Feature-Based Nonparametric Inventory Control with Censored Demand

Jingying Ding[a], Woonghee Tim Huh[b], and Ying Rong[a]

[a]Antai College of Economics and Management, Shanghai Jiao Tong University, Shanghai 200030, China.
[b]Operations and Logistics Division, Sauder School of Business, University of British Columbia, Vancouver, Canada V6T 1Z2.

Abstract

***Problem definition*:** We study stochastic periodic-review inventory systems with lost sales, where the decision maker has no access to the true demand distribution a priori and can only observe historical sales data (referred to as censored demand) and feature information about the demand. ***Academic/practical relevance*:** In an inventory system, excess demand is unobservable due to inventory constraints, and sales data alone cannot fully recover the true demand. Meanwhile, feature information about the demand is abundant to assist inventory decisions. We incorporate features for inventory systems with censored demand. ***Methodology*:** We propose two feature-based nonparametric inventory algorithms called the feature-based adaptive inventory algorithm and the dynamic shrinkage algorithm. Both algorithms are based on the stochastic gradient descent method. We measure the performance of the proposed algorithms through the average expected regret in finite periods: that is, the difference between the cost of our algorithms and that of a clairvoyant optimal policy with access to information, which is acting optimally. ***Results*:** We show that the average expected cost incurred under both algorithms converges to the clairvoyant optimal cost at the rate of $O(1/\sqrt{T})$. The feature-based adaptive inventory algorithm results in high volatility in the stochastic gradients, which hampers the initial performance of regret. The dynamic shrinkage algorithm uses a shrinkage parameter to adjust the gradients, which significantly improves the initial performance. ***Managerial implications*:** This paper considers feature information. The idea of dynamic shrinkage for the stochastic gradient descent method builds on a fundamental insight known as the bias-variance trade-off. Our research shows the importance of incorporating the bias-variance in a dynamic environment for inventory systems with feature information.

Keywords: Censored Demand, Learning algorithms, Bias and Variance Trade-off, Inventory

1 Introduction

Nowadays, the decision-makers (DMs) of retailers can access data from different sources, which include advertisement plans, product reviews, product display position on shelf, local events from social media, and so on. The abundance of the data brings an opportunity for DMs to analyze and

utilize the data to improve their inventory decisions. However, when excess demand is unobservable and only sales data (known as censored demand) is available, the data alone cannot fully recover the true demand, especially when the number of observations is limited. How can a DM dynamically learn the impact of the data on demand and simultaneously adjust inventory decisions to minimize cost over time?

To address this question, we consider a periodic-review lost-sales inventory system with demand based on feature information generated from data in T periods. The decision-maker has no access to the true demand distribution a priori and can only observe historical sales data (censored demand) and features about the demand. The DM knows that the demand is linear in the features, but the exact relationship is unknown to the DM. Therefore the DM has to learn not only the demand distribution, but also how the features impact the demand. The learning is much more complicated in the setting considering features. When a stockout occurs, it could be due to error associated with the feature parameters, but it is hard to tell which one or more of these parameters are responsible. For censored demand, all we know is whether the product stocked out or not. Based on this single binary indicator, we have the challenge of updating not only our inventory level, but also learning feature parameters, etc. We propose two feature-based nonparametric algorithms and measure their performance according to the average expected regret in T periods.

1.1 Literature Review

Our work is relevant to the following research streams.

Data-driven approaches for inventory systems without feature information: Various data-driven approaches for inventory systems have been proposed in the literature for nonparametric settings, where the demand distribution is not known.

Some approaches consider inventory problems using uncensored demand data. Sample average approximation (SAA) uses uncensored samples from the demand distributions to form empirical distribution (Levi et al., 2007, 2015). Liyanage and Shanthikumar (2005) used operational statistics to integrate parameter estimation and profit optimization. Subsequently, Chu et al. (2008) proposed a Bayesian analysis to show how to find the optimal operational statistic. Another approach is the bootstrap method, which estimates the newsvendor fractile of the demand distribution (Bookbinder and Lordahl, 1989). Some studies address the problem of inventory management from the perspective of robust optimization. (Bertsimas and Thiele, 2006; See and Sim, 2010; Mamani et al., 2017). Chen

et al. (2019) made replenishment and pricing decisions concurrently with backorders. Cheung and Simchi-Levi (2019) studied capacitated stochastic inventory control problems. Chen et al. (2020) proposed an algorithm for systems with random production capacity.

Because true demands may not be observed due to inventory constraint, censored demand has attracted more attention in recent years, where the firm only has access to past sales data rather than the true demand. One solution method is concave adaptive value estimation (CAVE) which utilizes censored data to approximate the cost function with a sequence of piecewise linear functions (Godfrey and Powell, 2001; Powell et al., 2004). Another solution method is based on online gradient descent (OGD) algorithms. Burnetas and Smith (2000) considered the combined problem of pricing and ordering and developed an adaptive policy for a perishable product. They showed the convergence result, but not the convergence rate. Huh and Rusmevichientong (2009) proposed algorithms based on the gradient descent and established the convergence rate. They also extended their result to the nonperishable inventory case. Huh et al. (2009) studied the problem of finding the best base-stock level in lost-sales systems with a positive lead time. Huh et al. (2011) developed a new class of adaptive data-driven policies based on the Kaplan-Meier estimator.

Inventory systems have been extended to the multiproduct case with a warehouse-capacity constraint by Shi et al. (2016). Zhang et al. (2018) proposed a learning algorithm to find the best base-stock policy in perishable inventory systems. Yuan et al. (2019) and Ban (2020) considered fixed ordering costs. Zhang et al. (2020) proposed learning algorithms for the lost-sales inventory system with positive lead times. However, these papers do not take feature information into consideration, while our work incorporates this information in inventory systems with censored demand.

Models with features: In the current literature, some papers incorporate exogenous feature information in inventory models. Hannah et al. (2010) proposed solution methods that depend on the weighted empirical stochastic optimization problem. Ban and Rudin (2019) incorporated features into the newsvendor problem and proposed single-step machine-learning algorithms. In these two papers, the decision is made in a static fashion and true demand can be fully observed. In contrast, our algorithms allow the DM to dynamically adjust the order quantities when more sales data is being observed.

The incorporation of features is also applied in fields other than inventory management. Ferreira et al. (2016) and Ban and Keskin (2020) used features data to optimize pricing decisions. Bertsimas and Kallus (2019) combined ideas from machine learning and operations research to develop a

framework for using data to make optimal decisions. Bastani and Bayati (2020) considered decisions at the individual level with high-dimensional data and formulated the problem as a multi-armed contextual bandit. Feng and Shanthikumar (2018) summarized the research in demand management and manufacturing with big data.

Online convex optimization: The framework of online convex optimization was first defined in the machine learning literature (Zinkevich, 2003) and has been significantly extended since. In online convex optimization, the DM does not know the associations between decisions and outcomes: namely the objective function is unknown. The goal is to minimize the regret, which is the difference between the cost incurred by the implemented decision and by the best decision in hindsight. Zinkevich (2003) has shown that the average T-period cost using an algorithm based on online gradient descent converges to the optimal cost at the rate of $O(1/\sqrt{T})$. Flaxman et al. (2005) extended this result to the bandit setting and showed that the estimator of the objective value is sufficient to approximate the gradient descent. The case where an unbiased estimator of the gradient is unknown has also been studied by Kleinberg (2004). Algorithms proposed by Hazan et al. (2006) achieve logarithmic regret $O(\log T/T)$ under the assumption that functions are strongly convex. We refer interested readers to Hazan (2016) and Shalev-Shwartz (2012) for an overview. Since the target levels in our problem may not be achieved due to inventory constraints in the nonperishable inventory case, the problem differs from the conventional online convex optimization problems.

1.2 Main Results and Contributions

We propose two feature-based nonparametric algorithms for stochastic inventory systems with censored demand and feature information: the feature-based adaptive inventory algorithm and the dynamic shrinkage algorithm. We show that the average expected regrets of both algorithms converge to zero at the rate of $O(1/\sqrt{T})$ and the result holds for both perishable and nonperishable inventory cases. As far as we know, there is no result ofthe convergence rate for inventory problems including features in the literature. Our algorithms are based on the stochastic gradient descent (SGD) method. The work is closest to Huh and Rusmevichientong (2009). They considered the independent and identically distributed (i.i.d) demand processes without features information and obtained the same asymptotic convergence rate of $O(1/\sqrt{T})$. Compared to Huh and Rusmevichientong (2009), our work is novel in terms of both the demand setting and algorithm design/analysis.

First, we allow the demands to have correlations over time through features in our setting. The majority of the literature in inventory management for censored demand assumes that the demands are independent and identically distributed, but they are naturally correlated over periods. The driving factor is the demand generation mechanism. For example, some features like weather condition and seasonality can have a temporal correlation, which results in the demands being correlated over time. In this paper, we incorporate features into models that can capture demand correlations. In addition, compared to other inventory works that include features (Hannah et al., 2010; Ban and Rudin, 2019), we consider a dynamic environment with partially observed demand (i.e., censored demand).

Second, bringing features into decisions can improve the quality of decisions eventually. However, in the initial periods, the limited number of sales observations can cause an issue of order quantity volatility, which may damage the performance. To address this issue, we introduce a shrinkage factor to adjust the magnitude of gradients in our proposed dynamic shrinkage algorithm. The diminishing shrinkage degree over time balances the short term performance (the initial performance is much improved numerically) with the long term performance (the regret remains $O(1/\sqrt{T})$).

Third, from the technical perspective, the key challenge in our analysis is to derive an upper bound of the difference between the target order-up-to level and the actual implemented order-up-to level (because of the positive inventory carry-over from previous periods in the nonperishable inventory case). To address this problem, we connect such difference with the waiting time process, which was first proposed by Huh and Rusmevichientong (2009). However, different from the identical and independent demand assumption made in Huh and Rusmevichientong (2009), nonstationary and correlated demands through the variability of features makes analysis more complex in our setting. We identify that strong convexity enables us to establish the convergence rate in solutions (in addition to objective value) so that we can set bounds for the waiting time process.

1.3 Organization of the Paper

The paper is organized as follows. In Section 2, we describe the problem formulation and settings in detail, in which we present assumptions and the clairvoyant optimal policy. We propose our basic algorithm, the feature-based adaptive inventory algorithm, in Section 3 and highlight the key idea of the algorithm. Then, we show the theoretic result that the average expected regret convergence rate of the feature-based adaptive inventory algorithm is $O(1/\sqrt{T})$ for both perishable and nonperishable

inventory cases in Section 4. Next, Section 5 shows the issue of the volatility of initial gradients and provides the dynamic shrinkage algorithm to overcome the issue. Subsequently, in Section 6, we present the results of numerical experiments and analyze the insights. Section 7 provides the concluding remarks.

