mathbb{I}\{\mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\}\\
&+ \mathbb{I}\{\mathcal{T}_1' = 1, \mathcal{T}_1'' = 1\} + \mathbb{I}\{\mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\}]\\
=&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1'' < \mathcal{M}_2, \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\} + 2\mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2 < \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&- 2\mathbb{I}\{\mathcal{M}_1'' \leq \mathcal{M}_2, \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\} - 2\mathbb{I}\{\mathcal{M}_1'' = \mathcal{M}_2'\}\\
&- \mathbb{I}\{\mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\} + \mathbb{I}\{\mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\}]\\
=&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1'' < \mathcal{M}_2, \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\} + \mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2 < \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&- \mathbb{I}\{\mathcal{M}_1'' \leq \mathcal{M}_2, \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\} - \mathbb{I}\{\mathcal{M}_1'' = \mathcal{M}_2'\}]\\
=&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\}\\
&+ \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_2' < \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&+ \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_2' = \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&+ \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_1'' < \mathcal{M}_2', \mathcal{T}_1' = 12\}\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_2' < \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_2' = \mathcal{M}_1'', \mathcal{T}_1' = 12\}\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_1'' < \mathcal{M}_2', \mathcal{T}_1' = 12\}\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' = \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\}\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\}\\
&- \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_1'' = \mathcal{M}_2'\}\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_1'' = \mathcal{M}_2'\}\\
&- \mathbb{I}\{\mathcal{M}_2 < \mathcal{M}_1' < \mathcal{M}_1'' = \mathcal{M}_2'\}]\\
=&2\mathbb{E}[\mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1\} && \equiv ①\\
&+ \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_2' < \mathcal{M}_1'', \mathcal{T}_1' = 12\} && \equiv ②\\
&+ \mathbb{I}\{\mathcal{M}_1' = \mathcal{M}_2 < \mathcal{M}_1'' < \mathcal{M}_2', \mathcal{T}_1' = 12\} && \equiv ③\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_2' < \mathcal{M}_1'', \mathcal{T}_1' = 12\} && \equiv ④\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_2' = \mathcal{M}_1'', \mathcal{T}_1' = 12\} && \equiv ⑤\\
&+ \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_1'' < \mathcal{M}_2', \mathcal{T}_1' = 12\} && \equiv ⑥\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' = \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\} && \equiv ⑦\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12\} && \equiv ⑧\\
&- \mathbb{I}\{\mathcal{M}_1' < \mathcal{M}_2 < \mathcal{M}_1'' = \mathcal{M}_2'\} && \equiv ⑨\\
&- \mathbb{I}\{\mathcal{M}_2 < \mathcal{M}_1' < \mathcal{M}_1'' = \mathcal{M}_2'\}] && \equiv ⑩\\
\geq& 0
\end{aligned}
$$

We first discuss the case where $D_1, D_2, D_{12} \geq 1$. Therefore, $\vec{Y} \in \{(Y_1, Y_2) : Y_1 \geq 2, Y_2 \geq 2\}$. Under this circumstance, see Figure A.1 for an illustration of the last inequality. Here, we provide an illustration on the meanings of arrows. In this figure, red arrows represent "classification", while blue ones stand for "one-to-one correspondence" or "injection". For example, ⑦ is connected with ⑦.1 and ⑦.2 through red arrows, it means that the case represented by ⑦ can be divided into two non-intersecting cases, i.e., cases ⑦.1 and ⑦.2. Note that the one-to-one correspondence, represented by a blue arrow, is established by exchanging the position of two or more orders. For example, the one-to-one correspondence from ⑧ to ① is obtained by exchanging the first type 1 order and type 12 order mentioned explicitly in case ⑧. By noting the "permutation-invariant" property of orders, we can establish that

$$\begin{aligned}&\mathbb{P}(\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 1, \mathcal{T}_1'' = 12)\\ \leq&\mathbb{P}(\mathcal{M}_1' < \mathcal{M}_1'' < \mathcal{M}_2 < \mathcal{M}_2', \mathcal{T}_1' = 12, \mathcal{T}_1'' = 1)\end{aligned}$$

From Figure A.1 we can see that it only remains ⑦.1.2.3 to find a sequence to which ⑦.1.2.3 corresponds. If there is a type-2 order before the first type-1 order mentioned explicitly in the sequence ⑦.1.2.3, i.e., the sequence is $(\cdots-)2-(\cdots-)-1-1-12-12(-\cdots)$, then it corresponds to a subcase of ⑤ that has two successive type-1 orders, i.e., a sequence like $(\cdots-)12-2-12-1-1(-\cdots)$. If there is no type-2 order before the first type-1 order, then by the assumption on demand lower bound, there must exist a type-2 order after the last type-12 order mentioned explicitly in the sequence, i.e., the sequence is $(\cdots-)1-1-12-12(-\cdots)-2(-\cdots)$, then it corresponds to a subcase of ⑤ that has type-12 order after the last type-12 order mentioned explicitly in the sequence ⑤, i.e., a sequence like $(\cdots-)12-2-12(-\cdots)-12(-\cdots)$, by noticing that it is impossible for ⑦.1.2.3 to consists of only type-1 orders at the head of it, that is, there must be at least one type-12 order before the first type-1 order mentioned explicitly in the sequence ⑦.1.2.3.

Under the scenario where $D_1 = 0$ or $D_2 = 0$ or $D_{12} = 0$, we can prove the inequality similarly. When $D_1 = 0$, all orders that consume item 1 are of type-12. Hence

We hence establish the inequality (A.2). Similarly, the inequality (A.3) is also satisfied by using a symmetric argument.

In summary, Q is $L^\natural$-convex, and hence convex-extensible in the region $D_1 + D_{12} > Y_1, D_2 + D_{12} > Y_2$.

It is easy to see that in the region $\{(Y_1, Y_2) : Y_1 \geq D_1 + D_{12}\}$ or $\{(Y_1, Y_2) : Y_2 \geq D_2 + D_{12}\}$, the conditional expected cost function is convex-extensible because they can be extended as a piecewise linear function. To establish the convex-extensibility of conditional expected cost function in the feasible region, it suffices to show that the conditional expected cost function is convex-extensible in the region $\{(Y_1, Y_2) : Y_1 <$

$D_1 + D_{12}, Y_2 < D_2 + D_{12}\}$ and convex-extensibility is preserved at the boundary. Note that when $Y_1 < D_1 + D_{12}, Y_2 < D_2 + D_{12}$, we have

$$\begin{aligned} &\pi^{sb}(Y_1+1, Y_2|D_1, D_2, D_{12}) - \pi^{sb}(Y_1, Y_2|D_1, D_2, D_{12}) \\ =&s \cdot \mathbb{E}[\mathbb{I}\{\mathcal{M}_1' > \mathcal{M}_2, \mathcal{T}_1' = 12\} - \mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2, \mathcal{T}_1' = 12\}] - b_1 \\ \leq&s - b_1 \end{aligned}$$

and as Y_1 grows larger (more than $D_1 + D_{12}$) with Y_2 fixed, a unit increment in Y_1 would lead to an increment of conditional expected cost by h_1. Therefore, with the assumption of $s < h_1 + b_1$ mentioned in Assumption 1, the convex-extensibility is preserved in Y_1 direction. Similarly, for fixed Y_1, we have $s < h_2 + b_2$ leads to the convex-extensibility in Y_2 direction even when Y_2 becomes larger than $D_2 + D_{12}$. Therefore, the convex-extensibility of cthe onditional expected cost function, and hence the expected cost function, in the feasible region is proved.

From the proof for submodularity of the cost function, we know that $\forall\{(Y_1, Y_2), (Y_1 + 1, Y_2), (Y_1, Y_2 + 1), (Y_1 + 1, Y_2 + 1)\} \subset U$,

$$\pi^{sb}(Y_1, Y_2) + \pi^{sb}(Y_1 + 1, Y_2 + 1) \leq \pi^{sb}(Y_1 + 1, Y_2) + \pi^{sb}(Y_1, Y_2 + 1).$$

We call this property *local submodularity*. This result follows from the $L^\natural$-convexity at each of the areas being separated (linear function is also $L^\natural$-convex function). Since $L^\natural$-convex function is also a submodular function, we reach the desired result. Note that local submodularity implies global submodularity, we hence conclude our proof. □

A.2 The Case of Lost Sales

Proof. Note that

$$\mathrm{NS}(Y_1 + 1, Y_2) - \mathrm{NS}(Y_1, Y_2) = \mathbb{E}[1 - \mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2, \mathcal{T}_1' = 12\}]$$

The proof in this section is even simpler than that in the case of backlog by noting that $\mathbb{P}(\mathcal{T}_1' = 1) = \mathbb{P}(\mathcal{T}_1'' = 1)$ in the case of backlog.

