

Using Stochastic Approximation Methods to Compute Optimal Base-Stock Levels in Inventory Control Problems

Sumit Kunnumkal

School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
sumit@orie.cornell.edu

Huseyin Topaloglu

School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
topaloglu@orie.cornell.edu

May 18, 2007

Abstract

In this paper, we consider numerous inventory control problems for which the base-stock policies are known to be optimal and we propose stochastic approximation methods to compute the optimal base-stock levels. The existing stochastic approximation methods in the literature guarantee that their iterates converge, but not necessarily to the optimal base-stock levels. In contrast, we prove that the iterates of our methods converge to the optimal base-stock levels. Moreover, our methods continue to enjoy the well-known advantages of the existing stochastic approximation methods. In particular, they only require the ability to obtain samples of the demand random variables, rather than to compute expectations explicitly and they are applicable even when the demand information is censored by the amount of available inventory.

1 INTRODUCTION

One approach for finding good solutions to stochastic optimization problems is to concentrate on a class of policies that are characterized by a number of parameters and to find a good set of values for these parameters by using stochastic approximation methods. This approach is quite flexible. We only need a sensible guess at the form of a good policy and stochastic approximation methods allow us to work with samples of the underlying random variables, rather than to compute expectations. Consequently, parameterized policies along with stochastic approximation methods are widely used in practice.

In this paper, we analyze stochastic approximation methods for several inventory control problems for which the base-stock policies are known to be optimal. For these problems, there exist base-stock levels $\{r_1^*, \dots, r_\tau^*\}$ such that it is optimal to keep the inventory position at time period t as close as possible to r_t^* . In other words, letting x_t be the inventory position at time period t and $[x]^+ = \max\{x, 0\}$, it is optimal to order $[r_t^* - x_t]^+$ units of inventory at time period t . This particular structure of the optimal policy generally arises from the fact that the value functions in the dynamic programming formulations of these problems are convex in the inventory position. In this case, the computation of the optimal base-stock levels through the Bellman equations requires solving a number of convex optimization problems.

On the other hand, we lose the appealing structure of the Bellman equations when we try to compute the optimal base-stock levels by using the existing stochastic approximation methods in the literature. As a result, the existing stochastic approximation methods can only guarantee that their iterates converge, but not necessarily to the optimal base-stock levels. Our main goal is to develop stochastic approximation methods that can indeed compute the optimal base-stock levels.

To illustrate the difficulties, we consider a two-period newsvendor problem with backlogged demands, zero lead times for the replenishments, and linear holding and backlogging costs. For this problem, it is known that the base-stock policies are optimal under fairly general assumptions. Assuming that the purchasing cost is zero and the initial inventory position is x_1 , the total expected cost incurred by a base-stock policy characterized by the base-stock levels $\{r_1, r_2\}$ can be written as

$$g(x_1, r_1, r_2) = h \mathbb{E} \left\{ [(x_1 \vee r_1) - d_1]^+ + [\max\{(x_1 \vee r_1) - d_1, r_2\} - d_2]^+ \right\} \\ + b \mathbb{E} \left\{ [d_1 - (x_1 \vee r_1)]^+ + [d_2 - \max\{(x_1 \vee r_1) - d_1, r_2\}]^+ \right\},$$

where $\{d_1, d_2\}$ are the demand random variables in the two time periods, h is the per unit holding cost and b is the per unit backlogging cost, and we let $x \vee y = \max\{x, y\}$. Since the inventory position after the replenishment decision at the first time period is $x_1 \vee r_1$ and the inventory position after the replenishment decision at the second time period is $\max\{(x_1 \vee r_1) - d_1, r_2\}$, the two expectations above respectively compute the total expected holding and backlogging costs. In this case, the optimal base-stock levels can be found by solving the problem

$$(r_1^*, r_2^*) = \underset{(r_1, r_2)}{\operatorname{argmin}} g(x_1, r_1, r_2). \tag{1}$$

One approach to solve this problem is to use stochastic gradients of $g(x_1, \cdot, \cdot)$ to iteratively search for a good set of base-stock levels. Under certain assumptions, it is possible to show that the iterates of

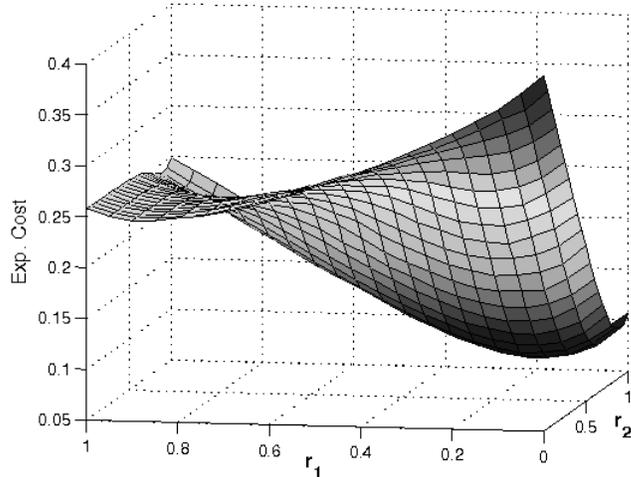


Figure 1: Total expected cost as a function of the base-stock levels for a two-period newsvendor problem. The problem parameters are $x_1 = 0$, $h = 0.25$, $b = 0.4$, $d_1 \sim \text{beta}(1, 5)$, $d_2 \sim \text{beta}(5, 1)$.

such a stochastic approximation method converge to a stationary point of $g(x_1, \cdot, \cdot)$ with probability 1 (w.p.1). However, $g(x_1, \cdot, \cdot)$ is not necessarily a convex function. In particular, a stationary point of $g(x_1, \cdot, \cdot)$ may not be an optimal solution to problem (1) and the solution obtained by a stochastic approximation method may not be very good. For example, Figure 1 shows the plot of $g(x_1, \cdot, \cdot)$ for a particular problem instance where $g(x_1, \cdot, \cdot)$ is not convex.

This is a rather surprising observation. If we assume nothing about the structure of the optimal policy and compute it through the Bellman equations, then the problem is “well-behaving” in the sense that all we need to do is to solve a number of convex optimization problems. On the other hand, if we exploit the information that the base-stock policies are optimal and use stochastic approximation methods to solve problem (1), then we can only obtain a stationary point of $g(x_1, \cdot, \cdot)$.

In this paper, we mainly consider variants of the multi-period newsvendor problem for which the base-stock policies are known to be optimal. Nevertheless, our results are fairly general and they can be applied on other problem classes whose optimal policies are characterized by a finite number of base-stock levels. To illustrate this point, we also consider a relatively nonstandard inventory purchasing problem where the price of the product changes randomly over time and we have to purchase a certain amount of product to satisfy the random demand that occurs at the end of the planning horizon. Although the problems that we work with are well-studied, our paper makes several substantial contributions. Most importantly, we offer a remedy for the aforementioned surprising observation by showing that it is indeed possible to compute the optimal base-stock levels through stochastic approximation methods. Apart from its theoretical value, this result allows us to exploit the well-known advantages of stochastic approximation methods when computing the optimal base-stock levels.