2 Problem Formulation

We consider a multiperiod stochastic inventory planning problem with censored demand and demand feature information. The true demand distribution is unknown to the DM, and he could only observe historical sales data and feature information at the beginning of the current period. We consider two separate cases in which the inventory at the end of a period either perishes or is carried over to the next period.

Random Demand: Our lost-sales inventory system is over a finite horizon of T periods, from $t = 1, ..., T$. We denote D_t and d_t to be the random demand and realized demand in period t, respectively. Feature information x_t is a N-dimension vector, which denotes exogenous variables that affect the demand in period t and can be observed before demand realization. The component of x_t can be a basic feature or any polynomial transformation of a basic feature like a square term or a cross term. The first component of x_t is fixed to be the constant 1 (an intercept term).

Assumptions: In this paper, we make the following assumptions about the demand.

(a) The demand D_t is linear in x_t. That is, $D_t = w \cdot x_t + \delta_t$,[1] where w is a N-dimension vector and δ_t is an independent and identically distributed (i.i.d) random variable. We do not require x_t to be i.i.d. Notice that both w and the distribution of δ is unknown to the DM.

(b) Let Ω_0 be a certain compact box set in $\mathbb{R}^N$, from which the value of w is chosen. Each component of w is bounded on the interval $[\underline{w}, \overline{w}]$, and Ω_0 is the cross product of these N closed intervals, i.e., $\Omega_0 = [\underline{w}, \overline{w}]^N$. Note that we do not require all components of w to sum up to 1.

(c) Assume the domain of x_t is a compact set $\mathbb{X}$. We can normalize the domain of x_t to be nonnegative, and the norm of x_t has an upper bound x_U, which is defined as $x_U = \max_t \|x_t\|$.

(d) Define a compact rectangle set Ω in $\mathbb{R}^N$, which is only different from the set Ω_0 in the first component in that it is bounded on the interval $[\underline{w}', \overline{w}']$ not $[\underline{w}, \overline{w}]$. For any $x_t \in \mathbb{X}$, $z \in \Omega$, and

[1] The notation " $\cdot$ " means the inner product of two vectors. For any real vectors, $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, $\boldsymbol{x} \cdot \boldsymbol{y} = \sum_{i=1}^{n} x_i y_i$.

$w \in \Omega_0$, there exists $\underline{\delta}$ and $\overline{\delta}$ such that

$$\underline{\delta} \leq z \cdot x_t - w \cdot x_t \leq \overline{\delta},\ \forall t \in T.$$

(e) The error term δ_t has a cumulative distribution function (CDF) $\Phi(\cdot)$ and a probability density function (PDF) $\phi(\cdot)$. With compact support $[\underline{\delta}, \overline{\delta}]$, for any $\delta_t \in [\underline{\delta}, \overline{\delta}]$, we assume $\phi(\delta_t) > \theta$, where θ is a positive constant.

(f) For any feature vector x, $inf\ E[D|x] > 0$.

Assumption (a) sets up the relationship between demand and features, which is common in the literature (Ban and Rudin, 2019). *Assumptions* (b) and (c) specify the condition of w and x_t, respectively, which is needed in the context of online convex optimization (Hazan, 2016). *Assumptions* (d) and (e) will be used later to guarantee the strongly convexity of the objective function in the proof of Lemma 3. *Assumption* (f) indicates the demand is always positive. Thus, in the nonperishable inventory case, the inventory would not be stuck to a certain level.

System Dynamics. For period $t \geq 1$, we assume the following sequence of events.

(i) At the beginning of each period t, the manager observes the feature information x_t and the initial on-hand inventory level $u_t \geq 0$. Moreover, we assume $u_1 = 0$.

(ii) The manager makes a replenishment decision based on the on-hand inventory and current feature information. The lead time is zero. The decision y_t represents the order-up-to level in period t.

(iii) The realized demand d_t occurs. The manager only observes the sales (i.e., the minimum of the realized demand and inventory level $\min\{d_t, y_t\}$). The initial on-hand inventory in the next period is given by $u_{t+1} = 0$ in the perishable case and by $u_{t+1} = [y_t - d_t]^+$ in the nonperishable case.

(iv) The overage cost and underage cost at the end of period t is generated by $h[y_t - d_t]^+ + b[d_t - y_t]^+$, where h and b represent the per-unit holding and lost-sales cost, respectively.

Clairvoyant Optimal Policy and Objectives. Suppose the DM knows the distribution of δ_t and the exact value of the vector w. For any horizon T, the optimal expected cost over T periods with no starting inventory can be solved by dynamic programming. For any order-up-to inventory

level y_t, let $\Pi_t(y_t)$ denote the expected one-period overage and underage cost in period t,[2], where:

$$\Pi_t(y_t) = h \cdot E[y_t - D_t]^+ + b \cdot E[D_t - y_t]^+. \tag{1}$$

Let $c_t(u_t)$ be the optimal cost in period t given the starting inventory is u_t. Let $C_{T+1}(\cdot) = 0$. Then, the dynamic programming formulation can be written as

$$c_t(u_t) = min_{y_t \geq u_t} \Pi_t(y_t) + E[c_{t+1}(u_{t+1}(y_t, D_t))], \tag{2}$$

where $u_{t+1}(y_t, D_t) = 0$ for the perishable case and $u_{t+1}(y_t, D_t) = (y_t - D_t)^+$ for the nonperishable case.

In the classical inventory model where the manager knows the demand distribution, a myopic solution is optimal to Equation (2) if D_t is i.i.d. Thus, the stationary multiperiod inventory model is analytically equivalent to the single-period newsvendor model. In our setting, D_t consists of two parts: the deterministic part $w \cdot x_t$ and the random error term δ_t. In the following proposition, we show the myopic solution $\arg\min_{y_t} \Pi_t(y_t)$ is still optimal.

Proposition 1. *Suppose the distribution of δ_t and the exact vector w are known to the DM. In period t, upon observing a N-dimension vector x_t, the optimal solution to Equation* (2) *is*

$$y_t^* = w^* \cdot x_t,$$

where w^* is also a N-dimension vector, which is defined below.

$$(w^*)_1 = (w)_1 + \Phi^{-1}(\frac{b}{b+h}), \tag{3}$$

$$(w^*)_j = (w)_j, \ j = 2, ..., N. \tag{4}$$

In Proposition 1, for each $j = 2, ..., N$, the j^{th} component of w^* is exactly equal to the j^{th} component of w. For $j = 1$, recall that the first component of x_t is the constant 1. The first component of w^* is the sum of the first component of w and the newsvendor fractile. Suppose that there is no feature information (i.e. x_t is always 0 except the first component, which is fixed at 1). Then Proposition 1 leads to traditional results. The proof of Proposition 1 is presented in the Appendix.

[2]We use index t since demand D_t is decided by features x_t in period t. It could be more accurate to write it as $\Pi_t(y_t|x_t)$, but to avoid being cumbersome, we will use $\Pi_t(y_t)$ when there is no ambiguity. Demands may not be independent and identically distributed across periods due to the process of x_t.

In Sections 3 and 5, we propose feature-based policies π to obtain a sequence of inventory levels $\{y_t \ : \ t \geq 1\}$, whose average expected cost converges to the clairvoyant optimal cost. We require that the inventory level y_t depends only on the sales quantities and the feature information observed before placing the order. To measure the performance of the proposed policy $\pi = \{y_t \ : \ t \geq 1\}$, we use the clairvoyant optimal cost as the benchmark. The T-period average expected regret is defined as follows:

$$Regret_T(\pi) = E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t^{\pi}) - \Pi_t(y_t^*))\right],$$

where the superscript π denotes the policy under consideration.

3 Feature-Based Adaptive Inventory Algorithm

In this section, we present the feature-based adaptive inventory (FAI) algorithm. In the algorithm, we define $\{\hat{y}_t : t \geq 1\}$ to be the *desired* order-up-to level in period t. We use $\{y_t : t \geq 1\}$ to represent the *implemented* order-up-to level. In the perishable case, the manager can always reach the desired order-up-to level, i.e., $y_t = \hat{y}_t$. In the nonperishable case, the inventory level after ordering is bounded below by the initial inventory level, i.e., $y_t = \max\{\hat{y}_t, u_t\}$, which means the manager could either order up to the desired order-up-to level or keep the inventory level intact.

We briefly describe the main idea of our algorithm. Recall from Section 2 that we show that $y_t^* = w^* \cdot x_t$ is the clairvoyant optimal order quantity in period t. In the algorithm, we use z_t to represent our guess in period t for w^*. In the FAI algorithm, we set $\hat{y}_t = z_t \cdot x_t$. The main purpose of the algorithm is to update the vector z_t towards w^*. At the beginning of period t, after observing the sales data in the previous period, we decide $\hat{y}_t$ based on z_t and feature information x_t. After the end of period t, after we observe sales, we adjust and obtain z_{t+1} by applying SGD. The detail of the FAI algorithm is as follows.

Algorithm 1: Feature-Based Adaptive Inventory Algorithm.

Initialization. Set $u_1 = 0$. Let z_1 be any value drawn from Ω and $y_1 = \hat{y}_1 = z_1 \cdot x_1$.

Main Step.

For each period $t \geq 1$, repeat the following procedure:

$$\begin{aligned} z_{t+1} &= P_{\Omega}(z_t - \varepsilon_t H_t(z_t)) \\ \hat{y}_{t+1} &= z_{t+1} \cdot x_{t+1} \\ y_{t+1} &= \max\{\hat{y}_{t+1}, u_{t+1}\} \end{aligned}$$

The function $P_{\Omega}(\cdot)$ projects any vector z to its closest vector in Ω. The step size ε_t and the

function $H_t(\cdot)$ are given below. Let $\mu = (h+b)\tilde{\theta}$, where $\tilde{\theta}$ is a positive constant satisfying $0 < \tilde{\theta} \leq \theta$ (θ is defined in *Assumption* (e)). Let the step size ε_t be given by:

$$\varepsilon_t = \frac{1}{(h+b)\tilde{\theta}t} = \frac{1}{\mu t}. \tag{5}$$

The random vector $H_t(z_t)$ has two points in its support for given x_t value and is defined as:

$$H_t(z_t) = \begin{cases} h \cdot x_t, & \text{if } D_t < \hat{y}_t, \\ -b \cdot x_t, & \text{if } D_t \geq \hat{y}_t. \end{cases} \tag{6}$$

In the FAI algorithm, $P_\Omega(\cdot)$ essentially maps $z_{t,1}$ to the closest point in $[\underline{w}', \overline{w}']$ and $z_{t,i}$ to the closest point in $[\underline{w}, \overline{w}]$ for $i = 2, ..., N$.[3] Moreover, we set the step size ε_t to be the same when updating all components of z_t. The random vector $H_t(z_t)$ is decided by the sales data in period t. This random vector $H_t(z_t)$ is used to update z_t to z_{t+1}, which in turn combines with the feature information in period $t+1$ to specify the desired order-up-to level for period $t+1$. In the case $D_t < \hat{y}_t$, we have strictly positive ending inventory and we will adjust vector z_{t+1} to be smaller. In the other case $D_t \geq \hat{y}_t$, it corresponds to a zero ending inventory, and we will adjust vector z_{t+1} to be larger.