Furthermore, notice that

$$\begin{aligned} &\pi^{s\ell}(Y_1+1, Y_2|D_1, D_2, D_{12}) - \pi^{s\ell}(Y_1, Y_2|D_1, D_2, D_{12}) \\ =&s \cdot \mathbb{E}[1 - \mathbb{I}\{\mathcal{M}_1' \leq \mathcal{M}_2, \mathcal{T}_1' = 12\}] - p_1 \\ \leq&s - p_1. \end{aligned}$$

By using Assumption 1 and a similar argument as in the case of backlog, we can derive the global convex-extensibility. Global submodularity is a natural result of the local submodularity. □

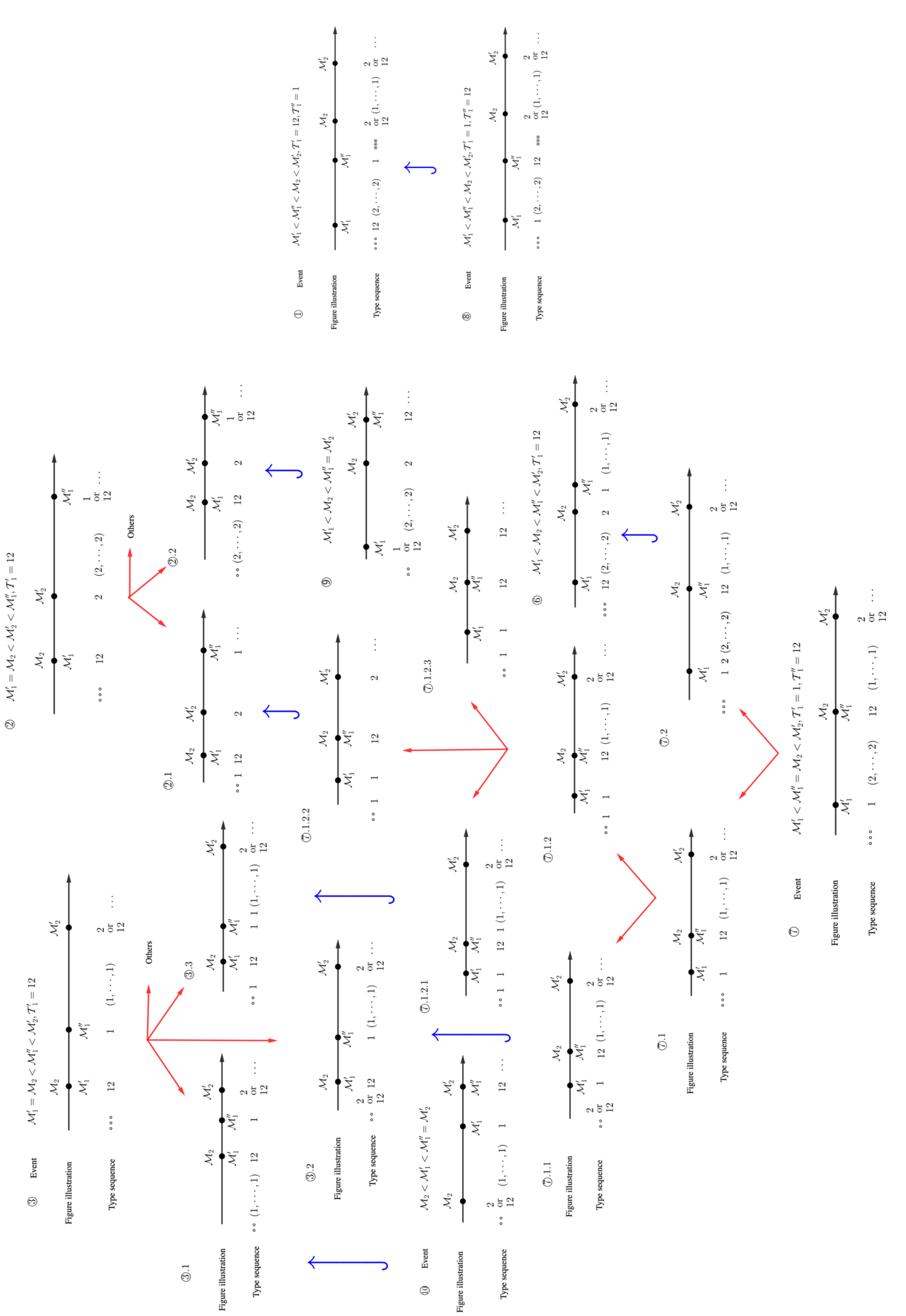


Figure A.1: Sequence correspondence

Appendix B

Proofs for Regret Upper Bound

B.1 Proof for Lemma 2

Proof. By using the iterating process, we have

$$\mathbb{E}||\vec{Z}_{t+1} - \vec{Z}^*||^2 = \mathbb{E}||\mathbf{P}_U(\vec{Z}_t - \alpha_t \vec{H}(\vec{Z}_t) - \vec{Z}^*)||^2 \leq \mathbb{E}||\vec{Z}_t - \alpha_t \vec{H}(\vec{Z}_t) - \vec{Z}^*||^2$$
$$= \mathbb{E}||(\vec{Z}_t - \vec{Z}^*) - \alpha_t \vec{H}(\vec{Z}_t)||^2 = \mathbb{E}||\vec{Z}_t - \vec{Z}^*||^2 + \alpha_t^2 \mathbb{E}||\vec{H}(\vec{Z}_t)||^2 - 2\alpha_t \mathbb{E}\left[\left\langle H(\vec{Z}_t), \vec{Z}_t - \vec{Z}^*\right\rangle\right]$$

Therefore, we obtain that

$$\begin{aligned}\mathbb{E}[\pi^b(\vec{Z}_t) - \pi^b(\vec{Z}^*)] \leq & \mathbb{E}\left[\left\langle \vec{H}(\vec{Z}_t), \vec{Z}_t - \vec{Z}^*\right\rangle\right] \\ \leq & \frac{\mathbb{E}||\vec{Z}_t - \vec{Z}^*||^2}{2\alpha_t} - \frac{\mathbb{E}||\vec{Z}_{t+1} - \vec{Z}^*||^2}{2\alpha_t} + \frac{\alpha_t}{2}\mathbb{E}||\vec{H}(\vec{Z}_t)||^2\end{aligned}$$

where in the second line, the expectation is taken over probabilistic rounding and sample paths. Note that the first inequality follows from the shape (convexity) of the cost function. Summing over t on both sides of the above inequality, we can get

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T} \pi^b(\vec{Z}_t)\right] - T\pi^b(\vec{Z}^*) \\ \leq & \sum_{t=1}^{T}\left[\frac{\mathbb{E}||\vec{Z}_t - \vec{Z}^*||^2}{2\alpha_t} - \frac{\mathbb{E}||\vec{Z}_{t+1} - \vec{Z}^*||^2}{2\alpha_t} + \frac{\alpha_t}{2}\mathbb{E}||\vec{H}(\vec{Z}_t)||^2\right] \\ \leq & \frac{\mathbb{E}||\vec{Z}_1 - \vec{Z}^*||^2}{2\alpha_1} + \frac{1}{2}\sum_{t=1}^{T}\left[\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t}\right]\mathbb{E}||\vec{Z}_{t+1} - \vec{Z}^*||^2 + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\alpha_t \\ \leq & \frac{diam(U)^2}{2}\left\{\frac{1}{\alpha_1} + \sum_{t=1}^{T}\left[\frac{1}{\alpha_{t+1}} - \frac{1}{\alpha_t}\right]\right\} + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\alpha_t \\ = & \frac{diam(U)^2}{2\alpha_{T+1}} + \frac{\bar{B}^2}{2}\sum_{t=1}^{T}\alpha_t\end{aligned}$$

where $\bar{B} := \max\{b_1 + s, b_2 + s, h_1, h_2\}$,

$$\frac{diam(U)^2}{2\alpha_{T+1}} = \frac{diam(U)^2}{2\gamma}\sqrt{T+1} \leq \frac{diam(U)^2}{\gamma}\sqrt{T}$$
$$\frac{\bar{B}^2}{2}\sum_{t=1}^{T}\alpha_t = \frac{\bar{B}^2\gamma}{2}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \frac{\bar{B}^2\gamma}{2}\int_0^T r^{-1/2}dr = \bar{B}^2\gamma\sqrt{T}$$

Take $\gamma = \frac{diam(U)}{\bar{B}}\gamma_0$, then

$$\mathbb{E}\left[\sum_{t=1}^{T}(\pi^b(\vec{Z}_t) - \pi^b(\vec{Z}^*))\right] \leq \left(\gamma_0 + \frac{1}{\gamma_0}\right) diam(U)\bar{B}\sqrt{T}$$

□

B.2 Proof for Lemma 3

We begin by establishing the relationship between the difference in cost functions and the difference in the two ordre-up-to levels involved in the inventory control problem.

Lemma 4. *The difference in the cost function*

$$\mathbb{E}[\pi^b(\vec{Y}_t) - \pi^b(\vec{Y}_t^\rho)] \leq (\max\{h_1, h_2, b_1, b_2\} + s)\mathbb{E}[\sum_{i=1}^{2}|Y_{t,i} - Y_{t,i}^\rho|]$$

A proof can be found in Appendix B.3.

Subsequently, we will investigate the steps involved in narrowing down the disparity between the actual implemented order-up-to level and the target order-up-to level.

Consider the stochastic process $(M_{t,i}|t \geq 1)$ defined by: $M_{1,i} = 0$, and

$$M_{t+1,i} = \left[M_{t,i} + \frac{S_{t,i}}{\sqrt{t}} - \check{D}_{t,i}\right]^+$$

where $S_{t,i} = \gamma h_i$, and $\check{D}_{t,i} = \tilde{D}_{t,i} - 1$ with $\tilde{D}_{t,i}$ being a nonnegative random variable satisfying $\tilde{D}_{t,i} \leq_{st} \hat{D}_i, i = 1, 2$. By constructing this stochastic process, we can confine the discrepancy between $Y_{t,i}$ and $Y_{t,i}^\rho$ in the following manner.

Lemma 5. *The total expected distance function*

$$\mathbb{E}\left[\sum_{t=1}^{T}|Y_{t,i} - Y_{t,i}^\rho|\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}M_{t,i}\right]$$

A proof can be found in Appendix B.4.