A primary advantage of stochastic approximation methods is that they allow working directly with the samples of the demand random variables, rather than the full demand distributions. In contrast,

using conventional inventory control models typically requires three steps. First, the historical demand data are collected. If the historical demand data include only the amount of inventory sold but not the amount of demand, then we have a situation where the demand information is censored by the amount of available inventory, and the historical demand data have to be “uncensored” to have access to the samples of the demand random variables. After this, parametric forms for the demand distributions are hypothesized and the parameters are fitted by using the historical demand data. Finally, the optimal base-stock levels are computed under the assumption that the fitted demand distributions characterize the demand random variables. In practice, the historical demand data are often “uncensored” by using heuristic approaches. Furthermore, the hypothesized forms for the demand distributions usually do not characterize the demand random variables accurately, causing errors just because wrong distributions are hypothesized to begin with. On the other hand, stochastic approximation methods work directly with the amount of inventory sold, rather than the amount of demand. Therefore, they do not require “uncensoring” the historical demand data. Also, since stochastic approximation methods work directly with the samples, they do not require hypothesizing parametric forms for the demand distributions.

The advantages mentioned in the previous paragraph unfortunately come at a cost. One difficulty with stochastic approximation methods is the choice of the step size parameters. In general, choosing the step size parameters requires some experimentation, and there are no hard and fast rules for making the choice. Although this difficulty is always a major concern, our stochastic approximation methods appear to be relatively robust to the choice of the step size parameters. In particular, we work with many different problem classes, demand distributions and cost parameters in our numerical experiments, but we use the same set of step size parameters throughout. The same set of step size parameters appear to work well in all of our numerical experiments. Another difficulty with stochastic approximation methods is the choice of the initial base-stock levels. A rough observation from our numerical experiments is that if our stochastic approximation methods start with base-stock levels having 80% optimality gap, then they take 10 to 40 iterations to obtain base-stock levels having 10% optimality gap. This performance itself may be satisfactory in certain settings, but we note that this performance is obtained without exploiting prior information about the demand distributions. In practice, we usually use some prior information to choose better initial base-stock levels and the role of stochastic approximation methods becomes that of only fine-tuning the base-stock levels by using the demand samples.

We also note that even if we have access to the demand distributions, numerically solving the Bellman equations requires discretization when the demand distributions are continuous. Under reasonable assumptions, the base-stock levels obtained by discretizing the demand distributions converge to the optimal base-stock levels as the discretization becomes finer, but our stochastic approximation methods can be used as an alternative method for computing the optimal base-stock levels.

The remainder of the paper is organized as follows. Section 2 briefly reviews the related literature. Sections 3 and 4 consider the multi-period newsvendor problem respectively with backlogged demands and lost sales, and develop stochastic approximation methods to compute the optimal base-stock levels. Section 5 shows that the proposed stochastic approximation methods are applicable when the demand information is censored. Section 6 develops a stochastic approximation method for an inventory pur-

chasing problem where we make purchasing decisions for a product whose price changes randomly over time and we use the product to satisfy the random demand at the end of the planning horizon. Section 7 presents numerical experiments.

2 RELEVANT LITERATURE

In this paper, we mainly consider the multi-period newsvendor problem with backlogged demands or lost sales. The problem involves controlling the inventory of a perishable (or fashion) product over a finite number of time periods. We have multiple opportunities to replenish the inventory of the product before the product reaches the end of its useful life. A classical example is controlling the inventory of a monthly magazine. We are allowed to replenish the magazines multiple times during the course of a month, but the left over magazines at the end of a month are discarded, possibly at a salvage value. For the multi-period newsvendor problem with lost sales, we assume that the lead times for the replenishments are zero. All cost functions we deal with are linear, although generalizations to convex cost functions are possible. The optimality of the base-stock policies for the variants of the multi-period newsvendor problem that we consider is well-known; see Arrow, Karlin and Scarf (1958), Porteus (1990) and Zipkin (2000). If the distribution of the demand is known, then the optimal base-stock levels can be computed through the Bellman equations.

Significant literature has evolved around the newsvendor problem under the assumption that the distribution of the demand is unknown. There may be different reasons for employing such an assumption. For example, we may not have enough data to fit a parametric demand distribution or it may be difficult to collect demand data since we are only able to observe the amount of inventory sold, but not the amount of demand. Scarf (1960), Iglehart (1964) and Azoury (1985) use Bayesian framework to estimate the demand parameters and to adaptively update the replenishment quantities as the demand information becomes available. Levi, Roundy and Shmoys (2005) provide bounds on how many demand samples are needed to obtain near-optimal base-stock levels with high probability. Conrad (1976), Braden and Freimer (1991) and Ding (2002) focus on the case where the demand information is censored by the amount of available inventory. Godfrey and Powell (2001) give a nice overview of the newsvendor problem with censored demands.

Stochastic approximation methods can deal with the uncertainty in the distribution of the demand and the censored demand information. They only require the ability to obtain samples from the demand distributions. Furthermore, they usually do not require having access to the exact values of the demand samples. Instead, only knowing the amount of inventory sold is often adequate. Consequently, stochastic approximation methods can be used under the assumption that a parametric form for the demand distribution is not available or the demand information is censored by the inventory availability.

The use of stochastic approximation methods for solving stochastic optimization problems is well-known. Kushner and Clark (1978) and Bertsekas and Tsitsiklis (1996) give a detailed coverage of stochastic approximation methods. As far as the applications are concerned, L'Ecuyer and Glynn (1994), Fu (1994), Glasserman and Tayur (1995), Bashyam and Fu (1998), Mahajan and van Ryzin (2001),

Karaesmen and van Ryzin (2004) and van Ryzin and Vulcano (2006) focus on queueing, inventory control and revenue management settings. Although the objective functions that are considered in many of these papers are not convex and we can only guarantee convergence to the stationary points, computational experience indicates that stochastic approximation methods provide good solutions in practice; see Mahajan and van Ryzin (2001) and van Ryzin and Vulcano (2006).

The traditional approach in the stochastic approximation literature is to concentrate on a class of policies that are characterized by a number of parameters. The hope is that this class of policies contain at least one good policy for the problem. In contrast, there are numerous reinforcement learning methods that try to avoid this shortcoming by explicitly approximating the value functions in the dynamic programming formulation of the problem. Q-learning and temporal differences learning use sampled state trajectories to approximate the value functions in problems with discrete state and decision spaces; see Sutton (1988) and Tsitsiklis (1994). Godfrey and Powell (2001), Topaloglu and Powell (2003) and Powell, Ruszczyński and Topaloglu (2004) propose sampling-based methods to approximate piecewise-linear convex value functions and these methods are convergent for certain stationary problems.