The FAI algorithm generalizes the method used in Huh and Rusmevichientong (2009) by incorporating feature information. In addition to learn the newsvendor fractile $\Phi^{-1}(\frac{b}{b+h})$, the FAI algorithm also needs to learn the value w over time. However, due to the discrepancy between z_t and w^*, the difference between the implemented order-up-to level y_t and the desired order-up-to $\hat{y}_t$ in the nonperishable case is driven not only by the randomness of error term δ_t but also by the variation of x_t over time. Since the variation of x_t can follow an arbitrary process, it will make the analysis more difficult. In the next section, we will overcome this difficulty to establish the analytic result for the regret.

4 Convergence Rate of the FAI Algorithm

The main result in this section is that the expected average cost of the FAI algorithm converges to the clairvoyant cost with vector w and the error distribution known at the rate of $O(\log T/T)$ in the perishable case and $O(1/\sqrt{T})$ in the nonperishable case. The result is stated in Theorem 2.

Theorem 2. *Under Assumption (a)-(f), the order-up-to levels* $\{y_t : t \geq 1\}$ *generated by the FAI algorithm have the following properties:*

[3] $z_{t,i}$ denotes the i-th component of vector z_t.

a. The perishable inventory case: For any $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi(y_t) - \Pi(y_t^*))\right] \leq \frac{(\max\{b,h\}x_U)^2}{2\mu T}(1+\log T).$$

b. The nonperishable inventory case: For any $T \geq 1$, there exists a constant C_1, such that

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi(y_t) - \Pi(y_t^*))\right] \leq \frac{(\max\{b,h\}x_U)^2}{2\mu T}(1+\log T) + \frac{C_1}{\sqrt{T}}.$$

For the remainder of this section, we prove Theorem 2. To find the convergence rate of the algorithm stated in the theorem, we express:

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t) - \Pi_t(y_t^*))\right] = \Lambda_1(T) + \Lambda_2(T), \tag{7}$$

where

$$\Lambda_1(T) = E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(\hat{y}_t) - \Pi_t(y_t^*))\right]$$

and

$$\Lambda_2(T) = E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t) - \Pi_t(\hat{y}_t))\right].$$

We will show the convergence of $\Lambda_1(T)$ and $\Lambda_2(T)$, respectively.

4.1 Perishable Inventory Case

In the case of perishable inventory, y_t will not be constrained by the on-hand inventory level, so the desired order-up-to level can always be implemented. Thus, we have $y_t = \hat{y}_t$, and Λ_2 becomes zero. Based on the definitions of the desired order-up-to level $\hat{y}_t = z_t \cdot x_t$ and the clairvoyant optimal order-up-to level $y_t^* = w^* \cdot x_t$, we can represent $\Pi_t(\hat{y}_t)$ and $\Pi_t(y_t^*)$ by defining $\Psi_t(z_t)$ and $\Psi_t(w^*)$, respectively, where

$$\Psi_t(z_t) = h \cdot E[z_t \cdot x_t - D_t]^+ + b \cdot E[D_t - z_t \cdot x_t]^+$$

and

$$\Psi_t(w^*) = h \cdot E[w^* \cdot x_t - D_t]^+ + b \cdot E[D_t - w^* \cdot x_t]^+.$$

Note that $\Psi_t(\cdot)$ and $\Pi_t(\cdot)$ are essentially the same, since $\Psi_t(z_t) = \Pi_t(z_t \cdot x_t)$. While $\Psi_t(\cdot)$ is the function of the estimated weight vector and $\Pi_t(\cdot)$ is the function of the inventory level.

Preliminary: strongly convexity. Recall the *Assumption* (e) that for any residual δ_t belongs to $[\underline{\delta}, \overline{\delta}]$, the PDF function of the residual satisfies $\phi(\delta_t) > \theta$, where θ is a positive constant. With this condition, we can prove that the cost function $\Psi_t(z_t)$ is strongly convex in vector z_t. The proof of Lemma 3 is given in the Appendix. Based on the definition of strong convexity (Boyd et al., 2004), we can say that a function is α-strongly convex if

$$f(y) \geq f(x) + \nabla f(x) \cdot (y - x) + \frac{\alpha}{2}\|y - x\|^2.$$

Lemma 3. *The cost function $\Psi_t(z_t)$ is μ-strongly convex in $z_t \in \Omega$, where $\mu = (h + b)\tilde{\theta}$, defined in Section 3.*

We first show bounds on $\Lambda_1(T)$ with the condition that the cost functions are strongly convex. Based on definitions $\Pi_t(\cdot)$ and $\Psi_t(\cdot)$, we have

$$\Lambda_1(T) = E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi(\hat{y}_t) - \Pi(y_t^*))\right] = E\left[\frac{1}{T}\sum_{t=1}^{T}(\Psi_t(z_t) - \Psi_t(w^*))\right].$$

When the exact gradient is accessible, Hazan (2016) establishes logarithmic bounds on the regret if the cost functions are strongly convex. Lemma 4 is a minor adaption of this result, which extends the case of the gradient to the case of a stochastic gradient. The proof of Lemma 4 is presented in the Appendix.

Lemma 4. *Let S be a compact convex set in R^n. Let $f : S \to R$ be an α-convex function defined on S. For any $v \in S$, let $H(v)$ be a stochastic gradient of f at v, i.e., $E[H(v)|v] = \nabla f(v)$. Suppose that there exists $\overline{G}$ such that $\|H(v)\| \leq \overline{G}$ with probability one for all $v \in S$. For any $t \geq 1$, recursively define:*

$$v_{t+1} = P_S(v_t - \eta_t H(v_t)), \tag{8}$$

where $P_S(\cdot)$ is the projection operator and step size $\eta_t = \frac{1}{\alpha t}$. Then, it achieves the following guarantee for all $T \geq 1$:

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(f(v_t) - f(v^*))\right] \leq \frac{\overline{G}^2}{2\alpha T}(1 + \log T),$$

where $v^* = \arg\min_{v \in S} f(v)$.

In our setting, it is easy to verify that $\|H_t(z_t)\| \leq \max\{b, h\}x_U$. Let $S = \Omega$ and $\overline{G} = \max\{b, h\}x_U$, and we have the following inequality from Lemma 4:

$$\Lambda_1(T) = E[\frac{1}{T}\sum_{t=1}^{T}(\Psi_t(z_t) - \Psi_t(w^*))] \leq \frac{(\max\{b, h\}x_U)^2}{2\mu T}(1 + \log T).$$

In the perishable inventory case, recall from $y_t = \hat{y}_t$ that $\Lambda_2(T)$ is always zero. Thus, the bound for the perishable case only depends on Λ_1. We complete the proof of Theorem 2(a) for the perishable inventory case.

4.2 Nonperishable Inventory Case

4.2.1 Connection between the stochastic gradient descent method and the waiting time process

To complete the proof of Theorem 2(b) for the nonperishable inventory case, we study the convergence rate of Λ_2 in this section. First, using the property of strongly convexity, we introduce Lemma 5, which will be used to set bounds for the distance between y_t and $\hat{y}_t$ in Theorem 6.

Lemma 5. *Consider a convex optimization problem*

$$\min_{a \in S} f(a),$$

where S is a convex set. Let a^* be the optimal solution. Let $\alpha_0 \in S$. The general stochastic gradient descent (SGD) algorithm recursively defines:

$$a_{k+1} = P_S(a_k - \gamma_k G_k),$$

where S is a convex set and G_k is a random vector satisfying $\mathbb{E}(G_k) = \nabla f(a_k)$. Assume that $\mathbb{E}(\|G_k\|^2) \leq \overline{G}_0^2$ and $f(a)$ is strongly convex, i.e., there exists a positive number $\alpha > 0$ satisfying, for any a_1 and a_2,

$$f(a_1) - f(a_2) \geq \nabla f(a_2) \cdot (a_1 - a_2) + \frac{\alpha}{2}\|a_1 - a_2\|^2.$$

Let $\gamma_k = \frac{1}{\alpha k}$, for each k. Then

$$\mathbb{E}(\|a_k - a^*\|^2) \leq \frac{\max\{\|a_1 - a^*\|^2, \frac{\overline{G}_0^2}{\alpha^2}\}}{k}.$$

The main idea of Lemma 5 is to find an upper bound of the difference between the optimal solution and the solution obtained by the SGD method in each iteration. This is a key enabler of the results in our paper. The existing SGD literature mainly focuses on the performance of the regret, while the convergence of the solutions has been little studied. We aim to use the solution convergence property to deal with the difficulty of feature changes across periods. The proof of Lemma 5 appears in the Appendix.

In this section, we want to establish a connection between the application of the stochastic gradient descent method and the waiting time process. Let ρ_0 be

$$\rho_0 = \frac{\max\{b,h\}x_U^2}{\mu} + \sqrt{\max\{\|z_1 - w^*\|^2, \frac{(\max\{b,h\}x_U)^2}{\mu^2}\}}x_U, \tag{9}$$

which is a constant. For any $\rho \geq \rho_0$, the stochastic process $(M_t(\rho) \mid t \geq 1)$ is defined as $M_0(\rho) = 0$ and

$$M_{t+1}(\rho) = [M_t(\rho) + \frac{\rho}{\sqrt{t}} - B_t]^+, \tag{10}$$

where $B_t's$ are defined as $B_t = w \cdot x_{t+1} + \delta_t$ for any $t \geq 1$. The random variable $M_t(\rho)$ can be interpreted as the waiting time of the t-th customer in the queuing system, where B_t is the inter-arrival time between the t-th and $(t+1)$-th customer, and $\frac{\rho}{\sqrt{t}}$ is the service time of the t-th customer.