Following the approach highlighted in Shi et al. (2016), we establish a connection between the stochastic process $(M_{t,i}|t \geq 1)$ and a $GI/G/1$ queue denoted as $(W_{t,i}|t \geq 1)$. The $GI/G/1$ queue is defined based on Lindley's equation: $W_{1,i} = 0$, and

$$W_{t+1,i} = [W_{t,i} + S_{t,i} - \check{D}_{t,i}]^+$$

where the sequences $S_{t,i}$ and $\check{D}_{t,i}$ represent independent and identically distributed random variables. Specifically, $S_{t,i}$ denotes the service time, while $\check{D}_{t,i}$ represents the inter-arrival time between the tth and $t+1$th customers in the $GI/G/1$ queue. Additionally, the random variable $W_{t,i}$ represents the waiting time of the tth customer in the queue. In their study, Shi et al. (2016) establish a connection between the stochastic process $(M_{t,i}|t \geq 1)$ and the $GI/G/1$ queue $(W_{t,i}|t \geq 1)$. However, before presenting their findings, it is important to establish a foundation by defining the busy period and server utilization, as well as discussing their properties.

Definition 1. *Let $\tau_{1,i} := 1$, and for $t \geq 1$, $\tau_{t+1,i} := \inf\{t > \tau_{t,i} : W_{t,i} = 0\}$. Let $B_{t,i} := \tau_{t+1,i} - \tau_{t,i}$. Then $B_{t,i}$ is the duration of the t^{th} busy period. Define $\upsilon_i := \mathbb{E}[S_{1,i}]/\mathbb{E}[\check{D}_{1,i}]$ as the server utilization.*

The following result is a well-established finding associated with the aforementioned definition.

Lemma 6. *In a $GI/G/1$ queue, when $\upsilon_i \leq 1$, the queue remains stable and the random variable $B_{k,i}$ is independent and identically distributed.*

It is important to highlight that the stability condition $\upsilon_i \leq 1$ can always be satisfied by appropriately scaling the units of the cost parameters. Let B_i be a random variable with the same distribution as $B_{t,i}$. Now, we can present the result proposed by Shi et al. (2016), which can be adapted to our current analysis with minor modifications. This result is built upon the connection between the two processes $(M_{t,i}|t \geq 1)$ and $(W_{t,i}|t \geq 1)$.

Lemma 7. *The total expected function*

$$\mathbb{E}\left[\sum_{t=1}^{T} M_{t,i}\right] \leq 2\mathbb{E}[B_i]S_i\sqrt{T}$$

The proof for this lemma can be found in Appendix B.5.

In what follows, a result characterizing the mean duration of the busy period under a relaxed assumption is established.

However, our research has yielded a noteworthy advancement as we have discovered a result that exclusively relies on the expectation of the demand. This represents a substantial improvement in our findings. A renowned result, originally presented by

Rice (1962) and Riordan (1962) and subsequently reiterated in Marshall (1968) Eq. (3), demonstrates that

$$a_{0,i}\mathbb{E}[I_i] = \mathbb{E}[\check{D}_{1,i}] - \mathbb{E}[S_{1,i}] = \mathbb{E}[\check{D}_{1,i}] - S_i$$

where $a_{0,i} = \mathbb{P}(\text{Arrival finds the system empty})$, I_i is the idle period. As is shown in Li (1997) Eq. (6), the expected busy period can be obtained by the identity

$$\mathbb{E}[B_i] = \frac{\upsilon_i}{1-\upsilon_i}\mathbb{E}[I_i]$$

when $\upsilon_i < 1$. Hence we obtain the following conclusion.

Lemma 8. *If $\upsilon_i < 1, E[\check{D}_{1,i}] < \infty$, then the expected busy period in a GI/G/1 queue is finite. Specifically,*

$$\mathbb{E}[B_i] = \frac{\upsilon_i}{1-\upsilon_i}\frac{\mathbb{E}[\check{D}_{1,i}] - S_i}{a_{0,i}}$$

Now we can establish the proof for Lemma 3, and it turns out that the constant $C_{1,i}$ only relies on h_1, h_2, s, γ, and the demand distribution.

Proof. Let $C_1 = 2(\max\{h_1, h_2, b_1, b_2\} + s)\sum_{i\in\{1,2\}}\mathbb{E}[B_i]S_i$. From Lemma 4, it suffices to show that

$$\mathbb{E}\left[\sum_{t=1}^{T}|\vec{Y}_{t,i} - \vec{Y}^{\rho}_{t,i}|\right] \leq 2\mathbb{E}[B_i]S_i\sqrt{T}$$

Lemma 5 and Lemma 7 naturally lead to this result. □

B.3 Proof for Lemma 4

Proof. Because $\vec{Y}_t = \vec{Y}^{\rho}_t \vee \vec{I}_t$, we have $Y_{t,i} \geq Y^{\rho}_{t,i}, i = 1, 2$.

By the convexity of the expected cost function, we know that

$$\begin{aligned}|\pi^b(\vec{Y}_t) - \pi^b(\vec{Y}^{\rho}_t)| &\leq \max\{|\nabla\pi^b(\vec{Y}_t)|, |\nabla\pi^b(\vec{Y}^{\rho}_t)|\}\sum_{i=1}^{2}|Y_{t,i} - Y^{\rho}_{t,i}| \\ &\leq (\max\{h_1, h_2, b_1, b_2\} + s)\sum_{i=1}^{2}|Y_{t,i} - Y^{\rho}_{t,i}|\end{aligned}$$

□

B.4 Proof for Lemma 5

Proof. We first prove that $|\vec{Y}_{t+1,i} - \vec{Y}^{\rho}_{t+1,i}| \leq_{st} \Big(|\vec{Y}_{t,i} - \vec{Y}^{\rho}_{t,i}| + \alpha_t h_i - (\hat{D}_i - 1)\Big)^+$.

If the carry-over inventory is no more than the target order-up-to level, i.e., $\vec{I}_{t+1} \leq \vec{Y}^{\rho}_{t+1}$, then $\vec{Y}_{t+1} = \max\{\vec{Y}^{\rho}_{t+1}, \vec{I}_{t+1}\} = \vec{Y}^{\rho}_{t+1}$, the above claim holds naturally.

Next, we consider the case when the carry-over inventory is greater than the target order-up-to level, i.e., $I_{t+1,i} > Y^{\rho}_{t+1,i}$ for some i, in which case $Y_{t+1,i} = I_{t+1,i} = Y_{t,i} - \hat{D}_{t,i}$ holds for i. Therefore,

$$\begin{aligned} & Y_{t+1,i} - Y^{\rho}_{t+1,i} \\ =& Y_{t+1,i} - \rho(\mathbf{P}_U(\vec{Z}_t - \alpha_t \vec{H}_t(\vec{Z}_t)))_i \\ =& Y_{t+1,i} - \rho(\vec{Z}_t - \alpha_t \vec{H}_t(\vec{Z}_t))_i. \end{aligned}$$

Since $Y_{t+1,i} = Y_{t,i} - \hat{D}_{t,i}$, it follows that

$$\begin{aligned} & Y_{t+1,i} - Y^{\rho}_{t+1,i} \\ \leq\ & Y_{t,i} - \rho(\vec{Z}_t - \alpha_t \vec{H}_t(\vec{Z}_t))_i - \hat{D}_{t,i} \\ \leq_{st}& Y_{t,i} - \rho(\vec{Z}_t - \alpha_t \vec{h})_i - \hat{D}_{t,i} \\ \leq_{st}& Y_{t,i} - \rho(Z_t)_i + \alpha_t h_i + 1 - \hat{D}_{t,i} \\ \overset{d}{=}\ & Y_{t,i} - Y^{\rho}_{t,i} + \alpha_t h_i + 1 - \hat{D}_{t,i} \end{aligned}$$

Note that $Y_{1,i} = Y^{\rho}_{1,i}$, by the definition of M_t, we have

$$|Y_{t,i} - Y^{\rho}_{t,i}| \leq_{st} M_t$$

which implies that

$$|Y_{t,i} - Y^{\rho}_{t,i}| \leq_{cv} M_t$$

By the property of convex order given in Shaked and Shanthikumar (2007), the lemma is proved.