The stochastic approximation methods that we propose in this paper embody the characteristics of the two types of approaches mentioned in the last two paragraphs. Similar to the standard stochastic approximation methods, we concentrate on the class of policies that are characterized by a finite number of base-stock levels, whereas similar to the value function approximation methods, we work with the dynamic programming formulation of the problem to search for the optimal base-stock levels.

3 MULTI-PERIOD NEWSVENDOR PROBLEM WITH BACKLOGGED DEMANDS

We want to control the inventory of a product over the time periods $\{1, \dots, \tau\}$. At time period t , we observe the inventory position x_t and place a replenishment order of $y_t - x_t$ units, which costs c per unit. The replenishment order arrives instantaneously and raises the inventory position to y_t . Following the arrival of the replenishment, we observe the random demand d_t and satisfy the demand as much as possible. We incur a cost of h per unit of held inventory per time period and a cost of b per unit of unsatisfied demand per time period. We assume that the revenue from the sales is zero without loss of generality. The goal is to minimize the total expected cost over the planning horizon.

Throughout, we assume that the demand random variables $\{d_t : t = 1, \dots, \tau\}$ are independent and have finite expectations, and their cumulative distribution functions are Lipschitz continuous. We assume that the cost parameters satisfy $b > c \geq 0$ and $h \geq 0$. The assumption that the cost parameters are stationary and the lead times for the replenishments are zero is for notational brevity. It is also possible to extend our analysis to the case where the distributions of the demand random variables are discrete. We note that the demand random variables do not have to be identically distributed. We let $v_t(x_t)$ be the minimum total expected cost incurred over the time periods $\{t, \dots, \tau\}$ when the inventory position at time period t is x_t and the optimal policy is followed over the time periods $\{t, \dots, \tau\}$. The functions $\{v_t(\cdot) : t = 1, \dots, \tau\}$ satisfy the Bellman equations

$$v_t(x_t) = \min_{y_t \geq x_t} c[y_t - x_t] + \mathbb{E} \{ h[y_t - d_t]^+ + b[d_t - y_t]^+ + v_{t+1}(y_t - d_t) \}, \quad (2)$$

with $v_{\tau+1}(\cdot) = 0$. If we let

$$f_t(r_t) = c r_t + \mathbb{E}\{h[r_t - d_t]^+ + b[d_t - r_t]^+ + v_{t+1}(r_t - d_t)\}, \quad (3)$$

then it can be shown that $f_t(\cdot)$ is a convex function with a finite unconstrained minimizer, say r_t^* . In this case, it is well-known that the optimal policy is a base-stock policy characterized by the base-stock levels $\{r_t^* : t = 1, \dots, \tau\}$. That is, if the inventory position at time period t is x_t , then it is optimal to order $[r_t^* - x_t]^+$ units. Therefore, we can write (2) as

$$\begin{aligned} v_t(x_t) &= \begin{cases} \mathbb{E}\{h[x_t - d_t]^+ + b[d_t - x_t]^+ + v_{t+1}(x_t - d_t)\} & \text{if } x_t \geq r_t^* \\ c[r_t^* - x_t] + \mathbb{E}\{h[r_t^* - d_t]^+ + b[d_t - r_t^*]^+ + v_{t+1}(r_t^* - d_t)\} & \text{if } x_t < r_t^* \end{cases} \\ &= \begin{cases} f_t(x_t) - c x_t & \text{if } x_t \geq r_t^* \\ f_t(r_t^*) - c x_t & \text{if } x_t < r_t^*. \end{cases} \end{aligned} \quad (4)$$

It can be shown that $f_t(\cdot)$ and $v_t(\cdot)$ are positive, Lipschitz continuous, differentiable and convex functions. We use $\dot{f}_t(\cdot)$ and $\dot{v}_t(\cdot)$ to respectively denote the derivatives of $f_t(\cdot)$ and $v_t(\cdot)$. The following lemma shows that $\dot{f}_t(\cdot)$ and $\dot{v}_t(\cdot)$ are also Lipschitz continuous.

Lemma 1 *There exists a constant L such that we have $|\dot{f}_t(\hat{x}_t) - \dot{f}_t(\tilde{x}_t)| \leq L|\hat{x}_t - \tilde{x}_t|$ and $|\dot{v}_t(\hat{x}_t) - \dot{v}_t(\tilde{x}_t)| \leq L|\hat{x}_t - \tilde{x}_t|$ for all $\hat{x}_t, \tilde{x}_t \in \mathbb{R}$, $t = 1, \dots, \tau$.*

Proof We show the result by induction over the time periods. Since $\dot{v}_{\tau+1}(\cdot) = 0$, this function is Lipschitz continuous. We assume that $\dot{v}_{t+1}(\cdot)$ is Lipschitz continuous. We have

$$\dot{f}_t(x_t) = c + h\mathbb{P}\{d_t < x_t\} - b\mathbb{P}\{d_t \geq x_t\} + \mathbb{E}\{\dot{v}_{t+1}(x_t - d_t)\}, \quad (5)$$

where the interchange of the expectation and the derivative above follows from Lemma 6.3.1 in Glasserman (1994). Since the composition of Lipschitz continuous functions is also Lipschitz continuous by Lemma 6.3.3 in Glasserman (1994), $\dot{f}_t(\cdot)$ is Lipschitz continuous. To see that $\dot{v}_t(\cdot)$ is Lipschitz continuous, we use (4) to obtain

$$\dot{v}_t(x_t) = \begin{cases} \dot{f}_t(x_t) - c & \text{if } x_t \geq r_t^* \\ -c & \text{if } x_t < r_t^*. \end{cases} \quad (6)$$

We assume that $\hat{x}_t \geq \tilde{x}_t$ without loss of generality and consider three cases. First, we assume that $\hat{x}_t \geq r_t^* \geq \tilde{x}_t$. Since r_t^* is the minimizer of $f_t(\cdot)$, we have $\dot{f}_t(r_t^*) = 0$, which implies that

$$|\dot{v}_t(\hat{x}_t) - \dot{v}_t(\tilde{x}_t)| = |\dot{f}_t(\hat{x}_t)| = |\dot{f}_t(\hat{x}_t) - \dot{f}_t(r_t^*)| \leq L|\hat{x}_t - r_t^*| \leq L|\hat{x}_t - \tilde{x}_t|,$$

where we use the Lipschitz continuity of $\dot{f}_t(\cdot)$ in the first inequality. The other two cases where we have $\hat{x}_t \geq \tilde{x}_t > r_t^*$ or $r_t^* > \hat{x}_t \geq \tilde{x}_t$ are easy to show. \square

We now consider computing the optimal base-stock levels $\{r_t^* : t = 1, \dots, \tau\}$ through a stochastic approximation method. Noting (5) and using $\mathbf{1}(\cdot)$ to denote the indicator function, we can compute a stochastic gradient of $f_t(\cdot)$ at x_t through

$$\Delta_t(x_t, d_t) = c + h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t). \quad (7)$$

In this case, letting $\{r_t^k : t = 1, \dots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k , $\{d_t^k : t = 1, \dots, \tau\}$ be the demand random variables at iteration k and α^k be a step size parameter, we can iteratively update the estimates of the optimal base-stock levels through

$$r_t^{k+1} = r_t^k - \alpha^k \Delta_t(r_t^k, d_t^k). \quad (8)$$

However, this approach is clearly not realistic because the computation in (7) requires the knowledge of $\{\dot{v}_t(\cdot) : t = 1, \dots, \tau\}$. The stochastic approximation method we propose is based on constructing tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$.