The following Theorem 6 is the main result in this section, which illustrates the connections between the excess inventory $y_t - \hat{y}_t$ and the waiting time of the t-th customer in the process $M_t(\rho)$ described above. The excess inventory is the difference between the implemented inventory level y_t and the target inventory level $\hat{y}_t$ given by the stochastic gradient descent method due to the positive inventory carry-over from the previous periods. The idea is similar to Huh and Rusmevichientong (2009), who established a relationship between the amount of inventory in excess of the target level and the waiting time process in a GI/D/1 queue. However, because Huh and Rusmevichientong (2009) do not consider that the impact of the demand feature and demands are identical and independently distributed across periods, they only need the bounds of gradients to set up such a connection. The difficulty we encounter is that once we incorporate features into the algorithm, the demands become nonstationary and correlated through the variability of features, which makes the analysis more complex in our setting. With the presence of the demand feature, we show that $y_{t+1} - \hat{y}_{t+1}$ is also related to the interaction between $(z_t - w^*)$, the difference between the SGD solution and the clairvoyant optimal solution, and $(x_t - x_{t+1})$, the evolution of the demand feature. Relying on the property of strong convexity, we can bound on $(z_t - w^*)$.[4] By Lemma 5, which shows the convergence rate of the solution by the SGD method to the optimal solution, we obtain the following Theorem 6.

Theorem 6. *Consider the nonperishable inventory. For any t, $y_t - \hat{y}_t \leq M_t(\rho)$ with probability one, where $M_t(\rho)$ is defined in Equation* (10).

[4] If we do not consider the features, the interaction is always zero.

Proof By our definition, the difference $y_t - \hat{y}_t$ is always nonnegative. We claim that for any $t \geq 1$,

$$y_{t+1} - \hat{y}_{t+1} \leq [y_t - \hat{y}_t + \frac{\rho_0}{\sqrt{t}} - w \cdot x_{t+1} - \delta_t]^+. \tag{11}$$

We consider the relationship between the starting inventory level u_{t+1} and the target inventory level $\hat{y}_{t+1}$. If $u_{t+1} \leq \hat{y}_{t+1}$, then $y_{t+1} = \hat{y}_{t+1}$ holds, which implies the above claim. Suppose $u_{t+1} > \hat{y}_{t+1}$. Then we have $y_{t+1} = y_t - d_t$ and

$$\begin{aligned} y_{t+1} - \hat{y}_{t+1} &= y_{t+1} - z_{t+1} \cdot x_{t+1} \\ &= y_{t+1} - P_{[\underline{w},\overline{w}]}(z_t - \varepsilon_t H_t(z_t)) \cdot x_{t+1} \\ &\leq y_{t+1} - (z_t - \varepsilon_t H_t(z_t)) \cdot x_{t+1} \\ &= y_t - d_t - (z_t - \varepsilon_t H_t(z_t)) \cdot x_{t+1}. \end{aligned}$$

Above, the inequality is due to the fact that stockout cannot happen in period t when $u_{t+1} > \hat{y}_{t+1}$, implying $H_t(z_t) = hx_t$, in which case $\varepsilon_t H_t(z_t)$ is positive. Since z_t is bounded by Ω, $z_t - \varepsilon_t H_t(z_t)$ is no greater than z_t, and we have $z_t - \varepsilon_t H_t(z_t) \leq P_\Omega(z_t - \varepsilon_t H_t(z_t))$. Since x_{t+1} is nonnegative, the above inequality holds. Then, since $\hat{y}_t = z_t \cdot x_t$ and $d_t = w \cdot x_t + \delta_t$, we have

$$\begin{aligned} y_{t+1} - \hat{y}_{t+1} &\leq y_t - d_t - (z_t - \varepsilon_t H_t(z_t)) \cdot x_{t+1} \\ &= y_t - d_t - (z_t - \varepsilon_t H_t(z_t)) \cdot (x_t + x_{t+1} - x_t) \\ &= y_t - \hat{y}_t - d_t + \varepsilon_t H_t(z_t) \cdot x_t - z_t \cdot (x_{t+1} - x_t) + \varepsilon_t H_t(z_t) \cdot (x_{t+1} - x_t) \\ &= y_t - \hat{y}_t - w \cdot x_t - \delta_t - z_t \cdot (x_{t+1} - x_t) + \varepsilon_t H_t(z_t) \cdot x_{t+1} \\ &= y_t - \hat{y}_t + \varepsilon_t H_t(z_t) \cdot x_{t+1} + (z_t - w) \cdot (x_t - x_{t+1}) - w \cdot x_{t+1} - \delta_t. \end{aligned} \tag{12}$$

Before we proceed, we set bounds for $\varepsilon_t H_t(z_t)$ and $(z_t - w) \cdot (x_t - x_{t+1})$ in the above expression. We first consider $\varepsilon_t H_t(z_t)$. Based on the definition of ε_t and $H_t(z_t)$ (see Equations (5) and (6)), it follows $\varepsilon_t H_t(z_t) \cdot x_{t+1} \leq \frac{\rho_1}{t}$, where $\rho_1 \triangleq \max\{h, b\} x_U^2/\mu$ is a constant.

We set a bound for $(z_t - w) \cdot (x_t - x_{t+1})$. Recall that the function $\Psi_t(z_t)$ has a strong convexity in $z_t \in \Omega$; z_t is updated by the stochastic gradient descent method, where the step size $\varepsilon_t = \frac{1}{\mu t}$; and we have $\mathbb{E}(\|H_t(z_t)\|^2) \leq (\max\{b, h\} x_U)^2$. According to Lemmas 3 and 5, we can obtain that

$$\mathbb{E}(\|z_t - w^*\|^2) \leq \frac{\max\left\{\|z_1 - w^*\|^2, \frac{(max\{b,h\}x_U)^2}{\mu^2}\right\}}{t} = \frac{\rho_2}{t},$$

where $\rho_2 \triangleq \max\{\|z_1 - w^*\|^2, \frac{(\max\{b,h\}x_U)^2}{\mu^2}\}$ is a constant. Recall $D_t = w \cdot x_t + \delta_t$. Note $y_t^* = w^* \cdot x_t$, and w and w^* only differ in the first component. In the term $(z_t - w) \cdot (x_t - x_{t+1})$, the first component of $x_t - x_{t+1}$ is zero, since the first term of feature vector x_t is a constant term and is always 1. Thus,

$$(z_t - w) \cdot (x_t - x_{t+1}) = (z_t - w^*) \cdot (x_t - x_{t+1}) \leq \sqrt{\rho_2/t} \|x_t - x_{t+1}\|,$$

showing the convergence rate $O(\frac{1}{\sqrt{t}})$.

Then, from Equation (12), we have

$$\begin{aligned} y_{t+1} - \hat{y}_{t+1} &\leq y_t - \hat{y}_t + \varepsilon_t H_t(z_t) x_{t+1} + (z_t - w) \cdot (x_t - x_{t+1}) - w \cdot x_{t+1} - \delta_t \\ &\leq y_t - \hat{y}_t + \frac{\rho_1}{t} + \frac{\sqrt{\rho_2}\|x_t - x_{t+1}\|}{\sqrt{t}} - w \cdot x_{t+1} - \delta_t \\ &\leq y_t - \hat{y}_t + \frac{\rho_0}{\sqrt{t}} - w \cdot x_{t+1} - \delta_t, \end{aligned}$$

where the last inequation holds since

$$\begin{aligned} \frac{\rho_1}{\sqrt{t}} + \sqrt{\rho_2}\|x_t - x_{t+1}\| &\leq \rho_1 + \sqrt{\rho_2} x_U \\ &= \frac{\max\{b,h\}x_U^2}{\mu} + \sqrt{\max\left\{\|z_1 - w^*\|^2, \frac{(\max\{b,h\}x_U)^2}{\mu^2}\right\}} x_U = \rho_0, \end{aligned}$$

from the definition of ρ_0 in Equation (9). This completes the proof of the claim shown in Equation (11).

Now, consider the stochastic process $(M_t(\rho)|t \geq 1)$ defined in Equation (10) by:

$$M_{t+1}(\rho) = [M_t(\rho) + \frac{\rho}{\sqrt{t}} - B_t]^+.$$

From the definition of B_t, we have $B_t = w \cdot x_{t+1} + \delta_t$. Since we assume the initial on-hand inventory is zero in the first period, the target inventory level can be achieved, so we have $y_1 - \hat{y}_1 = 0$. Since $y_{t+1} - \hat{y}_{t+1} \leq [y_t - \hat{y}_t + \frac{\rho_0}{\sqrt{t}} - w \cdot x_{t+1} - \delta_t]^+$ holds for all t from the above claim and $\rho \geq \rho_0$, it follows from the recursive definition of the M_t process that $y_t - \hat{y}_t \leq M_t(\rho)$ with probability one. □

4.2.2 Proof of Theorem 2

Theorem 6 provides an upper bound on $y_t - \hat{y}_t$ in terms of $M_t(\rho)$. Now we find an upper bound of $(M_t(\rho) \mid t \geq 1)$. The main idea is to use a stationary process to avoid the difficulty caused by the nonstationary M_t process. We can use the stochastic process $(W_t(\rho) \mid t \geq 1)$, where $W_{t+1}(\rho) = [W_t(\rho) + \rho - B_t]^+$. By Proposition 4 and Lemma 5 of Huh and Rusmevichientong (2009), for any $T \geq 1$, we have

$$E[\sum\nolimits_{t=1}^{T} M_t(\rho)] \leq \frac{14\rho\pi^2 E[(B_1 - E[B_1])^6]\sqrt{T}}{(E[B_1] - \rho)^6}. \tag{13}$$

Consider $\Lambda_2(T) = E[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t) - \Pi_t(\hat{y}_t))]$. Since $y_t = \max\{\hat{y}_t, u_t\}$, for each t, it follows that:

$$\begin{aligned}
\Pi_t(y_t) - \Pi_t(\hat{y}_t) &= h \cdot E[y_t - D_t]^+ + b \cdot E[D_t - y_t]^+ - h \cdot E[\hat{y}_t - D_t]^+ - b \cdot E[D_t - \hat{y}_t]^+ \\
&= h \cdot E[y_t - \max\{\hat{y}_t, D_t\}]^+ - b \cdot E[\min\{D_t, y_t\} - \hat{y}_t]^+ \\
&\leq h \cdot (y_t - \hat{y}_t).
\end{aligned}$$

It follows from Theorem 6 that $y_t - \hat{y}_t \leq M_t(\rho)$ with probability one. Therefore, for any $T \geq 1$,

$$\begin{aligned}
\Lambda_2(T) &= E[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t) - \Pi_t(\hat{y}_t))] \\
&\leq h \cdot E[\frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{y}_t)] \\
&\leq h \cdot E[\frac{1}{T}\sum_{t=1}^{T} M_t(\rho)] \\
&\leq \frac{14h\rho\pi^2 E[(B_1 - E[B_1])^6]}{(E[B_1] - \rho)^6\sqrt{T}},
\end{aligned}$$

where the last inequality follows from Equation (13).