□

B.5 Proof for Lemma 7

Proof. Let the random variable $L_i(t)$ be the index k in which $B'_{k,i} := \{\tau_{k-1,i}+1, \cdots, \tau_{k,i}\}$ contains t. It can be seen from the definitions of $M_{t,i}$ and $B'_{k,i}$, because the service rate of $(M_{t,i}|t \geq 1)$ is no less than that of $(W_{t,i}|t \geq 1)$, which means that $W_{t,i} = 0$ indicates $M_{t,i} = 0$, we have

$$M_{t,i} \leq \sum_{n=1}^{t} \frac{S_{n,i}}{\sqrt{n}} \mathbb{I}\{n \in B'_{L_i(t),i}\} \quad a.s.$$

Therefore

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T} M_{t,i}\right] \leq& \mathbb{E}\left[\sum_{t=1}^{T}\sum_{n=1}^{t}\frac{S_{n,i}}{\sqrt{n}}\mathbb{I}\{n \in B'_{L_i(t),i}\}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\frac{S_{t,i}}{\sqrt{t}}\sum_{n=1}^{T}\mathbb{I}\{n \in B'_{L_i(t),i}\}\right] \\ \leq& \mathbb{E}\left[\sum_{t=1}^{T}\frac{S_{t,i}}{\sqrt{t}}|B'_{L_i(t),i}|\right] = \sum_{t=1}^{T}\frac{1}{\sqrt{t}}\mathbb{E}[B_{1,i}]S_i \leq 2\mathbb{E}[B_i]S_i\sqrt{T}\end{aligned}$$

where the last inequality comes from integration, i.e., $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq \int_0^T \frac{1}{\sqrt{r}}dr = \sqrt{T}$. □

Appendix C

Proofs for the Regret Lower Bound

C.1 The Case of Backlog

C.1.1 Construction

Let $x_0 := \min\{\frac{b_2+\frac{1}{4}s}{b_2+h_2+\frac{1}{4}s}, \frac{2b_1+3b_2}{b_2+h_2}\}$. When T is big enough, specifically $T \geq \max\{\frac{1}{x_0^2}, \frac{1}{(1-x_0)^2}\}$, suppose that the demand follows one of the following two distributions: F_a: $D_1 = 1, D_{12} = 1$,

$$D_2 = \begin{cases} 1 & \text{w.p. } x_0 - \frac{1}{\sqrt{T}} \\ 2 & \text{w.p. } 1 - x_0 + \frac{1}{\sqrt{T}} \end{cases}$$

one can easily check that the optimal inventory decision (Y_1^{*a}, Y_2^{*a}) under F_a is $(2, 3)$. And F_b: $D_1 = 1, D_{12} = 1$,

$$D_2 = \begin{cases} 1 & \text{w.p. } x_0 + \frac{1}{\sqrt{T}} \\ 2 & \text{w.p. } 1 - x_0 - \frac{1}{\sqrt{T}} \end{cases}$$

whose optimal inventory decision (Y_1^{*b}, Y_2^{*b}) is either $(2, 2)$ or $(0, 0)$, depending on the parameters s, b_1, b_2, h_1, h_2, and T.

To see this, let us use $\{1, 2_1, 2_2, 12\}$ to denote the four orders placed by customers when $D_2 = 2$, hence the $4! = 24$ permutations are given by

$$\begin{aligned}
&1-2_1-2_2-12, \quad 2_1-1-2_2-12, \quad 2_2-1-2_1-12, \quad 12-2_1-2_2-1,\\
&1-12-2_1-2_2, \quad 2_1-12-1-2_2, \quad 2_2-12-1-2_1, \quad 12-1-2_1-2_2,\\
&1-2_2-12-2_1, \quad 2_1-2_2-12-1, \quad 2_2-2_1-12-1, \quad 12-2_2-1-2_1,\\
&1-2_2-2_1-12, \quad 2_1-2_2-1-12, \quad 2_2-2_1-1-12, \quad 12-2_2-2_1-1,\\
&1-12-2_2-2_1, \quad 2_1-12-2_2-1, \quad 2_2-12-2_1-1, \quad 12-1-2_2-2_1,\\
&1-2_1-12-2_2, \quad 2_1-1-12-2_2, \quad 2_2-1-12-2_1, \quad 12-2_1-1-2_2.
\end{aligned}$$

Moreover, let us use $\{1, 2, 12\}$ to denote the three orders places by customers when

$D_2 = 1$, then the $3! = 6$ permutations are given by

$$1-2-12, \quad 1-12-2, \quad 2-1-12, \quad 2-12-1, \quad 12-1-2, \quad 12-2-1$$

We have

Decision \ Demand	$\{1, 2_1, 2_2, 12\}$ w.p. $1 - x_0 + \frac{1}{\sqrt{T}}$			$\{1, 2, 12\}$ w.p. $x_0 - \frac{1}{\sqrt{T}}$		
(Y_1, Y_2)	holding cost	backorder cost	shipping cost	holding cost	backorder cost	shipping cost
$(0,0)$	0	$2b_1 + 3b_2$	$4s$	0	$2b_1 + 2b_2$	$3s$
$(0,1)$	0	$2b_1 + 2b_2$	$\frac{13}{3}s$	0	$2b_1 + b_2$	$\frac{7}{2}s$
$(0,2)$	0	$2b_1 + b_2$	$\frac{37}{8}s$	0	$2b_1$	$4s$
$(0,3)$	0	$2b_1$	$5s$	h_2	$2b_1$	$4s$
$(1,0)$	0	$b_1 + 3b_2$	$\frac{53}{12}s$	0	$b_1 + 2b_2$	$\frac{7}{2}s$
$(1,1)$	0	$b_1 + 2b_2$	$\frac{17}{4}s$	0	$b_1 + b_2$	$\frac{10}{3}s$
$(1,2)$	0	$b_1 + b_2$	$\frac{17}{4}s$	0	b_1	$\frac{7}{2}s$
$(1,3)$	0	b_1	$\frac{9}{2}s$	h_2	b_1	$\frac{7}{2}s$
$(2,0)$	0	$3b_2$	$5s$	0	$2b_2$	$4s$
$(2,1)$	0	$2b_2$	$\frac{14}{3}s$	0	b_2	$\frac{7}{2}s$
$(2,2)$	0	b_2	$\frac{17}{4}s$	0	0	$3s$
$(2,3)$	0	0	$4s$	h_2	0	$3s$

For simplicity, let $y_0 = x_0 - \frac{1}{\sqrt{T}}$. Then

$$\pi^{sb}(0,0) = 2b_1 + (3 - y_0)b_2 + (4 - y_0)s, \qquad \pi^{sb}(0,1) = 2b_1 + (2 - y_0)b_2 + (\frac{13}{3} - \frac{5}{6}y_0)s,$$
$$\pi^{sb}(0,2) = 2b_1 + (1 - y_0)b_2 + (\frac{37}{8} - \frac{5}{8}y_0)s, \quad \pi^{sb}(0,3) = 2b_1 + y_0 h_2 + (5 - y_0)s,$$
$$\pi^{sb}(1,0) = b_1 + (3 - y_0)b_2 + (\frac{53}{12} - \frac{11}{12}y_0)s, \quad \pi^{sb}(1,1) = b_1 + (2 - y_0)b_2 + (\frac{17}{4} - \frac{11}{12}y_0)s,$$
$$\pi^{sb}(1,2) = b_1 + (1 - y_0)b_2 + (\frac{17}{4} - \frac{3}{4}y_0)s, \quad \pi^{sb}(1,3) = b_1 + y_0 h_2 + (\frac{9}{2} - y_0)s,$$
$$\pi^{sb}(2,0) = (3 - y_0)b_2 + (5 - y_0)s, \qquad \pi^{sb}(2,1) = (2 - y_0)b_2 + (\frac{14}{3} - \frac{7}{6}y_0)s,$$
$$\pi^{sb}(2,2) = (1 - y_0)b_2 + (\frac{17}{4} - \frac{5}{4}y_0)s, \qquad \pi^{sb}(2,3) = y_0 h_2 + (4 - y_0)s.$$

It is obvious that

$$\pi^{sb}(0,1) > \pi^{sb}(1,1) > \pi^{sb}(2,2), \qquad \pi^{sb}(1,0) > \pi^{sb}(1,1) > \pi^{sb}(2,2),$$
$$\pi^{sb}(0,2) > \pi^{sb}(1,2) > \pi^{sb}(2,2), \qquad \pi^{sb}(0,3) > \pi^{sb}(1,3) > \pi^{sb}(2,3),$$
$$\pi^{sb}(2,0) > \pi^{sb}(2,1) > \pi^{sb}(2,2).$$

So, the possible optimal value could only be one of $\{\pi^{sb}(0,0), \pi^{sb}(2,2), \pi^{sb}(2,3)\}$. By using simple calculations, we can obtain the optimal decision as aforementioned.

C.1.2 Proof for Theorem 3 Under Our Construction

Proof. Denote the optimal ordering policy when the distribution follows F_ℓ be $(Y_1^{*\ell}, Y_2^{*\ell})$, where $\ell = a, b$. Note that $(Y_1^{*a}, Y_2^{*a}) \neq (Y_1^{*b}, Y_2^{*b})$. Let ϕ be an arbitrary policy. Denote $\mathcal{H}_t$ as the demand and decisions up to period t, i.e., $\mathcal{H}_t = \{(\vec{D}_s, \vec{Y}_s) : 1 \leq s \leq t\}$ for $t \geq 1$ and $\mathcal{H}_0 = \emptyset$. We further let $\psi_t(\mathcal{H}_{t-1})$ be the decision in the tth period under any policy ϕ. The worst-case expected regret of ϕ is bounded from below as follows

$$
\begin{aligned}
&\sup_{F\in\mathcal{F}}\{\pi^\phi(F,T) - \pi^*(F,T)\}\\
\geq &\sup_{F\in\{F_a,F_b\}}\{\pi^\phi(F,T) - \pi^*(F,T)\}\\
= &\max\left\{\sum_{t=1}^{T}\mathbb{E}_{F_a}^{\phi}[\pi(F_a, \psi_t(\mathcal{H}_{t-1})) - \pi(F_a, (Y_1^{*a}, Y_2^{*a}))],\right.\\
&\qquad\qquad\left.\sum_{t=1}^{T}\mathbb{E}_{F_b}^{\phi}[\pi(F_b, \psi_t(\mathcal{H}_{t-1})) - \pi(F_b, (Y_1^{*b}, Y_2^{*b}))]\right\}\\
\geq &\frac{b_2+h_2}{\sqrt{T}}\max\left\{\sum_{t=1}^{T}\mathbb{P}_a^{\phi}((Y_1,Y_2)\in A), \sum_{t=1}^{T}\mathbb{P}_b^{\phi}((Y_1,Y_2)\ \in B)\right\}
\end{aligned}
$$