Since r_t^* is the minimizer of $f_t(\cdot)$, (5) implies that $-c = \dot{f}_t(r_t^*) - c = h\mathbb{P}\{d_t < r_t^*\} - b\mathbb{P}\{d_t \geq r_t^*\} + \mathbb{E}\{\dot{v}_{t+1}(r_t^* - d_t)\}$. Therefore, using (5) and (6), we obtain

$$\dot{v}_t(x_t) = \begin{cases} h\mathbb{P}\{d_t < x_t\} - b\mathbb{P}\{d_t \geq x_t\} + \mathbb{E}\{\dot{v}_{t+1}(x_t - d_t)\} & \text{if } x_t \geq r_t^* \\ h\mathbb{P}\{d_t < r_t^*\} - b\mathbb{P}\{d_t \geq r_t^*\} + \mathbb{E}\{\dot{v}_{t+1}(r_t^* - d_t)\} & \text{if } x_t < r_t^*. \end{cases} \quad (9)$$

From this expression, it is clear that

$$\dot{v}_t(x_t, d_t) = \begin{cases} h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t) & \text{if } x_t \geq r_t^* \\ h\mathbf{1}(d_t < r_t^*) - b\mathbf{1}(d_t \geq r_t^*) + \dot{v}_{t+1}(r_t^* - d_t) & \text{if } x_t < r_t^* \end{cases} \quad (10)$$

gives a stochastic gradient of $v_t(\cdot)$ at x_t , satisfying $\dot{v}_t(x_t) = \mathbb{E}\{\dot{v}_t(x_t, d_t)\}$. To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$, we “mimic” the computation in (10) by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k : t = 1, \dots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k , we recursively define

$$\xi_t^k(x_t, d_t, \dots, d_\tau) = \begin{cases} h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau) & \text{if } x_t \geq r_t^k \\ h\mathbf{1}(d_t < r_t^k) - b\mathbf{1}(d_t \geq r_t^k) + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \dots, d_\tau) & \text{if } x_t < r_t^k, \end{cases} \quad (11)$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \dots, \cdot) = 0$. At iteration k , replacing $\dot{v}_{t+1}(x_t - d_t)$ in (7) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau)$, we approximate a stochastic gradient of $f_t(\cdot)$ at x_t by using

$$s_t^k(x_t, d_t, \dots, d_\tau) = c + h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau). \quad (12)$$

Consequently, we propose the following algorithm to search for the optimal base-stock levels.

Algorithm 1

Step 1. Initialize the estimates of the optimal base-stock levels $\{r_t^1 : t = 1, \dots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{d_t^k : t = 1, \dots, \tau\}$ be the demand random variables at iteration k , set

$$r_t^{k+1} = r_t^k - \alpha^k s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)$$

for all $t = 1, \dots, \tau$.

Step 3. Increase k by 1 and go to Step 2.

We let \mathcal{F}_k be the filtration generated by $\{\{r_1^1, \dots, r_\tau^1\}, \{d_1^1, \dots, d_\tau^1\}, \dots, \{d_1^{k-1}, \dots, d_\tau^{k-1}\}\}$. Given \mathcal{F}_k , we assume that the conditional distribution of $\{d_t^k : t = 1, \dots, \tau\}$ is the same as the distribution of

$\{d_t : t = 1, \dots, \tau\}$. For notational brevity, we use $\mathbb{E}_k\{\cdot\}$ to denote expectations and $\mathbb{P}_k\{\cdot\}$ to denote probabilities conditional on \mathcal{F}_k . We assume that the step size parameter α^k is \mathcal{F}_k -measurable, in which case the estimates of the optimal base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$ are also \mathcal{F}_k -measurable.

Comparing (7) and (12) indicates that if the functions $\mathbb{E}_k\{\xi_{t+1}^k(\cdot, d_{t+1}^k, \dots, d_\tau^k)\}$ and $\dot{v}_{t+1}(\cdot)$ are “close” to each other, then the step directions $\mathbb{E}_k\{s_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$ and $\mathbb{E}_k\{\Delta_t(\cdot, d_t^k)\}$ are “close” to each other, in which case using $s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)$ instead of $\Delta_t(r_t^k, d_t^k)$ does not bring too much error. In fact, our convergence proof is heavily based on analyzing the error function $\dot{v}_t(\cdot) - \mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$.

In this section, we show that $\lim_{k \rightarrow \infty} \dot{f}_t(r_t^k) = 0$ w.p.1 for all $t = 1, \dots, \tau$ for a sequence of base-stock levels $\{r_t^k : t = 1, \dots, \tau\}_k$ generated by Algorithm 1 and the total expected cost of the policy that uses the base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$ converges to the total expected cost of the optimal policy as $k \rightarrow \infty$. We begin with several preliminary lemmas.

3.1 PRELIMINARIES

In the following lemma, we derive bounds on $\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)$ and $s_t^k(\cdot, d_t^k, \dots, d_\tau^k)$.

Lemma 2 *There exists a constant M such that $|\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)| \leq M$ and $|s_t^k(x_t, d_t^k, \dots, d_\tau^k)| \leq M$ w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, $k = 1, 2, \dots$*

Proof We let $N = \max\{c, h, b\}$. We show by induction over the time periods that $|\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)| \leq 2[\tau - t + 1]N$ w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, $k = 1, 2, \dots$. The result holds for time period τ by (11). Assuming that the result holds for time period $t + 1$, we have $|\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)| \leq h + b + 2[\tau - t]N \leq 2[\tau - t + 1]N$ w.p.1 and this establishes the result. Therefore, we have $|s_t^k(x_t, d_t^k, \dots, d_\tau^k)| \leq c + h + b + 2[\tau - t]N \leq 2[\tau - t + 2]N$ w.p.1 by (12). The result follows by letting $M = 2[\tau + 1]N$. \square

We note that $\dot{v}_t(\cdot)$, being the derivative of the convex function $v_t(\cdot)$, is increasing. The following lemma shows that $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$ also satisfies this property.