Recall that we express the average expected cost as $\Lambda_1(T)+\Lambda_2(T)$ in Equation (7). In Section 4.1, we analyzed $\Lambda_1(T) = E[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(\hat{y}_t) - \Pi_t(y_t^*))]$ and proved Theorem 2(a) for the perishable inventory case. Here, we focus on the nonperishable inventory case based on the results established thus far. Then,

$$\begin{aligned}
E[\frac{1}{T}\sum_{t=1}^{T}(\Pi_t(y_t) - \Pi_t(y_t^*))] &= \Lambda_1(T) + \Lambda_2(T) \\
&\leq \frac{(\max\{b,h\}x_U)^2}{2\mu T}(1 + \log T) + \frac{14h\rho\pi^2 E[(B_1 - E[B_1])^6]}{(E[B_1] - \rho)^6\sqrt{T}}.
\end{aligned}$$

This completes the proof of Theorem 2 for the nonperishable inventory case.

5 Dynamic Shrinkage Algorithm

In this section, we propose a new shrinkage-based algorithm to adjust the gradient by introducing a shrinkage parameter. We make this extension based on our observation of numerical experiments. We have observed that using all features often provides a less satisfactory result than not using features when t is small. First, we compare the regret incurred by using all features with the regret of the case not using features, which motivates our new algorithm that can address the issue (Section 5.1).

Then, we present the details of the algorithm and show that the convergence result still holds under the new algorithm (Section 5.2).

5.1 Motivation

To measure the value of incorporating the feature information into the inventory decisions, we compare cases which differ in their utilization of feature information. In particular, we consider the following two cases: (i) in the first case, the DM makes inventory decisions incorporating all feature information, and (ii) in the second case, no feature information is leveraged. The latter case is common in practice, where since the information may be limited or costly to access, inventory decisions are made based on historical sales alone. In Figure 1, we assume the true demand is generated by 20 features, and in both cases, there are 20 features which influence the demand. For the case which incorporates features, we use the feature-based data-driven algorithm in Section 3 to compute order-up-to levels. Figure 1 plots the relationship between the regret and time periods in a log-log form.

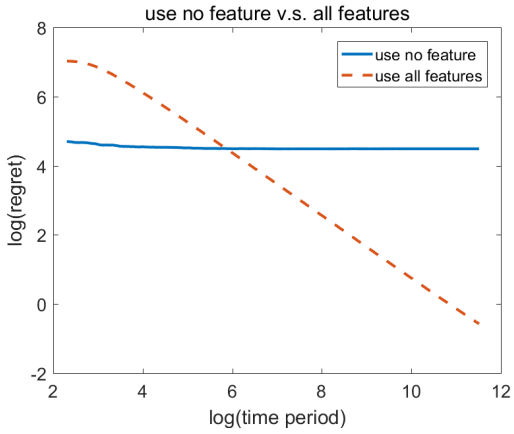


Figure 1: Capture part of features

From Figure 1, we observe that in the early periods, the case not using feature information outperforms the case using all features. However, as t increases, the case using all features prevails. It seems that using all feature information does harm the performance initially. At the beginning of the selling season, the DM has a few data points, and the number of features exceeds the number of observations. Consequently, it is difficult to obtain the coefficient of each feature accurately.

Furthermore, the step size is relatively large in the initial periods, which leads to the large volatility of z_t. The initial volatility has a prolonged impact on the following periods. As z_t is adjusted based on the previous z_{t-1}, the volatility will be carried over to the later periods. Due to the decreasing step size, the adjustment slows, and it needs more periods to eliminate the initial impact. In addition, consider the mechanism of the SGD method. We descend in a direction determined by a single binary outcome that $D_t < \hat{y}_t$ or not. Regardless of the importance of each feature, the components of z_t update based on the magnitude of the realized value for each feature. Thus, a one-dimension outcome leads to a high dimension change, which greatly increases the volatility of z_t. These factors cause the undesirable performance of the case using all features in the initial stage. Yet, as t increases and more data accumulates, the model including all features outperforms the other model. The case not using features does not converge to an optimal solution, while the case using all features converges to the clairvoyant optimal solution as shown in Section 4.

While our focus has been on the theoretical performance of using all features, we also hope the algorithm will have good performance in practice. The above-mentioned observation gives us some insights to reduce the volatility of z_t and then to improve the performance of the SGD method in practice in early periods. Considering much volatility is associated with the initial z_t, we borrow the idea of shrinkage in statistics that builds on a fundamental insight known as the bias-variance trade-off to address this issue. The idea of shrinkage has been applied in many problems. In regression methods, regularization terms like lasso or ridge terms are added to play the role of shrinkage. Lasso/ridge regression places a particular form of penalty on the parameters, which shrinks the coefficient estimates towards zero relative to the least squares estimates (Hoerl and Kennard, 1970; Tibshirani, 1996). The shrinkage method abandons the requirement of an unbiased estimator to reduce variance, and it can also perform variable selection. In statistics, James and Stein (1992) explicitly proposed a class of so-called James-Stein (JS) estimators, which introduced a shrinkage factor that can be written in the form $1 - \frac{b}{a+\|x\|^2}$. It is based on a result in statistics. The sample mean as the best unbiased estimator is strictly dominated under the expected utility criterion. Thus, it cannot be a rational choice. From a geometric perspective, the estimator shrinks each component of X toward the origin by a common factor of less than 1. The result is a biased estimation. The extent of shrinkage is specified by the parameters a and b.

All these methods are typically used in a static way. Considering the multi-period problem, we employ a dynamic adjusted shrinkage factor that changes the extent of shrinkage according to period t. We introduce a dynamic shrinkage factor β_t into the gradient, and it is a function of t. Similar to

the idea of the JS estimator, the factor β_t is less than 1 and "shrinks" each component of z_t (except the first term) towards zero from a geometric perspective: resulting in a biased gradient. Moreover, β_t approaches 1 as t increases. As demonstrated above, the initial order quantities are largely affected by volatile gradients. Thus, high degree shrinkage can help to address bias and variance trade-off. On the other hand, as t increases, we need to let β_t approach 1 so that the algorithm can converge to the true optimal order quantity.

In the following subsection, we will demonstrate our dynamic shrinkage method in detail and prove that the convergence result still holds under the new method. The numerical part will show that the shrinkage method helps to improve the initial performance of our algorithm in practice.

5.2 Dynamic shrinkage approach

We introduce a shrinkage factor β_t to adjust the gradient in period t. Let β_t be dependent on t and be defined as $\beta_t = 1 - e^{-\lambda t}$, where λ is a positive constant. Then, β_t is increasing in t and satisfies $0 < \beta_t \leq 1$. Define a N-dimension parameter vector $\boldsymbol{\beta_t} = [1, \beta_t, ..., \beta_t]$. Recall the definition of $H_t(z_t)$ (see Equation (6)). The random variable $\tilde{H}_t(z_t)$ is defined as:

$$\tilde{H}_t(z_t) = \boldsymbol{\beta_t} \cdot H_t(z_t).$$

Algorithm 2: Dynamic Shrinkage Algorithm.

Initialization. Set $u_1 = 0$. Let z_1 be any value drawn from Ω and $y_1 = \hat{y}_1 = z_1 \cdot x_1$.

Main Step.

For each period $t \geq 1$, repeat the following procedure:

$$\begin{aligned} z_{t+1} &= P_\Omega(z_t - \varepsilon_t \tilde{H}_t(z_t)) \\ \hat{y}_{t+1} &= z_{t+1} \cdot x_{t+1} \\ y_{t+1} &= \max\{\hat{y}_{t+1}, u_{t+1}\} \end{aligned}$$

Above, $\varepsilon_t = \frac{1}{\mu t}$ is the same as Equation (5).

The main idea of the dynamic shrinkage (DS) approach is to use a shrinkage parameter β_t to limit the initial volatility due to the gradient. As t increases, β_t converges to 1, and we do not need shrinkage any more. Theorem 7 shows that under the modified algorithm, the convergence rate still holds.

Theorem 7. *The order-up-to levels $\{y_t : t \geq 1\}$ generated by the DS algorithm has the following properties:*

a. The perishable inventory case: For any $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi(y_t) - \Pi(y_t^*))\right] \leq \left[(\frac{1}{\beta_1} - 1)\frac{\mu L}{T} + \frac{B}{\mu\beta_1 T}\right](1 + \log T).$$

b. The nonperishable inventory case: There exist L and B such that, for any $T \geq 1$,

$$E\left[\frac{1}{T}\sum_{t=1}^{T}(\Pi(y_t) - \Pi(y_t^*))\right] \leq \left[(\frac{1}{\beta_1} - 1)\frac{\mu L}{T} + \frac{B}{\mu\beta_1 T}\right](1 + \log T) + \frac{C_1}{\sqrt{T}},$$

where C_1 is defined in Theorem 2.

The proof of Theorem 7 follows the structure of the proof of Theorem 2, and we divide the regret into two parts $\Lambda_1(T)$ and $\Lambda_2(T)$ as in Equation (7). The main difference between the two algorithms is the updating rule for z_t, in which we use $\tilde{H}_t(\cdot)$ rather than $H_t(\cdot)$ where $\tilde{H}_t(z_t) = \boldsymbol{\beta_t} \cdot H_t(z_t)$. To show that the convergence rate of both $\Lambda_1(T)$ and $\Lambda_2(T)$ still holds under the DS algorithm, we note the following inequality used in the proof. The complete proof appears in the Appendix.