where $(Y_1^{*a}, Y_2^{*a}) \notin A, (Y_1^{*b}, Y_2^{*b}) \notin B$, $A \bigcap B = \emptyset$, and $A \bigcup B = U$. The last inequality is obtained by checking the expected cost associated with decisions apart from $(Y_1^{*\ell}, Y_2^{*\ell})$. The above equation can further be bounded from below by

$$
\frac{b_2+h_2}{2\sqrt{T}}\sum_{t=1}^{T}\max\{\mathbb{P}_a^{\phi}((Y_1,Y_2)\in A), \mathbb{P}_b^{\phi}((Y_1,Y_2)\ \in B)\}
$$

using the inequality $\max\{\sum_{t=1}^T y_t, \sum_{t=1}^T z_t\} \geq (1/2)\sum_{t=1}^T (y_t + z_t) \geq (1/2)\sum_{t=1}^T \max\{y_t, z_t\}$ for nonnegative y_t, z_t. Consider the following two hypotheses,

$$
\begin{aligned}
H_a: &\quad F = F_a\\
H_b: &\quad F = F_b
\end{aligned}
$$

and let γ_t be a decision rule, mapping $\mathcal{H}_{t-1}$ into set $\{a, b\}$, where a represents that H_a is true and b represents that H_b is true. By using Tsybakov (2008) Theorem 2.2, similar to that in Besbes and Muharremoglu (2013), we can get the probability of making a mistake for any decision rule is bounded from below as follows

$$\inf_{\gamma_t} \max\{\mathbb{P}_a(\gamma_t = b), \mathbb{P}_b(\gamma_t = a)\} \geq \frac{1}{4}\exp\{-\mathscr{K}_{t-1}(\mathbb{P}_a, \mathbb{P}_b)\}$$

where

$$\mathscr{K}_t(\mathbb{P}_a, \mathbb{P}_b) = \mathbb{E}_a\left[\log \frac{\mathbb{P}_a(\vec{D}_1, \cdots, \vec{D}_t)}{\mathbb{P}_b(\vec{D}_1, \cdots, \vec{D}_t)}\right]$$

is the Kullback-Leibler divergence between the distributions of $\{\vec{D}_1, \cdots, \vec{D}_t\}$ under F_a and under F_b. Therefore,

$$\max\{\mathbb{P}_a^{\phi}((Y_1, Y_2) \in A), \mathbb{P}_b^{\phi}((Y_1, Y_2) \ \in B)\} \geq \frac{1}{4}\exp\{-\mathscr{K}_{t-1}(\mathbb{P}_a, \mathbb{P}_b)\}$$

Note that by our construction, the Kullback-Leibler divergence $\mathscr{K}_t(\mathbb{P}_a, \mathbb{P}_b)$ can be written as

$$\mathscr{K}_t(\mathbb{P}_a, \mathbb{P}_b) = t\left[(x_0 - \frac{1}{\sqrt{T}})\log\left(\frac{x_0 - \frac{1}{\sqrt{T}}}{x_0 + \frac{1}{\sqrt{T}}}\right) + (1 - x_0 + \frac{1}{\sqrt{T}})\log\left(\frac{1 - x_0 + \frac{1}{\sqrt{T}}}{1 - x_0 - \frac{1}{\sqrt{T}}}\right)\right]$$

By using the inequality

$$2z \leq \log\frac{1+z}{1-z} \leq 2z + 2z^2$$

we can show that $\mathscr{K}_t(\mathbb{P}_a, \mathbb{P}_b) \leq \frac{\frac{2}{x_0(1-x_0)^2}t}{T}$ for $T \geq 1$. This helps to yield that

$$\max\{\mathbb{P}_a^{\phi}((Y_1, Y_2) \in A), \mathbb{P}_b^{\phi}((Y_1, Y_2) \ \in B)\} \geq \frac{1}{4}\exp\{-\frac{\frac{2}{x_0(1-x_0)^2}(t-1)}{T}\} \geq \frac{1}{4}e^{-\frac{2}{x_0(1-x_0)^2}}$$

Consequently, the worst-case regret is bounded from below by

$$\frac{b_2 + h_2}{\sqrt{T}}\sum_{t=1}^{T}\frac{1}{4}e^{-\frac{2}{x_0(1-x_0)^2}} = \frac{b_2 + h_2}{4}e^{-\frac{2}{x_0(1-x_0)^2}}\sqrt{T}.$$

□

C.2 The Case of Lost Sales

C.2.1 Construction

Let $x_0 := \min\{\frac{h_2}{h_2+p_2-\frac{2}{3}s}, \frac{3(2p_2-s)}{4s}, \frac{2(2p_1-s)}{s}\}$. When T is big enough, specifically $T \geq \max\{\frac{1}{x_0^2}, \frac{1}{(1-x_0)^2}\}$, suppose that the demand follows one of the following two distributions:

F_a: $D_1 = 1, D_{12} = 1$,

$$D_2 = \begin{cases} 1 & \text{w.p. } x_0 - \frac{1}{\sqrt{T}} \\ 2 & \text{w.p. } 1 - x_0 + \frac{1}{\sqrt{T}} \end{cases}$$

one can easily check that the optimal inventory decision (Y_1^{*a}, Y_2^{*a}) under F_a is $(2, 2)$. And F_b: $D_1 = 1, D_{12} = 1$,

$$D_2 = \begin{cases} 1 & \text{w.p. } x_0 + \frac{1}{\sqrt{T}} \\ 2 & \text{w.p. } 1 - x_0 - \frac{1}{\sqrt{T}} \end{cases}$$

whose optimal inventory decision (Y_1^{*b}, Y_2^{*b}) is one of $(1, 2), (2, 0)$ and $(2, 3)$, depending on the parameters s, b_1, b_2, h_1, h_2, and T.

This is because

Demand / Decision	$\{1, 2_1, 2_2, 12\}$ w.p. $1 - x_0 + \frac{1}{\sqrt{T}}$			$\{1, 2, 12\}$ w.p. $x_0 - \frac{1}{\sqrt{T}}$		
(Y_1, Y_2)	holding cost	lost-sales cost	shipping cost	holding cost	lost-sales cost	shipping cost
$(0, 0)$	0	$2p_1 + 3p_2$	0	0	$2p_1 + 2p_2$	0
$(0, 1)$	0	$2p_1 + 2p_2$	s	0	$2p_1 + p_2$	s
$(0, 2)$	0	$2p_1 + p_2$	$2s$	0	$2p_1$	$2s$
$(0, 3)$	0	$2p_1$	$3s$	h_2	$2p_1$	$2s$
$(1, 0)$	0	$p_1 + 3p_2$	s	0	$p_1 + 2p_2$	s
$(1, 1)$	0	$p_1 + 2p_2$	$\frac{7}{4}s$	0	$p_1 + p_2$	$\frac{5}{3}s$
$(1, 2)$	0	$p_1 + p_2$	$\frac{31}{12}s$	0	p_1	$\frac{5}{2}s$
$(1, 3)$	0	p_1	$\frac{7}{2}s$	h_2	p_1	$\frac{5}{2}s$
$(2, 0)$	0	$3p_2$	$2s$	0	$2p_2$	$2s$
$(2, 1)$	0	$2p_2$	$\frac{8}{3}s$	0	p_2	$\frac{5}{2}s$
$(2, 2)$	0	p_2	$\frac{10}{3}s$	0	0	$3s$
$(2, 3)$	0	0	$4s$	h_2	0	$3s$

For simplicity, let $y_0 = x_0 - \frac{1}{\sqrt{T}}$. Then

$$\pi^{s\ell}(0,0) = 2p_1 + 2p_2 + y_0 p_2, \qquad \pi^{s\ell}(0,1) = 2p_1 + p_2 + s + y_0 p_2,$$
$$\pi^{s\ell}(0,2) = 2p_1 + 2s + y_0 p_2, \qquad \pi^{s\ell}(0,3) = h_2 + 2p_1 + 2s + y_0(s - h_2),$$
$$\pi^{s\ell}(1,0) = p_1 + 2p_2 + s + y_0 p_2, \qquad \pi^{s\ell}(1,1) = p_1 + p_2 + \frac{5}{3}s + y_0(p_2 + \frac{1}{12}s),$$
$$\pi^{s\ell}(1,2) = p_1 + \frac{5}{2}s + y_0(p_2 + \frac{1}{12}s), \qquad \pi^{s\ell}(1,3) = h_2 + p_1 + \frac{5}{2}s + y_0(s - h_2),$$
$$\pi^{s\ell}(2,0) = 2p_2 + 2s + y_0(p_2 - s), \qquad \pi^{s\ell}(2,1) = p_2 + \frac{5}{2}s + y_0(p_2 + \frac{1}{6}s),$$
$$\pi^{s\ell}(2,2) = 3s + y_0(p_2 + \frac{1}{3}s), \qquad \pi^{s\ell}(2,3) = h_2 + 3s + y_0(s - h_2).$$

It is obvious from $s \leq p_i$ and $y_0 < 1$ that

$$\pi^{s\ell}(0,0) > \pi^{s\ell}(0,1) > \pi^{s\ell}(0,2) > \pi^{s\ell}(1,2), \qquad \pi^{s\ell}(0,3) > \pi^{s\ell}(1,3) > \pi^{s\ell}(2,3),$$
$$\pi^{s\ell}(1,0) > \pi^{s\ell}(1,2) > \pi^{s\ell}(1,2), \qquad \pi^{s\ell}(2,1) > \pi^{s\ell}(2,2).$$

One can easily check that $(2,2)$ is the optimal decision under F_a and is not optimal under F_b.