Lemma 3 *If \hat{x}_t, \tilde{x}_t satisfy $\hat{x}_t \leq \tilde{x}_t$, then we have $\mathbb{E}_k\{\xi_t^k(\hat{x}_t, d_t^k, \dots, d_\tau^k)\} \leq \mathbb{E}_k\{\xi_t^k(\tilde{x}_t, d_t^k, \dots, d_\tau^k)\}$ w.p.1 for all $t = 1, \dots, \tau$, $k = 1, 2, \dots$*

Proof We show the result by induction over the time periods. We first show that the result holds for time period τ . We consider three cases. First, we assume that $r_\tau^k \leq \hat{x}_\tau \leq \tilde{x}_\tau$. Using (11), we have $\mathbb{E}_k\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\} = h\mathbb{P}_k\{d_\tau^k < \hat{x}_\tau\} - b\mathbb{P}_k\{d_\tau^k \geq \hat{x}_\tau\} = [h + b]\mathbb{P}_k\{d_\tau^k < \hat{x}_\tau\} - b \leq [h + b]\mathbb{P}_k\{d_\tau^k < \tilde{x}_\tau\} - b = \mathbb{E}_k\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\}$. Second, we assume that $\hat{x}_\tau < r_\tau^k \leq \tilde{x}_\tau$. In this case, (11) and the argument in the previous sentence imply that $\mathbb{E}_k\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\} = \mathbb{E}_k\{\xi_\tau^k(r_\tau^k, d_\tau^k)\} \leq \mathbb{E}_k\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\}$. Third, we assume that $\hat{x}_\tau \leq \tilde{x}_\tau < r_\tau^k$. We have $\mathbb{E}_k\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\} = \mathbb{E}_k\{\xi_\tau^k(r_\tau^k, d_\tau^k)\} = \mathbb{E}_k\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\}$. Therefore, the result holds for time period τ . Assuming that the result holds for time period $t + 1$, it is easy to check in a similar fashion that the result holds for time period t by considering the three cases $r_t^k \leq \hat{x}_t \leq \tilde{x}_t$ or $\hat{x}_t < r_t^k \leq \tilde{x}_t$ or $\hat{x}_t \leq \tilde{x}_t < r_t^k$. \square

As mentioned above, our convergence proof analyzes the error function $\dot{v}_t(\cdot) - \mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$ extensively. For notational brevity, we let

$$e_t^k(x_t) = \dot{v}_t(x_t) - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\}, \quad (13)$$

with $e_{\tau+1}^k(\cdot) = 0$. In the following lemma, we establish a bound on the error function. This result is a direct implication of the fact that $f_t(\cdot)$ is convex and $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$ is increasing.

Lemma 4 *We have*

$$|e_t^k(x_t)| \leq \max \left\{ |\dot{f}_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}, \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\} \right\} \quad (14)$$

w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, $k = 1, 2, \dots$

Proof Using (5) and (11), we obtain

$$\begin{aligned} \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} &= \begin{cases} h \mathbb{P}_k\{d_t^k < x_t\} - b \mathbb{P}_k\{d_t^k \geq x_t\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} & \text{if } x_t \geq r_t^k \\ h \mathbb{P}_k\{d_t^k < r_t^k\} - b \mathbb{P}_k\{d_t^k \geq r_t^k\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} & \text{if } x_t < r_t^k \end{cases} \\ &= \begin{cases} \dot{f}_t(x_t) - c - \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k)\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} & \text{if } x_t \geq r_t^k \\ \dot{f}_t(r_t^k) - c - \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k)\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} & \text{if } x_t < r_t^k. \end{cases} \end{aligned} \quad (15)$$

We consider four cases. First, we assume that $x_t \geq r_t^k$ and $x_t \geq r_t^*$. Using (6) and (15), we have $e_t^k(x_t) = \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} = \mathbb{E}_k\{e_{t+1}^k(x_t - d_t^k)\}$. Therefore, we obtain $|e_t^k(x_t)| \leq \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\}$ by Jensen's inequality.

Second, we assume that $x_t \geq r_t^k$ and $x_t < r_t^*$. We have $\mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \geq \mathbb{E}_k\{\xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\}$ by Lemma 3. Using this inequality, (6) and (15), we obtain

$$\begin{aligned} e_t^k(x_t) &= -c - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \leq -c - \mathbb{E}_k\{\xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}. \end{aligned}$$

Since $x_t < r_t^*$ and r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $\dot{f}_t(x_t) \leq 0$. Using (15), we also obtain

$$\begin{aligned} e_t^k(x_t) &= -c - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \\ &= -\dot{f}_t(x_t) + \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} \geq \mathbb{E}_k\{e_{t+1}^k(x_t - d_t^k)\}. \end{aligned}$$

The last two chains of inequalities imply that

$$|e_t^k(x_t)| \leq \max \left\{ |\dot{f}_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\} \right\}.$$

Third, we assume that $x_t < r_t^k$ and $x_t \geq r_t^*$. Since $f_t(\cdot)$ is convex, we have $\dot{f}_t(r_t^k) \geq \dot{f}_t(x_t) \geq \dot{f}_t(r_t^*) = 0$. Using (6) and (15), we obtain $e_t^k(x_t) = \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} = \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}$, which implies that

$$-\dot{f}_t(r_t^k) + \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\} \leq e_t^k(x_t) \leq \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}.$$

Therefore, we obtain

$$|e_t^k(x_t)| \leq \max\left\{|\dot{f}_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)\}|\right\}.$$

Fourth, we assume that $x_t < r_t^k$ and $x_t < r_t^*$. In this case, (6) and (15) imply that $e_t^k(x_t) = -\dot{f}_t(r_t^k) + \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} = -\dot{f}_t(r_t^k) + \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}$. Therefore, we obtain $|e_t^k(x_t)| = |\dot{f}_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}|$. The result follows by combining the four cases. \square

3.2 CONVERGENCE PROOF

We have the following convergence result for Algorithm 1.

Proposition 5 *Assume that the sequence of step size parameters $\{\alpha^k\}_k$ satisfy $\alpha_k \geq 0$ for all $k = 1, 2, \dots$, $\sum_{k=1}^{\infty} \alpha^k = \infty$ and $\sum_{k=1}^{\infty} [\alpha^k]^2 < \infty$ w.p.1. If the sequence of base-stock levels $\{r_t^k : t = 1, \dots, \tau\}_k$ are generated by Algorithm 1, then the sequence $\{f_t(r_t^k)\}_k$ converges w.p.1 for all $t = 1, \dots, \tau$ and we have $\lim_{k \rightarrow \infty} \dot{f}_t(r_t^k) = 0$ w.p.1 for all $t = 1, \dots, \tau$.*

Proof All statements in the proof are in w.p.1 sense. We use induction over the time periods to show that the following results hold for all $t = 1, \dots, \tau$.

(A.1) The sequence $\{f_t(r_t^k)\}_k$ converges.

(A.2) We have $\sum_{k=1}^{\infty} \alpha^k [|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}]^+ < \infty$.

(A.3) We have $\lim_{k \rightarrow \infty} \dot{f}_t(r_t^k) = 0$.