Using the new updating rule, we have the following inequality that is analogous to the proof of Lemma 4:

$$\begin{aligned}
E\|z_{t+1} - w^*\|^2 &= E\|P_\Omega(z_t - \varepsilon_t\tilde{H}(z_t)) - w^*\|^2 \\
&\leq E\|z_t - \varepsilon_t\tilde{H}(z_t) - w^*\|^2 \\
&= E\|(z_t - w^*) - \varepsilon_t\boldsymbol{\beta_t} \cdot H_t(z_t)\|^2 \\
&= E\|z_t - w^*\|^2 - 2\varepsilon_t\beta_t E[H(z_t) \cdot (z_t - w^*)] \\
&\quad +\varepsilon_t^2 E\|\tilde{H}(z_t)\|^2 - 2\varepsilon_t(1 - \beta_t)E[H^1(z_t)(z_t^1 - (w^*)^1)], \qquad (14)
\end{aligned}$$

where the superscript i denotes the ith component of the vector. The last equality stems from the definition of $\boldsymbol{\beta_t}$. We claim that there exists a constant B such that

$$E\|\tilde{H}(z_t)\|^2 - 2\frac{1 - \beta_t}{\varepsilon_t}E[H^1(z_t) \cdot (z_t^1 - (w^*)^1)] \leq B.$$

With the definition of β_t and ε_t, we have

$$\begin{aligned}
E\|\tilde{H}(z_t)\|^2 - 2\frac{1 - \beta_t}{\varepsilon_t}E[H^1(z_t) \cdot (z_t^1 - (w^*)^1)] &= E\|\tilde{H}(z_t)\|^2 - 2\mu t e^{-\lambda t}E[H^1(z_t) \cdot (z_t^1 - (w^*)^1)] \\
&\leq E\|H(z_t)\|^2 - 2\mu t e^{-\lambda t}E[H^1(z_t) \cdot (z_t^1 - (w^*)^1)].
\end{aligned}$$

Since $2\mu t e^{-\lambda t} > 0$ and is decreasing in $t > 1/\lambda$, the second term can be bounded. Also, $\|H(z_t)\|^2$ is bounded. Therefore, there exists B to upper bound $E\|\tilde{H}(z_t)\|^2 - 2\frac{1-\beta_t}{\varepsilon_t}E[H^1(z_t) \cdot (z_t^1 - (w^*)^1)]$.

Then from Equation (14), we have

$$E\|z_{t+1} - w^*\|^2 \leq E\|z_t - w^*\|^2 - 2\varepsilon_t\beta_t E[H(z_t) \cdot (z_t - w^*)] + \varepsilon_t^2 B. \tag{15}$$

The above inequality is used in the proof of Theorem 7 to account for the impact of shrinkage. More details are given in the Appendix.

6 Numerical Experiment

To measure the performance of the feature-based adaptive inventory (FAI) algorithm and the dynamic shrinkage (DS) algorithm, we compare them with the clairvoyant optimal policy, which uses more information and provides a lower bound on the best achievable performance of any policy. Our results show that the average cost generated by both the FAI and DS algorithm will converge to the clairvoyant cost at a certain rate. Since the performance may be influenced by various parameters, we also discuss situations with different values of parameters such as demand variability.

6.1 Performance of Algorithms

We study the performance of the FAI algorithm and the DS algorithm, comparing with the case which neglects feature information. We use the empirical risk minimization (ERM) algorithm as a benchmark, which is proposed by Ban and Rudin (2019). The ERM approach with regularization solves the newsvendor problem with feature data under uncensored demand. It is equivalent to high-dimensional quantile regression and the order-up-to quantity is obtained by solving the linear program (NV-ERM2) in Ban and Rudin (2019). They provided performance bounds on the out-of-sample cost of the ordering decisions chosen by the algorithm. Note that the ERM approach uses uncensored data rather than censored data, and thus the comparison shows the impact of censored demand.

In our experiments, we randomly generated $N = 100$ problem instances. Each problem instance consists of independent demand samples and parameters over a time horizon of 2000 periods. We compute the average cost till period t, where the one-period cost $\Pi_t(y_t)$ is given by

$$\Pi_t(y_t) = h \cdot [y_t - d_t]^+ + b \cdot [d_t - y_t]^+.$$

The instances are generated as follows: demand $D_t = w \cdot x_t + \delta_t$, in which w is a constant vector across N instances and each component is chosen from the interval $[1, 10]$. The components of x_t

are randomly drawn from a uniform distribution $U[1,2]$. The error term δ_t is normally distributed as $N(0,40)$. In all experiments, we set $h=1$ and $b=3$. The number of features is 20. With the knowledge of the true w and underlying δ distribution, we should set the clairvoyant optimal inventory level as $y_t^* = w \cdot x_t + \delta^*$, where $\delta^* = \min\{\delta : \Phi(\delta) \geq b/(b+h)\}$. Since the result of the perishable inventory case stems from classic online convex optimization, we focus on the nonperishable inventory case in this section.

The numerical experiments indicate that both the FAI algorithm and the DS algorithm work well and converge to the clairvoyant optimal cost at a certain rate. Note that $Regret \leq C/t^{\kappa}$ implies that $log(Regret)$ should be approximately linear in $log(t)$ with slope $-\kappa$. To validate the convergence rate, we show the figure of $log(Regret)$ as a function of $log(t)$ in Figure 2, which confirms our theoretical results.

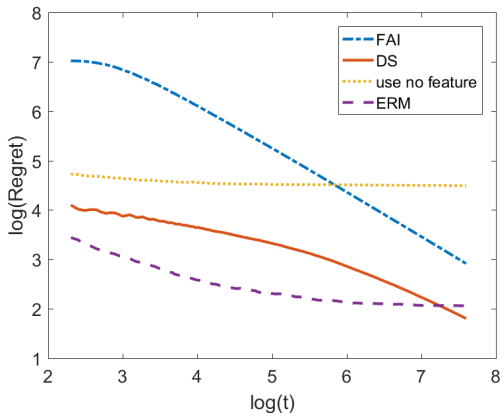


Figure 2: Performance of algorithms

Figure 2 shows that the DS algorithm outperforms the FAI algorithm significantly in the beginning periods. We have analyzed the performance of the case using no feature data and the case using all features in Figure 1. The good performance of the DS approach shows that the reduction of the volatility of the gradients is effective. In the initial periods, we adjust the gradient in a relative small range by using a dynamic shrinkage factor, and in this way, we efficiently decrease the cost loss due to fluctuant order quantities. As t increases, the FAI approach and the DS approach will approach the same convergence rate since the shrinkage parameter goes to 1. The ERM approach is the uncensored demand benchmark, where we observe the realization of uncensored demand in each period. The uncensored demand provides more information, so the ERM approach performs

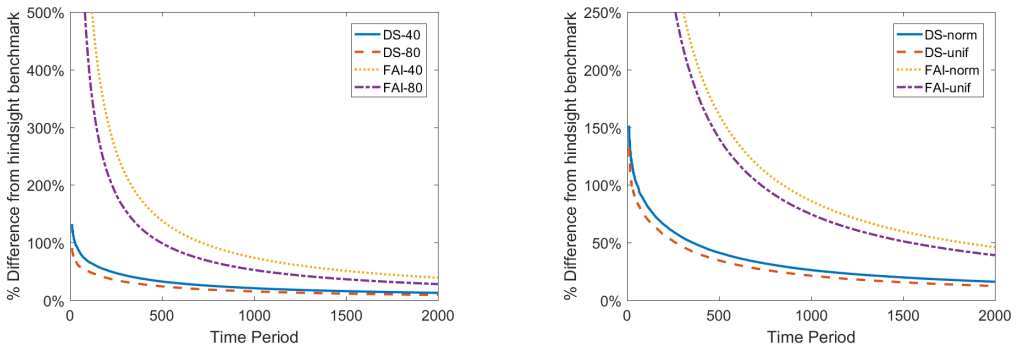
best initially. However, as the regularization parameter of the ERM approach is a constant, not decreasing in t, the existence of the regularization term leads to the biased solution, which causes the ERM to perform worse than the DS approach eventually.

6.2 Performance on Other Parameters

We investigate the influence of demand variability on the performance of the FAI algorithm and the DS algorithm. We first consider that the error term δ is normally distributed, where the mean is 0 and the standard deviation is 40 and 80, respectively. In Figure 3(a), the figure of the legend represents the standard deviation of δ. It shows that as the variance increases, the regret converges to zero faster under both algorithms, which indicates that when the random part accounts more, the first term dominates and the importance of the features decreases. We again consider the setting where there is a normal distribution with mean 0 and standard variance 40 and a uniform distribution in $[-70, 70]$. Figure 3(b) shows that with uniform distribution, the regret converges slightly faster than it does with normal distribution. It is similar to the first case. Since δ in uniform distribution is more diffuse due to the characteristics of distribution (PDF), the influence of the first term is larger than the normal distribution case. In other words, if the random part of demand is large enough, it is similar to the case with no features, since the influence of features on demand is relatively small and can be neglected. The only term which needs to learn is the newsvendor fractile of δ distribution. Thus, the clairvoyant optimal order quantities for different features are close, and the performance is good even though it is not the optimal solution w^*. On the other hand, the small standard deviation of δ indicates that the clairvoyant optimal quantity would be very close to the true demand (demand realization) due to the small randomness. Thus, in this case, each component of the solution needs to be close to the optimal solution w^*, rather than the first term. It takes a longer time to learn the impact of each feature.

7 Conclusion

We investigated two feature-based nonparametric algorithms for stochastic inventory systems when the DM has historical sales data and feature information about demand. Our paper is the first to present the convergence rate for an inventory problem with features when the demand is censored. The algorithms are based on the results of online convex optimization and are applicable for both perishable and nonperishable inventory cases. Our algorithm aims to determine the order-up-to



a b

Figure 3: average cost with varying standard deviations

inventory level conditioning on features. We establish the convergence rate, and the key challenge in our analysis is to derive the upper bound of the distance between the target order-up-to level and the actual implemented order-up-to level. To overcome the difficulty associated with changing feature values, we borrow the property of strong convexity to build the convergence of solutions. Motivated by this numerical result that order quantities fluctuate dramatically at the beginning due to the volatility of the gradients, we propose another algorithm which uses a shrinkage parameter to adjust the gradients and significantly improves the initial performance. We believe the approach used in this paper will lead to more applications to address operations problems.

References

Ban, G.-Y. (2020). Confidence intervals for data-driven inventory policies with demand censoring. *Operations Research*, 68(2):309–326.

Ban, G.-Y. and Keskin, N. B. (2020). Personalized dynamic pricing with machine learning: High dimensional features and heterogeneous elasticity. *Forthcoming, Management Science*.

Ban, G.-Y. and Rudin, C. (2019). The big data newsvendor: Practical insights from machine learning. *Operations Research*, 67(1):90–108.

Bastani, H. and Bayati, M. (2020). Online decision making with high-dimensional covariates. *Operations Research*, 68(1):276–294.

Bertsimas, D. and Kallus, N. (2019). From predictive to prescriptive analytics. *Management Science*, 66(3):1025–1044.

Bertsimas, D. and Thiele, A. (2006). A robust optimization approach to inventory theory. *Operations Research*, 54(1):150–168.