C.2.2 Proof for Theorem 5 Under Our Construction

The proof is exactly the same as that in the case of backlog. We omit it here.

Appendix D

Proofs for Cyc-PIO and Est-PIO Algorithms

D.1 Regret Upper Bound for Cyc-PIO Algorithm

D.1.1 Single-Product Case with Lost Sales

Proof. As δ_k is the probability for Z_k to be rounded to $\lceil Z_k \rceil$ (when $\delta_k > 0$), we have

$$\mathbb{P}(\tau_{k+1} - \tau_k = \ell) = \begin{cases} \delta_k (1-\delta_k)^{\ell-1} & \text{if } \ell < n_k \\ \delta_k (1-\delta_k)^{n_k - 1} + (1-\delta_k)^{n_k} & \text{if } \ell = n_k \end{cases}$$

The expected loss in the kth cycle is

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1} (\pi(Y_t) - \pi(Z^*))\right] \\
=&\mathbb{E}\left[\sum_{\ell=1}^{n_k} \mathbb{P}(\tau_{k+1} - \tau_k = \ell)[(\pi(\lceil Z_k \rceil) - \pi(Z^*)) + (\ell - 1)(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))]\right] \\
=&\mathbb{E}\left[\sum_{\ell=1}^{n_k} \delta_k (1-\delta_k)^{\ell-1}[(\pi(\lceil Z_k \rceil) - \pi(Z^*)) + (\ell - 1)(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))]\right] \\
&+ \mathbb{E}\left[(1-\delta_k)^{n_k}[[(\pi(\lceil Z_k \rceil) - \pi(Z^*)) + (n_k - 1)(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))]]\right] \\
=&\mathbb{E}[\pi(\lceil Z_k \rceil) - \pi(Z^*)] + \mathbb{E}[(n_k - 1)(1-\delta_k)^{n_k}(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))] \\
&+ \mathbb{E}\left[\delta_k \sum_{\ell=1}^{n_k - 1} \ell (1-\delta_k)^{\ell} \pi(\lfloor Z_k \rfloor) - \pi(Z^*))\right] \\
=&\mathbb{E}[\pi(\lceil Z_k \rceil) - \pi(Z^*)] + \mathbb{E}[(n_k - 1)(1-\delta_k)^{n_k}(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))]
\end{aligned}$$

$$
\begin{aligned}
&+\mathbb{E}\left[\frac{(1-\delta_k)[-n_k(1-\delta_k)^{n_k-1}+(n_k-1)(1-\delta_k)^{n_k}+1]}{\delta_k}(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))\right] \\
=&\mathbb{E}\left[(\pi(\lceil Z_k\rceil)-\pi(Z^*))+\frac{-(1-\delta_k)^{n_k}+(1-\delta_k)}{\delta_k}(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))\right] \\
=&\mathbb{E}\left[\frac{[\delta_k(\pi(\lceil Z_k\rceil)-\pi(Z^*))+(1-\delta_k)(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))]-(1-\delta_k)^{n_k}(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))}{\delta_k}\right] \\
=&\mathbb{E}\left[\frac{(\pi(Z_k)-\pi(Z^*))-(1-\delta_k)^{n_k}(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))}{\delta_k}\right]
\end{aligned}
$$

From the performance of our PIO algorithm (regret upper bound and lower bound), we know that $\mathbb{E}[\pi(Z_k)-\pi(Z^*)] = O(\frac{1}{\sqrt{k}})$. Hence $\mathbb{E}|Z_k-Z^*| = O(\frac{1}{\sqrt{k}})$. Therefore, there exists a k_0 such that for $k > k_0$, we have $\mathbb{E}|Z_k - Z^*| < \frac{1}{2}$. For those Z_k such that $|Z_k - Z^*| < \frac{1}{2}$, from the piecewise-linearity, we know that for $k > k_0$, there exists constants c_1, c_2 such that

$$
\pi^\ell(Z_k) - \pi^\ell(Z^*) = \begin{cases} c_1(Z_k - Z^*) & \text{if } Z_k \geq Z^* \\ c_2(Z_k - Z^*) & \text{if } Z_k < Z^* \end{cases}
$$

where $c_1 \geq 0$ and $c_2 \leq 0$. Furthermore, since $\delta_k = Z_k - \lfloor Z_k \rfloor$, we have for $k > k_0$,

$$
\delta_k = \begin{cases} Z_k - Z^* & \text{if } Z_k \geq Z^* \\ 1-(Z^* - Z_k) & \text{if } Z_k < Z^* \end{cases}
$$

Therefore, we can rewrite the expected loss in the kth cycle as

$$
\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t)-\pi(Z^*))\right] = \begin{cases} c_1 & \text{if } Z_k \geq Z^* \\ c_2 \cdot \mathbb{E}\left[\frac{(Z^*-Z_k)^k+(Z^*-Z_k)}{1-(Z^*-Z_k)}\right] & \text{if } Z_k < Z^* \end{cases}
$$

for $k > k_0$ and $|Z_k - Z^*| < \frac{1}{2}$.

For $k > k_0$, we claim that

$$
\mathbb{P}(|Z_k - Z^*| \geq \frac{1}{2}) = O(\frac{1}{\sqrt{k}}).
$$

This follows from the fact that $\mathbb{E}|Z_k - Z^*| = O(\frac{1}{\sqrt{k}})$. Therefore, if $|Z_k - Z^*| \geq \frac{1}{2}$, the expected loss in the kth cycle is upper bounded by a constant $c^{>0.5}$ because

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t)-\pi(Z^*))\middle| |Z_k-Z^*|\geq\frac{1}{2}\right]\mathbb{P}(|Z_k-Z^*|\geq\frac{1}{2}) \\
=&\mathbb{E}\left[(\pi(\lceil Z_k\rceil)-\pi(Z^*))+\frac{-(1-\delta_k)^{n_k}+(1-\delta_k)}{\delta_k}(\pi(\lfloor Z_k\rfloor)-\pi(Z^*))\middle| |Z_k-Z^*|\geq\frac{1}{2}\right] \\
&\cdot\mathbb{P}(|Z_k-Z^*|\geq\frac{1}{2}) \\
=&O(1)
\end{aligned}
$$

where the last equality follows from

$$\frac{-(1-\delta_k)^{n_k}+(1-\delta_k)}{\delta_k} = (1-\delta_k)[1+(1-\delta_k)+(1-\delta_k)^2+\cdots+(1-\delta_k)^{n_k-2}] < n_k = O(\sqrt{k})$$

and that $\mathbb{P}(|Z_k - Z^*| \geq \frac{1}{2}) = O(\frac{1}{\sqrt{k}})$. Moreover,

$$\mathbb{E}[\#\{k < K : |Z_k - Z^*| \geq \frac{1}{2}\}] = O(\sqrt{T})$$

which is obvious from the fact that $\sum_{k=1}^{K} \mathbb{E}[\pi(Z_K) - \pi(Z^*)] = O(\sqrt{T})$.

In the following, we examine the expected cycle length of the kth cycle, which is denoted as m_k, when $|Z_k - Z^*| < \frac{1}{2}$. For $k > k_0$ such that $Z_k > Z^*$, we have

$$\begin{aligned}
\mathbb{E}[m_k] &= \sum_{\ell=1}^{n_k} \ell\mathbb{P}(m_k = \ell) \\
&= \sum_{\ell=1}^{n_k-1} \ell\delta_k(1-\delta_k)^{\ell-1} + n_k(\delta_k(1-\delta_k)^{n_k-1} + (1-\delta_k)^{n_k}) \\
&= \frac{1-(1-\delta_k)^{n_k}}{\delta_k} \\
&= \Omega(\sqrt{k})
\end{aligned}$$

where the last equality is derived by monotonicity with respect to δ_k, L'Hospital's rule, and the fact that $\delta_k = Z_k - Z^* = O(\frac{1}{\sqrt{k}})$. On the other hand, for $k > k_0$ and $Z_k < Z^*$, we have

$$\mathbb{E}[m_k] = \frac{1-(1-\delta_k)^{\sqrt{k}}}{\delta_k} = \Theta(1)$$

Let $u_1, u_2, \cdots, u_q$ ($u_1 < u_2 < \cdots < u_q$) denote the indices of cycles such that $0 < Z_{u_i} - Z^* < \frac{1}{2}$. Then we have

$$T \geq (u_1-1)+c\sqrt{u_1}+(u_2-u_1-1)+c\sqrt{u_2}+(u_3-u_2-1)+c\sqrt{u_3}+\cdots+c\sqrt{u_q}+(T-u_q-1)$$

where $\sqrt{u_k} \geq \sqrt{k}$, $q \leq T$. Therefore, $q = O(T^{2/3})$ because $\sum_{k=1}^{q} \sqrt{k} \leq \int_0^q \sqrt{t}dt = \frac{2}{3}q^{\frac{3}{2}}$.