(A.4) We have $|e_t^k(x_t)| \leq \sum_{s=t}^{\tau} |\dot{f}_s(r_s^k)|$ for all $x_t \in \mathbb{R}$, $k = 1, 2, \dots$

(A.5) There exists a constant A_t such that we have

$$|e_t^k(x_t)|^2 \leq A_t \sum_{s=t}^{\tau} [|\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k) \mathbb{E}_k\{e_{s+1}^k(r_s^k - d_s^k)\}]^+ \quad (16)$$

for all $x_t \in \mathbb{R}$, $k = 1, 2, \dots$

We begin by showing that (A.1)-(A.5) hold for time period τ . Since we have $r_\tau^{k+1} = r_\tau^k - \alpha^k s_\tau^k(r_\tau^k, d_\tau^k)$, using Lemma 1 and the Taylor series expansion of $f_\tau(\cdot)$ at r_τ^{k+1} , a standard argument yields

$$f_\tau(r_\tau^{k+1}) \leq f_\tau(r_\tau^k) - \alpha^k \dot{f}_\tau(r_\tau^k) s_\tau^k(r_\tau^k, d_\tau^k) + \frac{1}{2} [\alpha^k]^2 L |s_\tau^k(r_\tau^k, d_\tau^k)|^2; \quad (17)$$

see (3.39) in Bertsekas and Tsitsiklis (1996). Since we have $\mathbb{E}_k\{s_\tau^k(r_\tau^k, d_\tau^k)\} = c + h \mathbb{P}_k\{d_\tau^k < r_\tau^k\} - b \mathbb{P}_k\{d_\tau^k \geq r_\tau^k\} = \dot{f}_\tau(r_\tau^k)$, taking expectations in (17) and using Lemma 2 yield

$$\mathbb{E}_k\{f_\tau(r_\tau^{k+1})\} \leq f_\tau(r_\tau^k) - \alpha^k [\dot{f}_\tau(r_\tau^k)]^2 + \frac{1}{2} [\alpha^k]^2 L M^2. \quad (18)$$

Since $f_\tau(\cdot)$ is positive and $\sum_{k=1}^{\infty} [\alpha^k]^2 < \infty$, we can now use the supermartingale convergence theorem to conclude that the sequence $\{f_\tau(r_\tau^k)\}_k$ converges and $\sum_{k=1}^{\infty} \alpha^k [\dot{f}_\tau(r_\tau^k)]^2 < \infty$; see Neveu (1975). Therefore, since $e_{\tau+1}^k(\cdot) = 0$ by definition, (A.1) and (A.2) hold for time period τ . Since we have $\mathbb{E}_k \{s_\tau^k(r_\tau^k, d_\tau^k)\} = \dot{f}_\tau(r_\tau^k)$, the iteration $r_\tau^{k+1} = r_\tau^k - \alpha^k s_\tau^k(r_\tau^k, d_\tau^k)$ is a standard stochastic approximation method to minimize $f_\tau(\cdot)$ and we have $\lim_{k \rightarrow \infty} \dot{f}_\tau(r_\tau^k) = 0$; see Proposition 4.1 in Bertsekas and Tsitsiklis (1996). Therefore, (A.3) holds for time period τ . Since $e_{\tau+1}^k(\cdot) = 0$, Lemma 4 shows that (A.4) and (A.5) hold for time period τ . Therefore, (A.1)-(A.5) hold for time period τ .

Assuming that (A.1)-(A.5) hold for time periods $t+1, \dots, \tau$, Lemmas 6-8 below show that (A.1)-(A.5) also hold for time period t . This completes the induction argument, and the result follows by (A.1) and (A.3). \square

Lemmas 6-8 complete the induction argument given in the proof of Proposition 5. All statements in their proofs should be understood in w.p.1 sense.

Lemma 6 *If (A.1)-(A.5) hold w.p.1 for time periods $t+1, \dots, \tau$, then (A.1) and (A.2) hold w.p.1 for time period t .*

Proof Using (5) and (12), we have

$$\begin{aligned} \mathbb{E}_k \{s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\} &= c + h \mathbb{P}_k \{d_t^k < r_t^k\} - b \mathbb{P}_k \{d_t^k \geq r_t^k\} + \mathbb{E}_k \{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} \\ &= \dot{f}_t(r_t^k) - \mathbb{E}_k \{\dot{v}_{t+1}(r_t^k - d_t^k)\} + \mathbb{E}_k \{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)\}. \end{aligned}$$

Similar to (17) and (18), using the equality above, Lemma 1 and the Taylor series expansion of $f_t(\cdot)$ at r_t^{k+1} , we have

$$\begin{aligned} \mathbb{E}_k \{f_t(r_t^{k+1})\} &\leq f_t(r_t^k) - \alpha^k \dot{f}_t(r_t^k) \mathbb{E}_k \{s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\} + \frac{1}{2} [\alpha^k]^2 L M^2 \\ &= f_t(r_t^k) - \alpha^k \dot{f}_t(r_t^k) [\dot{f}_t(r_t^k) - \mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}] + \frac{1}{2} [\alpha^k]^2 L M^2. \end{aligned}$$

Letting $X^k = \alpha^k \dot{f}_t(r_t^k) [\dot{f}_t(r_t^k) - \mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}]$, the expression above is of the form $\mathbb{E}_k \{f_t(r_t^{k+1})\} \leq f_t(r_t^k) - [X^k]^+ + [-X^k]^+ + [\alpha^k]^2 L M^2 / 2$. Therefore, if we can show that $\sum_{k=1}^{\infty} [-X^k]^+ < \infty$, then we can use the supermartingale convergence theorem to conclude that the sequence $\{f_t(r_t^k)\}_k$ converges and $\sum_{k=1}^{\infty} [X^k]^+ < \infty$.

We now show that $\sum_{k=1}^{\infty} [-X^k]^+ < \infty$. If $[-X^k]^+ > 0$, then we have

$$0 \leq [\dot{f}_t(r_t^k)]^2 < \dot{f}_t(r_t^k) \mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\} \leq |\dot{f}_t(r_t^k)| |\mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}|.$$

Dividing the expression above by $|\dot{f}_t(r_t^k)|$, we obtain $|\dot{f}_t(r_t^k)| < |\mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}|$. Therefore, having $[-X^k]^+ > 0$ implies that

$$\begin{aligned} [-X^k]^+ &= \alpha^k [\dot{f}_t(r_t^k) \mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\} - [\dot{f}_t(r_t^k)]^2]^+ \leq \alpha^k [\dot{f}_t(r_t^k) \mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}]^+ \\ &\leq \alpha^k |\dot{f}_t(r_t^k)| |\mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}| \leq \alpha^k |\mathbb{E}_k \{e_{t+1}^k(r_t^k - d_t^k)\}|^2 \leq \alpha^k \mathbb{E}_k \{|e_{t+1}^k(r_t^k - d_t^k)|^2\}. \end{aligned}$$