Bookbinder, J. H. and Lordahl, A. E. (1989). Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions*, 21(4):302–312.

Boyd, S., Boyd, S. P., and Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

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

Chen, B., Chao, X., and Ahn, H.-S. (2019). Coordinating pricing and inventory replenishment with nonparametric demand learning. *Operations Research*, 67(4):1035–1052.

Chen, W., Shi, C., and Duenyas, I. (2020). Optimal learning algorithms for stochastic inventory systems with random capacities. *Production and Operations Management*, 29(7):1624–1649.

Cheung, W. C. and Simchi-Levi, D. (2019). Sampling-based approximation schemes for capacitated stochastic inventory control models. *Mathematics of Operations Research*, 44(2):668–692.

Chu, L. Y., Shanthikumar, J. G., and Shen, Z.-J. M. (2008). Solving operational statistics via a Bayesian analysis. *Operations Research Letters*, 36(1):110–116.

Feng, Q. and Shanthikumar, J. G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, 27(9):1670–1684.

Ferreira, K. J., Lee, B. H. A., and Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88.

Flaxman, A. D., Kalai, A. T., and Mcmahan, H. B. (2005). Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proc. 16th Annual ACM-SIAM Sympos. Discrete Algorithms*, pages 385–394, Vancouver, British Columbia, Canada.

Godfrey, G. A. and Powell, W. B. (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.

Hannah, L., Powell, W., and Blei, D. (2010). Nonparametric density estimation for stochastic optimization with an observable state variable. *Advances in Neural Information Processing Systems*, 23:820–828.

Hazan, E. (2016). Introduction to online convex optimization. *Foundations and Trends in Optimization*, 2(3-4):157–325.

Hazan, E., Kalai, A., Kale, S., and Agarwal, A. (2006). Logarithmic regret algorithms for online convex optimization. In *International Conference on Computational Learning Theory*, pages 499–513. Springer.

Hoerl, A. E. and Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55–67.

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

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

Huh, W. T. and Rusmevichientong, P. (2014). Online sequential optimization with biased gradients: Theory and applications to censored demand. *INFORMS Journal on Computing*, 26(1):150–159.

James, W. and Stein, C. (1992). Estimation with quadratic loss. In *Breakthroughs in Statistics*, pages 443–460. Springer.

Kleinberg, R. D. (2004). Nearly tight bounds for the continuum-armed bandit problem. *Advances in Neural Information Processing Systems*, 17:697–704.

Levi, R., Perakis, G., and Uichanco, J. (2015). The data-driven newsvendor problem: New bounds and insights. *Operations Research*, 63(6):1294–1306.

Levi, R., Roundy, R. O., and Shmoys, D. B. (2007). Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research*, 32(4):821–839.

Liyanage, L. H. and Shanthikumar, J. G. (2005). A practical inventory control policy using operational statistics. *Operations Research Letters*, 33(4):341–348.

Mamani, H., Nassiri, S., and Wagner, M. R. (2017). Closed-form solutions for robust inventory management. *Management Science*, 63(5):1625–1643.

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

See, C.-T. and Sim, M. (2010). Robust approximation to multiperiod inventory management. *Operations Research*, 58(3):583–594.

Shalev-Shwartz, S. (2012). Online learning and online convex optimization. *Foundations and Trends in Machine Learning*, 4(2):107–194.

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

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Yuan, H., Luo, Q., and Shi, C. (2019). Marrying stochastic gradient descent with bandits: Learning algorithms for inventory systems with fixed costs. *Working paper*.

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

Zhang, H., Chao, X., and Shi, C. (2020). Closing the gap: A learning algorithm for lost-sales inventory systems with lead times. *Management Science*, 66(5):1962–1980.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, pages 928–936.

Appendix

Proof of Proposition 1 First, consider myopic solution $\arg\min_{y_t} \Pi_t(y_t)$. The myopic order-up-to level is the minimizer of $\Pi_t(y_t)$. By setting $\Pi_t'(y_t) = 0$, we obtain that the order-up-to level is given by

$$S_t^* = w \cdot x_t + \Phi^{-1}(\frac{b}{b+h}) = w^* \cdot x_t = y_t^*,$$

where the second last equality holds since the first component of x_t is 1, and the last equality holds from the definition of y_t^*. Without initial inventory constraints, we can always attain the target order-up-to level, and the total cost over T periods is a lower bound of the dynamic programming value given in Equation (2):

$$c_1(u_1) \geq \sum_{t=1}^{T} \Pi_t(y_t^*). \tag{16}$$

Next, we claim that in each period t, the myopic order-up-to level y_t^* can be reached, i.e., $y_t^* \geq u_t$. We use induction to prove this claim. Consider period $t = 1$. Since $u_1 = 0$, y_1^* can be achieved. Assume $y_t^* \geq u_t$ holds in period t and we order up to y_t^* in period t. Let us consider period $t+1$. It can verify that

$$\begin{aligned} y_{t+1}^* - u_{t+1} &= y_{t+1}^* - (y_t^* - D_t)^+ \\ &= y_{t+1}^* - [\Phi^{-1}(\frac{b}{b+h}) - \delta_t]^+. \end{aligned}$$

If $\Phi^{-1}(\frac{b}{b+h}) - \delta_t \leq 0$, then $y_{t+1}^* - u_{t+1} = y_{t+1}^* > 0$, the claim holds. If $\Phi^{-1}(\frac{b}{b+h}) - \delta_t > 0$, we have $y_{t+1}^* - u_{t+1} = y_{t+1}^* - (\Phi^{-1}(\frac{b}{b+h}) - \delta_t) = w \cdot x_{t+1} + \delta_t$. By *Assumption* (f) that states that demands are always positive, $w \cdot x_{t+1} + \delta_t > 0$, then $y_{t+1}^* - u_{t+1} > 0$. Therefore, we prove that in each period t, $y_t^* \geq u_t$ holds. The myopic order-up-to level y_t^* is achievable. From Equation (16), we have $c_1(u_1) = \sum_{t=1}^{T} \Pi_t(y_t^*)$, and thus y_t^* is the optimal order-up-to level in period t.

Proof of Lemma 3. By the definition of $\Psi_t(\cdot)$ and D_t, we have the following equality:

$$\begin{aligned} \Psi_t(z_t) &= hE[(z_t \cdot x_t - D_t)^+] + bE[(D_t - z_t \cdot x_t)^+] \\ &= hE[(z_t \cdot x_t - w \cdot x_t - \delta_t)^+] + bE[(w \cdot x_t + \delta_t - z_t \cdot x_t)^+]. \\ &= h \int_{-\infty}^{z_t \cdot x_t - w \cdot x_t} (z_t \cdot x_t - w \cdot x_t - u)\phi(u)du \\ &\quad + b \int_{z_t \cdot x_t - w \cdot x_t}^{\infty} (u + w \cdot x_t - z_t \cdot x_t)\phi(u)du. \end{aligned}$$

Then, note z_t is a vector and the gradient of the cost function is:

$$\begin{aligned}\nabla\Psi_t(z_t) &= hx_t\int_{-\infty}^{z_t\cdot x_t - w\cdot x_t}\phi(u)du - bx_t\int_{z_t\cdot x_t - w\cdot x_t}^{\infty}\phi(u)du \\ &= (h+b)x_t\Phi(z_t\cdot x_t - w\cdot x_t) - bx_t \\ &= [(h+b)\Phi(z_t\cdot x_t - w\cdot x_t) - b]x_t.\end{aligned}$$

For any pair of vectors $z_1, z_2 \in \Omega$, we have

$$\begin{aligned}(\nabla\Psi_t(z_1) - \nabla\Psi_t(z_2))\cdot(z_1 - z_2) &= (h+b)\left[\Phi(z_1\cdot x_t - w\cdot x_t) - \Phi(z_2\cdot x_t - w\cdot x_t)\right]x_t\cdot(z_1 - z_2) \\ &= \left[(h+b)\int_{z_2\cdot x_t - w\cdot x_t}^{z_1\cdot x_t - w\cdot x_t}\phi(u)du\right]x_t\cdot(z_1 - z_2) \\ &> (h+b)\theta\|x_t\|^2\|z_1 - z_2\|^2 \\ &\geq \mu\|z_1 - z_2\|^2.\end{aligned}$$

The first inequality holds from *Assumptions* (d) and (e), since $\phi(\delta_t) > \theta$ for any $\delta_t \in [\underline{\delta}, \overline{\delta}]$, and $z_1\cdot x_t - w\cdot x_t$, $z_2\cdot x_t - w\cdot x_t$ are both in the region $[\underline{\delta}, \overline{\delta}]$. Therefore, the integration is larger than $\theta x_t\cdot(z_1 - z_t)$. Since the first component of x_t is the constant 1, we have $\|x_t\|^2 \geq 1$, which implies that the second inequality holds since μ satisfies $0 \leq \mu \leq (h+b)\theta$.

Therefore, the above inequality shows that Ψ_t is strongly convex in $z_t \in \Omega$ by Boyd et al. (2004). □

Proof of Lemma 4. Applying the definition of strong convexity to the pair of points v_t, v^*, we have

$$2E[f(v_t) - f(v^*)] \leq 2E[H(v_t)\cdot(v_t - v^*)] - \alpha E\|v^* - v_t\|^2. \tag{17}$$

Using the update rule for v_{t+1} (Equation (8)), and since the projection operator will not increase the distance between two points (Pythagorean theorem), we get

$$\begin{aligned}E\|v_{t+1} - v^*\|^2 &= E\|P_S(v_t - \eta_t H(v_t)) - v^*\|^2 \\ &\leq E\|v_t - \eta_t H(v_t) - v^*\|^2 \\ &= E\|(v_t - v^*) - \eta_t H(v_t)\|^2 \\ &= E\|v_t - v^*\|^2 + \eta_t^2 E\|H(v_t)\|^2 - 2\eta_t E[H(v_t)\cdot(v_t - v^*)].\end{aligned}$$

Hence,

$$2E[H(v_t)\cdot(v_t - v^*)] \leq \frac{E\|v_t - v^*\|^2 - E\|v_{t+1} - v^*\|^2}{\eta_t} + \eta_t E\|H(v_t)\|^2. \tag{18}$$