By Wald's equation, for $k > k_0$, if $Z^* < Z_k < Z^* + \frac{1}{2}$, the expected loss in the kth cycle is

$$\begin{aligned}
&\mathbb{E}[\sum_{t=\tau_k}^{\tau_{k+1}-1} (\pi(Y_t) - \pi(Z^*))] \\
=&\mathbb{E}[\sum_{t=\tau_k}^{\tau_{k+1}-2} (\pi(\rho(Z_k)) - \pi(Z^*)) + (\pi(\lceil Z_k \rceil) - \pi(Z^*))]
\end{aligned}$$

$$
\begin{aligned}
&\leq \mathbb{E}[\sum_{t=\tau_k}^{\tau_k+\tau}(\pi(\rho(Z_k)) - \pi(Z^*))] + \mathbb{E}[(\pi(\lceil Z_k \rceil) - \pi(Z^*))] \\
&= \mathbb{E}[\mathbb{E}[\tau](\pi(Z_k) - \pi(Z^*))] + \mathbb{E}[(\pi(\lceil Z_k \rceil) - \pi(Z^*))] \\
&= \mathbb{E}[\frac{1}{\delta_k}(\pi(Z_k) - \pi(Z^*))] + \mathbb{E}[(\pi(\lceil Z_k \rceil) - \pi(Z^*))] \\
&= c_1 + c_1 \\
&= 2c_1
\end{aligned}
$$

where τ is a stopping time which is defined as the first time of reaching $\lceil Z_k \rceil$ if we do probabilistic rounding for infinite times. If $Z^* - \frac{1}{2} < Z_k < Z^*$, the expected loss in the kth cycle is

$$
\begin{aligned}
&\mathbb{E}[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))] \\
=&\mathbb{E}\left[\frac{(\pi(Z_k) - \pi(Z^*)) - (1-\delta_k)^{n_k}(\pi(\lfloor Z_k \rfloor) - \pi(Z^*))}{\delta_k}\right] \\
\leq&\mathbb{E}[2((\pi(Z_k) - \pi(Z^*))] + 2\frac{c'}{\sqrt{k}}\mathbb{E}[\pi(\lfloor Z_k \rfloor) - \pi(Z^*)] \\
=&\mathbb{E}[2((\pi(Z_k) - \pi(Z^*))] + 2\frac{c' \cdot (-c_2)}{\sqrt{k}}
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{k=1}^{K}\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\right] \\
=&\mathbb{E}\left[\sum_{k=1}^{k_0}\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\right] + \mathbb{E}\left[\sum_{k=k_0+1}^{K}\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\right] \\
\leq&\sum_{k=1}^{k_0} k\max\{\pi(Z) - \pi(Z^*)\} + \mathbb{E}\left[\sum_{k=k_0+1}^{K}\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\middle| |Z_k - Z^*| < \frac{1}{2}\right]\right] \\
&+ \mathbb{E}\left[\sum_{k=k_0+1}^{K}\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\middle| |Z_k - Z^*| \geq \frac{1}{2}\right]\mathbb{P}(|Z_k - Z^*| \geq \frac{1}{2})\right] \\
=&\sum_{k=1}^{k_0} k\max\{\pi(Z) - \pi(Z^*)\} + \mathbb{E}\left[\sum_{\substack{k_0<k\leq K \\ Z_k > Z^*}}\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\middle| |Z_k - Z^*| < \frac{1}{2}\right]\right] \\
&+ \mathbb{E}\left[\sum_{\substack{k_0<k\leq K \\ Z_k < Z^*}}\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t) - \pi(Z^*))\middle| |Z_k - Z^*| < \frac{1}{2}\right]\right]
\end{aligned}
$$

$$
\begin{aligned}
&+\mathbb{E}\left[\sum_{k=k_0+1}^{K}\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}(\pi(Y_t)-\pi(Z^*))\middle| |Z_k-Z^*|\geq\frac{1}{2}\right]\mathbb{P}(|Z_k-Z^*|\geq\frac{1}{2})\right]\\
\leq&\sum_{k=1}^{k_0}k\max\{\pi(Z)-\pi(Z^*)\}+2c_1\mathbb{E}[\#\{k<K:Z^*<Z_k<Z^*+\frac{1}{2}\}]\\
&+2\sum_{k=k_0+1}^{K}\left(\mathbb{E}[\pi(Z_k)-\pi(Z^*)]+\frac{c'}{\sqrt{k}}\mathbb{E}[\pi(Z^*-1)-\pi(Z^*)]\right)\\
&+c^{>0.5}\mathbb{E}[\#\{k\leq K:|Z_k-Z^*|\geq\frac{1}{2}\}]\\
=&\sum_{k=1}^{k_0}k\max\{\pi(Z)-\pi(Z^*)\}+2c_1q\\
&+2\sum_{k=k_0+1}^{K}\left(\mathbb{E}[\pi(Z_k)-\pi(Z^*)]+\frac{c'}{\sqrt{k}}\mathbb{E}[\pi(Z^*-1)-\pi(Z^*)]\right)\\
&+c^{>0.5}\mathbb{E}[\#\{k\leq K:|Z_k-Z^*|\geq\frac{1}{2}\}]\\
=&O(1)+O(T^{2/3})+O(T^{1/2})+O(T^{1/2})\\
=&O(T^{2/3})
\end{aligned}
$$

where $\sum_{k=k_0+1}^{K}\mathbb{E}[\pi(Z_k)-\pi(Z^*)]=O(K^{1/2})$ and $O(K^{1/2})=O(T^{1/2})$. Thus, the regret is $O(T^{2/3})$. □

D.1.2 Two-Product Case with Lost Sales and Fulfillment Dynamics

Proof. The kth cycle's truncated length is defined to be u_k, i.e., $\tau_{k+1}=\min\{t>\tau_k:\vec{Y}_t=(\lceil Z_{k,1}\rceil,\lceil Z_{k,2}\rceil)\}\wedge(\tau_k+n_k)$. Let $\delta_k=\min\{Z_{k,1}-\lfloor Z_{k,1}\rfloor, Z_{k,2}-\lfloor Z_{k,2}\rfloor\}$, which is the probability for $\vec{Z}_k$ being rounded to $\lceil\vec{Z}_k\rceil$. Therefore,

$$
\mathbb{P}(\tau_{k+1}-\tau_k=\ell)=\begin{cases}\delta_k(1-\delta_k)^{\ell-1} & \text{if } \ell<n_k\\ \delta_k(1-\delta_k)^{n_k-1}+(1-\delta_k)^{n_k} & \text{if } \ell=n_k\end{cases}
$$

So, the expected loss in the kth cycle is

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=\tau_k}^{\tau_{k+1}-1}\mathbb{E}[\pi(\vec{Y}_t)-\pi(\vec{Z}^*))]\right]\\
=&\mathbb{E}\left[\sum_{\ell=1}^{n_k}\mathbb{P}(\tau_{k+1}-\tau_k=\ell)\right[(\pi(\lceil\vec{Z}_k\rceil)-\pi(\vec{Z}^*))
\end{aligned}
$$

$$
\begin{aligned}
&+ (\ell - 1)\left(\frac{\xi_k}{1-\delta_k}(\pi(\lfloor \vec{Z}_k \rfloor) - \pi(\vec{Z}^*)) + \frac{1-\xi_k-\delta_k}{1-\delta_k}(\pi(\lfloor \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right)\Bigg]\Bigg] \\
=&\mathbb{E}\left[\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*)\right] \\
&+ \mathbb{E}\Bigg[\left(\delta_k \sum_{\ell=1}^{n_k-1} \ell(1-\delta_k)^{\ell} + (n_k-1)(1-\delta_k)^{n_k}\right) \cdot \\
&\qquad \left(\frac{\xi_k}{1-\delta_k}(\pi(\lfloor \vec{Z}_k \rfloor) - \pi(\vec{Z}^*)) + \frac{1-\xi_k-\delta_k}{1-\delta_k}(\pi(\lfloor \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right)\Bigg] \\
=&\mathbb{E}\left[\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*)\right] \\
&+ \mathbb{E}\Bigg[\left(\delta_k \sum_{\ell=1}^{n_k-1} \ell(1-\delta_k)^{\ell-1} + (n_k-1)(1-\delta_k)^{n_k-1}\right) \cdot \\
&\qquad \left((\pi(\vec{Z}_k) - \pi(\vec{Z}^*)) - \delta_k(\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right)\Bigg] \\
=&\mathbb{E}\left[\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*)\right] \\
&+ \mathbb{E}\Bigg[\left(\frac{(n_k-1)(1-\delta_k)^{n_k} - n_k(1-\delta_k)^{n_k-1} + 1}{\delta_k} + (n_k-1)(1-\delta_k)^{n_k-1}\right) \cdot \\
&\qquad \left((\pi(\vec{Z}_k) - \pi(\vec{Z}^*)) - \delta_k(\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right)\Bigg] \\
=&\mathbb{E}\left[\left(\frac{(n_k-1)(1-\delta_k)^{n_k} - n_k(1-\delta_k)^{n_k-1} + 1}{\delta_k} + (n_k-1)(1-\delta_k)^{n_k-1}\right)(\pi(\vec{Z}_k) - \pi(\vec{Z}^*))\right] \\
&- \mathbb{E}\left[\left(\frac{(n_k-1)(1-\delta_k)^{n_k} - n_k(1-\delta_k)^{n_k-1}}{\delta_k} + (n_k-1)(1-\delta_k)^{n_k-1}\right)\delta_k(\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right] \\
=&\mathbb{E}\left[\frac{1-(1-\delta_k)^{n_k-1}}{\delta_k}(\pi(\vec{Z}_k) - \pi(\vec{Z}^*))\right] + \mathbb{E}\left[(1-\delta_k)^{n_k-1}(\pi(\lceil \vec{Z}_k \rceil) - \pi(\vec{Z}^*))\right]
\end{aligned}
$$

where

$$
\lfloor \vec{Z}_k \rceil = \begin{cases} (\lfloor Z_1 \rfloor, \lceil Z_2 \rceil) & \text{if } Z_1 - \lfloor Z_1 \rfloor \le Z_2 - \lfloor Z_2 \rfloor \\ (\lceil Z_1 \rceil, \lfloor Z_2 \rfloor) & \text{if } Z_1 - \lfloor Z_1 \rfloor > Z_2 - \lfloor Z_2 \rfloor \end{cases}
$$

and $\xi_k = \min\{\lceil Z_{k,1} \rceil - Z_{k,1}, \lceil Z_{k,2} \rceil - Z_{k,2}\}$, which is the probability of $\vec{Z}_k$ being rounded to $\lfloor \vec{Z}_k \rfloor$.