We note that the expression on the right side of (16) does not depend x_t and it is \mathcal{F}_k -measurable. In this case, using the chain of inequalities above and the induction hypothesis (A.5), we obtain

$$\begin{aligned} \sum_{k=1}^{\infty} [-X^k]^+ &= \sum_{k=1}^{\infty} \mathbf{1}([-X^k]^+ > 0) [-X^k]^+ \leq \sum_{k=1}^{\infty} \alpha^k \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)|^2 \} \\ &\leq \sum_{k=1}^{\infty} \sum_{s=t+1}^{\tau} \alpha^k A_{t+1} [|\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k) \mathbb{E}_k \{ e_{s+1}^k(r_s^k - d_s^k) \}]^+ \\ &= \sum_{s=t+1}^{\tau} \sum_{k=1}^{\infty} \alpha^k A_{t+1} [|\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k) \mathbb{E}_k \{ e_{s+1}^k(r_s^k - d_s^k) \}]^+ < \infty, \end{aligned}$$

where exchanging the order of the sums in the second equality is justified by Fubini's theorem and the last inequality follows from the induction hypothesis (A.2). Therefore, we can use the supermartingale convergence theorem to conclude that $\{f_t(r_t^k)\}_k$ converges and $\sum_{k=1}^{\infty} [X^k]^+ < \infty$, which is to say that $\sum_{k=1}^{\infty} \alpha^k [|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}]^+ < \infty$. \square

Lemma 7 *If (A.1)-(A.5) hold w.p.1 for time periods $t+1, \dots, \tau$, then (A.3) holds w.p.1 for time period t .*

Proof We first show that $\liminf_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| = 0$. By the induction hypothesis (A.4), we have $|e_{t+1}^k(r_t^k - d_t^k)| \leq \sum_{s=t+1}^{\tau} |\dot{f}_s(r_s^k)|$. Taking expectations and limits, and using the induction hypothesis (A.3), we obtain $\lim_{k \rightarrow \infty} \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \} = 0$. Therefore, for given $\epsilon > 0$, there exists a finite iteration number K such that $\mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \} \leq 2\epsilon$ for all $k = K, K+1, \dots$

By Lemma 6, (A.2) holds for time period t . Since we have $\sum_{k=1}^{\infty} \alpha^k = \infty$, (A.2) implies that $\liminf_{k \rightarrow \infty} [|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}]^+ = 0$. In particular, we must have $[|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}]^+ \leq 3\epsilon^2$ for infinite number of iterations. Therefore, after iteration number K , we must have

$$|\dot{f}_t(r_t^k)|^2 - 2|\dot{f}_t(r_t^k)|\epsilon \leq |\dot{f}_t(r_t^k)|^2 - |\dot{f}_t(r_t^k)| \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \} \leq 3\epsilon^2$$

for infinite number of iterations, which implies that $|\dot{f}_t(r_t^k)| \in [-\epsilon, 3\epsilon]$ for infinite number of iterations. Since ϵ is arbitrary, we obtain $\liminf_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| = 0$.

By examining the so-called upcrossings of the interval $[\epsilon/2, \epsilon]$ by the sequence $\{ |\dot{f}_t(r_t^k)| \}_k$ and following an argument similar to the one used to show Proposition 4.1 in Bertsekas and Tsitsiklis (1996), we can also show that $\limsup_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| = 0$ and this establishes the result. We defer the proof of this part to the appendix. \square

Lemma 8 *If (A.1)-(A.5) hold w.p.1 for time periods $t+1, \dots, \tau$, then (A.4) and (A.5) hold w.p.1 for time period t .*

Proof The induction hypothesis (A.4) implies that $|e_{t+1}^k(x_t - d_t^k)| \leq \sum_{s=t+1}^{\tau} |\dot{f}_s(r_s^k)|$ for all $x_t \in \mathbb{R}$ and $|e_{t+1}^k(r_t^k - d_t^k)| \leq \sum_{s=t+1}^{\tau} |\dot{f}_s(r_s^k)|$. Taking expectations and using these expectations in (14), it is easy

to see that (A.4) holds for time period t . For all $x_t \in \mathbb{R}$, squaring (14) also implies that

$$\begin{aligned}
|e_t^k(x_t)|^2 &\leq [\dot{f}_t(r_t^k)]^2 - 2\dot{f}_t(r_t^k)\mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\} + [\mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}]^2 \\
&\quad + [\mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}]^2 + [\mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\}]^2 \\
&\leq 2[\dot{f}_t(r_t^k)]^2 - \dot{f}_t(r_t^k)\mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\} + 2\mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|^2\} + \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|^2\} \\
&\leq 2[\dot{f}_t(r_t^k)]^2 - \dot{f}_t(r_t^k)\mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\} + \\
&\quad + 3A_{t+1}\sum_{s=t+1}^{\tau} [\dot{f}_s(r_s^k)]^2 - \dot{f}_s(r_s^k)\mathbb{E}_k\{e_{s+1}^k(r_s^k - d_s^k)\} +,
\end{aligned}$$

where we use the induction hypothesis (A.5) in the last inequality. Letting $A_t = \max\{2, 3A_{t+1}\}$, (A.5) holds for time period t . \square

We close this section by investigating the performances of the policies characterized by the base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$. The policy characterized by the base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$ keeps the inventory position at time period t as close as possible to r_t^k . We let $V_t^k(x_t)$ be the total expected cost incurred by this policy over the time periods $\{t, \dots, \tau\}$ when the inventory position at time period t is x_t . The functions $\{V_t^k(\cdot) : t = 1, \dots, \tau\}$ satisfy

$$V_t^k(x_t) = \begin{cases} \mathbb{E}_k\{h[x_t - d_t^k]^+ + b[d_t^k - x_t]^+ + V_{t+1}^k(x_t - d_t^k)\} & \text{if } x_t \geq r_t^k \\ c[r_t^k - x_t] + \mathbb{E}_k\{h[r_t^k - d_t^k]^+ + b[d_t^k - r_t^k]^+ + V_{t+1}^k(r_t^k - d_t^k)\} & \text{if } x_t < r_t^k. \end{cases} \quad (19)$$

In contrast, the function $v_1(\cdot)$ gives the minimum total expected cost incurred over the time periods $\{1, \dots, \tau\}$. Proposition 9 shows that $\lim_{k \rightarrow \infty} |V_1^k(x_1) - v_1(x_1)| = 0$ w.p.1 for all $x_1 \in \mathbb{R}$ and establishes that the policies characterized by the base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$ are asymptotically optimal.