Summing up Equation (18) from $t = 1$ to T, with $\eta_t = \frac{1}{\alpha t}$, and combining it with Equation (17), we have:

$$\begin{aligned}
2\sum_{t=1}^{T} E[f(v_t) - f(v^*)] &\leq \sum_{t=1}^{T}\{\frac{E\|v_t - v^*\|^2 - E\|v_{t+1} - v^*\|^2}{\eta_t} + \eta_t E\|H(v_t)\|^2 - \alpha E\|v_t - v^*\|^2\} \\
&\leq \sum_{t=1}^{T} E\|v_t - v^*\|^2(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} - \alpha) + \overline{G}^2\sum_{t=1}^{T}\eta_t \\
&\leq \overline{G}^2\sum_{t=1}^{T}\eta_t = \overline{G}^2\sum_{t=1}^{T}\frac{1}{\alpha t} \\
&\leq \frac{\overline{G}^2}{\alpha}(1 + \log T).
\end{aligned}$$

Thus, $E[\frac{1}{T}\sum_{t=1}^{T}(f(v_t) - f(v^*))] \leq \frac{\overline{G}^2}{2\alpha T}(1 + \log T)$. □

Proof of Lemma 5. This proof stems from the class notes by Ji Liu (2015),[5] where the SGD updating rule we use needs to be projected to the feasible region S, but he does this without the constraint. According to the strong convexity property, we have

$$f(a^*) - f(a_k) \geq \nabla f(a_k)\cdot(a^* - a_k) + \frac{\alpha}{2}\|a_k - a^*\|^2$$

$$f(a_k) - f(a^*) \geq \nabla f(a^*)\cdot(a_k - a^*) + \frac{\alpha}{2}\|a_k - a^*\|^2.$$

Summing up the above inequalities gives

$$(\nabla f(a_k) - \nabla f(a^*))\cdot(a_k - a^*) = \nabla f(a_k)\cdot(a_k - a^*) \geq \alpha\|a_k - a^*\|^2. \tag{19}$$

Also, we have

$$\begin{aligned}
\mathbb{E}(\|a_{k+1} - a^*\|^2) &= \mathbb{E}(\|P_S(a_k - \gamma_k G_k) - a^*\|^2) \\
&\leq \mathbb{E}(\|a_k - a^*\|^2) - 2\gamma_k\mathbb{E}[G_k\cdot(a_k - a^*)] + {\gamma_k}^2\mathbb{E}(\|G_k\|^2) \\
&\leq \mathbb{E}(\|a_k - a^*\|^2) - 2\gamma_k\mathbb{E}[\nabla f(a_k)\cdot(a_k - a^*)] + {\gamma_k}^2\overline{G}_0^2.
\end{aligned}$$

Applying Equation (19), it follows

$$\begin{aligned}
\mathbb{E}(\|a_{k+1} - a^*\|^2) &\leq \mathbb{E}(\|a_k - a^*\|^2) - 2\alpha\gamma_k\mathbb{E}(\|a_k - a^*\|^2) + {\gamma_k}^2\overline{G}_0^2 \\
&= (1 - 2\alpha\gamma_k)\mathbb{E}(\|a_k - a^*\|^2) + {\gamma_k}^2\overline{G}_0^2.
\end{aligned} \tag{20}$$

[5] https://www.cs.rochester.edu/u/jliu/CSC-576/class-note-10.pdf

We prove the convergence rate by induction. First, it is easy to see that

$$\|a_1 - a^*\|^2 \leq \frac{\max\{\|a_1 - a^*\|^2, \frac{\overline{G}_0^2}{\alpha^2}\}}{1}$$

Then, we assume that the convergence rate holds with k. Next, we show that it holds with $k+1$. Denote $L = \max\{\|a_1 - a^*\|^2, \frac{\overline{G}_0^2}{\alpha^2}\}$. From Equation (20), we have

$$\begin{aligned}
\mathbb{E}(\|a_{k+1} - a^*\|^2) &\leq (1 - \frac{2}{k})\mathbb{E}(\|a_k - a^*\|^2) + \frac{1}{\alpha^2 k^2}\overline{G}_0^2 \\
&\leq (1 - \frac{2}{k})\frac{L}{k} + \frac{1}{\alpha^2 k^2}\overline{G}_0^2 \\
&\leq (\frac{1}{k} - \frac{2}{k^2})L + \frac{L}{k^2} \\
&= (\frac{1}{k} - \frac{1}{k^2})L \\
&\leq \frac{L}{k+1}
\end{aligned}$$

This completes the proof of Lemma 5.

Proof of Theorem 7. Following the structure of the proof of Theorem 2, we consider $\Lambda_1(T)$ and $\Lambda_2(T)$, respectively. Two ingredients which lead to Theorem 2 are Lemmas 4 and 5. In the following, we show that the variants of these two lemmas also hold under the DS algorithm. We remark that under the FAI algorithm, Lemma 4 was proved directly without the result of the solution convergence rate (corresponding to Lemma 5). However, as β_t is introduced, we need to apply this result to prove the variant of Lemma 4. Thus, we first show the variant of Lemma 5 that the convergence rate of solutions holds under the DS algorithm. We introduce a new function $g(t)$, where $g(t) \triangleq 2\beta_t - \frac{t}{t+1} = 2 - 2e^{-\lambda t} - \frac{t}{t+1}$. There exists $\tilde{t} > 0$, such that $g(t)$ is increasing in $t \geq \tilde{t}$ and $g(\tilde{t}) > 0$. Define $\bar{g} = g(\tilde{t})$. We state a variant of Lemma 5: for any t, we have

$$\|z_t - w^*\|^2 \leq \frac{\max\{\xi, \frac{B}{\mu^2 \bar{g}}\}}{t}, \tag{21}$$

where $\xi \triangleq \tilde{t} \max_{a_1, a_2 \in \Omega} \|a_1 - a_2\|^2$. Denote $L = \max\{\xi, \frac{B}{\mu^2 \bar{g}}\}$.

The proof is similar to Lemma 5. For any $t \leq \tilde{t}$, it is easy to see that Equation (21) holds.

Suppose $t > \tilde{t}$, then we proceed by induction. From Equation (15), we have

$$\begin{aligned} E\|z_{t+1} - w^*\|^2 &\leq E\|z_t - w^*\|^2 - 2\varepsilon_t\beta_t E[H(z_t)\cdot(z_t - w^*)] + \varepsilon_t^2 B \\ &\leq (1 - 2\mu\varepsilon_t\beta_t)E\|z_t - w^*\|^2 + \varepsilon_t^2 B \\ &= (1 - \frac{2\beta_t}{t})E\|z_t - w^*\|^2 + \frac{B}{\mu^2 t^2} \\ &\leq (1 - \frac{2\beta_t}{t})\frac{L}{t} + \frac{\bar{g}}{t^2}L \\ &\leq (\frac{1}{t} + \frac{g(t) - 2\beta_t}{t^2})L \\ &= (\frac{1}{t} - \frac{1}{t(t+1)})L = \frac{L}{t+1}. \end{aligned}$$

The second inequality uses the property of strong convexity, where we apply Equation (19) and Lemma 3 that $\Psi_t(\cdot)$ is μ-strongly convex. The next equality follows from the definition of ε_t. The subsequent inequality holds due to the result in period t by induction, as well as the definition of L. The second last equality comes from the definition of $g(t)$. This completes the induction step. Thus, we prove the variant of Lemma 5 that

$$E\|z_t - w^*\|^2 \leq \frac{L}{t}. \tag{22}$$

Then, we need to modify Lemma 4 under the DS algorithm to bound $\Lambda_1(T)$. Applying the definition of strong convexity to the pair of points z_t and w^*, we have

$$2E_{z_t}[\Phi_t(z_t) - \Phi_t(w^*)] \leq 2E_{z_t}[H(z_t)\cdot(z_t - w^*)] - \mu E_{z_t}\|w^* - z_t\|^2. \tag{23}$$

From Equation (15), we get

$$2E[H(z_t)\cdot(z_t - w^*)] \leq \frac{E\|z_t - w^*\|^2 - E\|z_{t+1} - w^*\|^2}{\varepsilon_t\beta_t} + \frac{\varepsilon_t B}{\beta_t}. \tag{24}$$

Summing Equation (24) from $t = 1$ to T and combining Equation (23), we have

$$\begin{aligned} 2\sum_{t=1}^{T} E[\Phi(z_t) - \Phi(w^*)] &\leq \sum_{t=1}^{T}\{\frac{E\|z_t - w^*\|^2 - E\|z_{t+1} - w^*\|^2}{\varepsilon_t\beta_t} + \frac{\varepsilon_t}{\beta_t}B - \mu E\|z_t - w^*\|^2\} \\ &\leq \sum_{t=1}^{T} E\|z_t - w^*\|^2(\frac{1}{\varepsilon_t\beta_t} - \frac{1}{\varepsilon_{t-1}\beta_{t-1}} - \mu) + B\sum_{t=1}^{T}\frac{1}{\mu\beta_t t} \\ &\leq (\frac{1}{\beta_1} - 1)\mu L\sum_{t=1}^{T}\frac{1}{t} + \frac{B}{\mu\beta_1}\sum_{t=1}^{T}\frac{1}{t} \\ &\leq [(\frac{1}{\beta_1} - 1)\mu L + \frac{B}{\mu\beta_1}](1 + \log T). \end{aligned} \tag{25}$$

The third inequality holds due to Equation (22) and $\varepsilon_t = \frac{1}{\mu t}$, where

$$\begin{aligned}\frac{1}{\varepsilon_t\beta_t} - \frac{1}{\varepsilon_{t-1}\beta_{t-1}} - \mu &= \frac{\mu t}{\beta_t} - \frac{\mu(t-1)}{\beta_{t-1}} - \mu \\ &< \frac{\mu t}{\beta_t} - \frac{\mu(t-1)}{\beta_t} - \mu \\ &= \frac{\mu}{\beta_t} - \mu < (\frac{1}{\beta_1} - 1)\mu.\end{aligned}$$

According to Equation (25), we show that $\Lambda_1(T)$ is $O(\log T/T)$. This completes the proof of Theorem 7(a), since in the perishable inventory case, $\Lambda_2(T)$ is always zero. Similarly, Huh and Rusmevichientong (2014) extended the classical result in online convex optimization to allow for biased gradient estimators. They considered a setting of sequentially convex functions where an unbiased estimate of the gradient is unavailable. The biased gradients are due to the properties of the function naturally. However, we take the initiative to introduce biased gradients to improve practical performance.

To complete the proof of Theorem 7(b), we consider $\Lambda_2(T)$. With Equation (22), the proof of Theorem 6 can go through by replacing $H_t(z_t)$ with $\tilde{H}_t(z_t)$ (see Section 4.2.1). Thus, we still have

$$\Lambda_2(T) \leq \frac{14h\rho\pi^2 E[(B_1 - E[B_1])^6]}{(E[B_1] - \rho)^6\sqrt{T}}.$$

The proof of Theorem 7(b) is completed. □