The remaining proof is mostly like the single-product case.

□

D.2 Regret Upper Bound for Est-PIO Algorithm

D.2.1 Single-Product Case with Lost Sales

Proof. We claim that the left-subderivative at $\{1, 2, \cdots, Z^*, Z^*+1\}$ will be updated in the correct direction (with correct sign) after K iterations with high probability $1 - O(\frac{1}{\sqrt{K}})$.

Suppose the suboptimality gap (i.e., the expected cost difference between the optimal decision and the best suboptimal decision), which is strictly greater than 0, is ε.

Note that Kolmogorov's law of iterated logarithm implies uniform convergence, i.e., $\frac{S_{n(K)}}{n(K)} \rightrightarrows \mathbb{P}(D < Y_0)$, *a.s.*. Therefore, for any Y_0 in the admissible region, there exists a $N(Y_0)$ such that with probability 1,

$$\left| \frac{\sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Y_0, D_t < Y_0\}}{\sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Y_0\}} - \mathbb{P}(D < Y_0) \right| < \frac{\varepsilon}{2},$$

for any $n(K) = \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Y_0\} \geq N(Y_0)$.

Hence, if $Y_k \geq Y_0$ is satisfied for no less than $\tau := \max_{Y_0}\{N(Y_0)\}$ times, then the non-zero left-subderivative is updated in the corrected direction. From now on, we prove that $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Y_0\} = \infty) = 1$ for any $Y_0 \in \{1, 2, \cdots, Z^*, Z^*+1\}$. We prove this by using mathematical induction.

It is obvious that $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq 1\} = \infty) = 1$ by the lower bound of demand and hence admissible region. If the hypothesis holds for $Y_0 \leq z$ where $z \leq Z^*$, i.e., $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Y_0\} = \infty) = 1$ for $Y_0 \leq z$ with $z \leq Z^*$, we check for $z+1$.

If the estimated left-subderivative at $z+1$ is currently incorrect, we differentiate two cases: $z < Z^*$ and $z = Z^*$. For $z < Z^*$, the left-subderivative at $z+1$ would be non-negative. By our hypothesis, we assume that the left-subgradients at $\{1, 2, \cdots, z\}$ are now updated in the correct direction. If Z is greater than $z+1$ at the current iteration k, then it will certainly be rounded to a value greater than or equal to $z+1$; if Z is smaller than z, because

$$\frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{k+1}} + \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{k+2}} + \cdots + \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{2k}} \geq k \cdot \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{2k}} \geq \frac{\varepsilon\gamma}{4}$$
$$\frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{2k+1}} + \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{2k+2}} + \cdots + \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{4k}} \geq 2k \cdot \frac{\frac{\varepsilon}{2} \cdot \gamma}{\sqrt{4k}} \geq \frac{\varepsilon\gamma}{4}$$
$$\cdots$$

we know that Z would reach $(z, \beta]$ in less than or equal to $2^{\lceil \frac{4z}{\varepsilon\gamma} \rceil} + 1$ iterations, resulting in more than $\frac{1}{\sqrt{k+2^{\lceil \frac{4z}{\varepsilon\gamma} \rceil}+1}}$ chance to be rounded to a value no less than $z+1$. If Z will never come back to a value smaller than or equal to z after finitely many iterations, then the left-subderivative at $z+1$ shall be updated correctly in finite iterations, as the expected number of iterations for Z to be rounded to a value larger than or equal to $z+1$ is less than or equal to

$$\sum_{n=1}^{\infty} n(1-\frac{c}{n})^{n+k+2^{\lceil \frac{4z}{\varepsilon\gamma} \rceil}+1},$$

which is finite. In this case, $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq z+1\} = \infty) = 1$. Otherwise, Z will be smaller than or equal to z for infinitely many times. When k is large enough, the last iteration before Z comes into the interval $[1,z]$, Z would lie in $(z, z+1]$. As the estimated subgradient lies within $[-b,h]$, we know that when $k > \frac{2\sqrt{2}h}{\varepsilon}$, Z would be updated to a value more than z in less than k steps as

$$k \cdot \frac{\frac{\varepsilon}{2}\cdot\gamma}{\sqrt{2k}} \geq \frac{\gamma h}{\sqrt{k}}$$

If the left-subderivative at $z+1$ is updated in the correct direction, i.e., becomes negative, then Z will never be back to $[1,z]$. Otherwise, $z+1$ still has nonnegative left-subderivative and the expected number of steps being rounded to $z+1$ is less than or equal to

$$\begin{aligned}\sum_{n=1}^{\infty} n\frac{c+\gamma}{\sqrt{k_n}} \prod_{m=1}^{n-1}(1-\frac{c}{\sqrt{k_m}}) \leq \sum_{n=1}^{\infty} n\frac{c+\gamma}{\sqrt{k}} \prod_{m=1}^{n-1}(1-\frac{c}{\sqrt{mk}}) \leq \sum_{n=1}^{\infty} \frac{c+\gamma}{\sqrt{k}} \frac{n}{e^{\frac{c}{\sqrt{k}}\sum_{m=1}^{n-1}\frac{1}{\sqrt{m}}}} \\ \leq \sum_{n=1}^{\infty} \frac{c+\gamma}{\sqrt{k}} \frac{n}{e^{\frac{c}{\sqrt{k}}(n-1)\frac{1}{\sqrt{n-1}}}} = \sum_{n=1}^{\infty} \frac{c+\gamma}{\sqrt{k}} \frac{n}{e^{\frac{c}{\sqrt{k}}\sqrt{n-1}}}\end{aligned}$$

which is finite. Here, we assume $\prod_{m=1}^{0}(1-\frac{c}{\sqrt{mk}}) = 1$ and make use of the inequality $\log(1+x) \leq x$ for $x > -1$. In the above formula, k_n is the iteration that Z lies in the interval $(z, z+1)$ after k, and $k_n < nk$ as

$$k \cdot \frac{\frac{\varepsilon}{2}\cdot\gamma}{\sqrt{(n+1)k}} \geq \frac{\gamma h}{\sqrt{nk}}.$$

This implies that

$$\mathbb{P}(\eta_{N(z+1)}(z) > K) = O(\frac{1}{\sqrt{K}})$$

where $\eta_k(z) = \min\{n : \sum_{m=1}^{n} \mathbb{I}\{Y_m \geq z+1\} \geq k\}$ (prove by contradiction by using $\mathbb{E}[\eta_{N(z+1)}(z)] < \infty$). Therefore, after finite iterations, the left-subderivative at $z+1$ would be updated in the correct direction, also resulting in $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq z+1\} = \infty) = 1$ for infinitely many times.

For $z = Z^*$, because the left-subderivatives at $\{1, 2, \cdots, Z^*\}$ are updated in the correct direction and are all negative, we can conclude that $\mathbb{P}(\lim_{K\to\infty} \sum_{k=1}^{K} \mathbb{I}\{Y_k \geq Z^*+1\} =$

$\infty) = 1$ as analysis before. Therefore, we have proved that the left-subderivative at $\{1, 2, \cdots, Z^*, Z^* + 1\}$ will be updated in the correct direction (with the correct sign) after finite iterations.

We also claim that Z will be in $[Z^* + 1, \beta]$ for finitely many times. Otherwise, the left-subderivatives at $Z^* + 2$ and hence at $Z^* + 1$ would be updated in the correct direction, i.e., with a positive value. Thereby, when k is large enough, Z will not be in $[Z^* + 1, \beta]$.

Next, we prove that the regret is $O(\sqrt{T})$. Note that Z will be in $(Z^* - 1, Z^* + 1)$ within finitely many steps, and all the left-subderivatives at $\{1, 2, \cdots, Z^*, Z^* + 1\}$ will be updated in the correct direction with an error less than $\frac{\varepsilon}{2}$ after some iteration K_0 with high probability $1 - O(\frac{1}{\sqrt{T}})$. If there is a single optimal decision, then the regret would be less than

$$\sum_{k=K_0}^{T} \left[\frac{\max\{b, h\}\gamma}{\sqrt{k}} + \frac{c(\pi(Z^* + 1) - \pi(Z^*)}{\sqrt{k}} + \bar{C}\frac{\max_Z\{\pi(Z)\} - \pi(Z^*)}{\sqrt{k}} \right] = O(\sqrt{T}).$$

where the first term in the summation is caused by the fluctuation of Z around the optimal decision, the second term originates from the probability rounding rule and the third term results from the incorrect update of the estimated left-subderivative. □

D.2.2 Two-Product Case with Lost Sales and Fulfillment Dynamics

Proof. Under the single-optimal assumption, we can easily conclude that the left-subderivative along any axis is non-zero by using convexity. Therefore, the proof for the two-product scenario assembles that for the single-product. We omit the proof here for brevity. □