Proposition 9 *Assume that the sequence of step size parameters $\{\alpha^k\}_k$ satisfy $\alpha_k \geq 0$ for all $k = 1, 2, \dots$, $\sum_{k=1}^{\infty} \alpha^k = \infty$ and $\sum_{k=1}^{\infty} [\alpha^k]^2 < \infty$ w.p.1. If the sequence of base-stock levels $\{r_t^k : t = 1, \dots, \tau\}_k$ are generated by Algorithm 1, then we have $\lim_{k \rightarrow \infty} |V_1^k(x_1) - v_1(x_1)| = 0$ w.p.1 for all $x_1 \in \mathbb{R}$.*

Proof All statements are in w.p.1 sense. We first show that $\lim_{k \rightarrow \infty} f_t(r_t^k) = f_t(r_t^*)$ for all $t = 1, \dots, \tau$. By Proposition 5, the sequence $\{f_t(r_t^k)\}_k$ converges, which implies that there exists a subsequence $\{r_t^{k_j}\}_j$ with the limit point \hat{r}_t . Since the sequence $\{\dot{f}_t(r_t^k)\}_k$ converges to 0 by Proposition 5, we must have $\dot{f}_t(\hat{r}_t) = 0$. Therefore, since $f_t(\cdot)$ is convex, \hat{r}_t is a minimizer of $f_t(\cdot)$ satisfying $f_t(\hat{r}_t) = f_t(r_t^*)$. In this case, the subsequence $\{f_t(r_t^{k_j})\}_j$ converges to $f_t(r_t^*)$. Since the sequence $\{f_t(r_t^k)\}_k$ converges, we conclude that this sequence converges to $f_t(r_t^*)$.

Noting (3), we can write (19) as

$$V_t^k(x_t) = \begin{cases} f_t(x_t) - cx_t + \mathbb{E}_k\{V_{t+1}^k(x_t - d_t^k) - v_{t+1}(x_t - d_t^k)\} & \text{if } x_t \geq r_t^k \\ f_t(r_t^k) - cx_t + \mathbb{E}_k\{V_{t+1}^k(r_t^k - d_t^k) - v_{t+1}(r_t^k - d_t^k)\} & \text{if } x_t < r_t^k. \end{cases} \quad (20)$$

We now use induction over the time periods to show that $0 \leq V_t^k(x_t) - v_t(x_t) \leq \sum_{s=t}^{\tau} [f_s(r_s^k) - f_s(r_s^*)]$ for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, in which case the final result follows by noting that $\{f_t(r_t^k)\}_k$ converges to

$f_t(r_t^*)$ for all $t = 1, \dots, \tau$. It is easy to show the result for the last time period. Assuming that the result holds for time period $t + 1$, we now show that the result holds for time period t . We consider four cases. Assume that $r_t^k \leq x_t < r_t^*$. Since r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $f_t(x_t) \leq f_t(r_t^k)$. In this case, using (4), (20) and the induction hypothesis, we obtain

$$\begin{aligned} 0 \leq V_t^k(x_t) - v_t(x_t) &= f_t(x_t) - f_t(r_t^*) + \mathbb{E}_k\{V_{t+1}^k(x_t - d_t^k) - v_{t+1}(x_t - d_t^k)\} \\ &\leq f_t(r_t^k) - f_t(r_t^*) + \sum_{s=t+1}^{\tau} [f_s(r_s^k) - f_s(r_s^*)]. \end{aligned}$$

The other three cases where we have $r_t^* \leq x_t < r_t^k$, or $r_t^* \leq x_t$ and $r_t^k \leq x_t$, or $r_t^* \geq x_t$ and $r_t^k \geq x_t$ can be shown similarly. \square

4 MULTI-PERIOD NEWSVENDOR PROBLEM WITH LOST SALES

This section shows how to extend the ideas in Section 3 to the case where the unsatisfied demand is immediately lost. We use the same assumptions for the cost parameters and the demand random variables. In particular, we have $b > c \geq 0$, $h \geq 0$ and the demand random variables at different time periods are independent, but not necessarily identically distributed. However, we need to strictly impose the assumption that the lead times for the replenishments are zero. Otherwise, the base-stock policies are not necessarily optimal. In addition, our presentation here strictly imposes the assumption that the cost parameters are stationary, but the online supplement extends our analysis to the case where the cost parameters are nonstationary. Letting $v_t(x_t)$ have the same interpretation as in Section 3, the functions $\{v_t(\cdot) : t = 1, \dots, \tau\}$ satisfy the Bellman equations

$$v_t(x_t) = \min_{y_t \geq x_t} c[y_t - x_t] + \mathbb{E}\{h[y_t - d_t]^+ + b[d_t - y_t]^+ + v_{t+1}([y_t - d_t]^+)\}, \quad (21)$$

with $v_{\tau+1}(\cdot) = 0$. We also let

$$f_t(r_t) = c r_t + \mathbb{E}\{h[r_t - d_t]^+ + b[d_t - r_t]^+ + v_{t+1}([r_t - d_t]^+)\}.$$

It can be shown that $v_t(\cdot)$ and $f_t(\cdot)$ are positive, Lipschitz continuous, differentiable and convex functions, and $f_t(\cdot)$ has a finite unconstrained minimizer. In this case, the optimal base-stock levels $\{r_t^* : t = 1, \dots, \tau\}$ are the minimizers of the functions $\{f_t(\cdot) : t = 1, \dots, \tau\}$.

Since we have

$$\dot{f}_t(r_t) = c + h\mathbb{P}\{d_t < r_t\} - b\mathbb{P}\{d_t \geq r_t\} + \mathbb{E}\{\dot{v}_{t+1}(r_t - d_t)\mathbf{1}(d_t < r_t)\}, \quad (22)$$

we can compute a stochastic gradient of $f_t(\cdot)$ at x_t through

$$\Delta_t(x_t, d_t) = c + h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t)\mathbf{1}(d_t < x_t) \quad (23)$$

and iteratively search for the optimal base-stock levels through (8). However, this approach requires the knowledge of $\{v_t(\cdot) : t = 1, \dots, \tau\}$. We now use ideas similar to those in Section 3 to approximate the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$ in a tractable manner.

Using the optimal base-stock level r_t^* , we write (21) as

$$v_t(x_t) = \begin{cases} \mathbb{E}\{h[x_t - d_t]^+ + b[d_t - x_t]^+ + v_{t+1}([x_t - d_t]^+)\} & \text{if } x_t \geq r_t^* \\ c[r_t^* - x_t] + \mathbb{E}\{h[r_t^* - d_t]^+ + b[d_t - r_t^*]^+ + v_{t+1}([r_t^* - d_t]^+)\} & \text{if } x_t < r_t^*. \end{cases} \quad (24)$$

Since r_t^* is the minimizer of $f_t(\cdot)$, (22) implies that $-c = \dot{f}_t(r_t^*) - c = h\mathbb{P}\{d_t < r_t^*\} - b\mathbb{P}\{d_t \geq r_t^*\} + \mathbb{E}\{\dot{v}_{t+1}(r_t^* - d_t)\mathbf{1}(d_t < r_t^*)\}$. Therefore, using this expression in (24), we obtain

$$\dot{v}_t(x_t) = \begin{cases} h\mathbb{P}\{d_t < x_t\} - b\mathbb{P}\{d_t \geq x_t\} + \mathbb{E}\{\dot{v}_{t+1}(x_t - d_t)\mathbf{1}(d_t < x_t)\} & \text{if } x_t \geq r_t^* \\ h\mathbb{P}\{d_t < r_t^*\} - b\mathbb{P}\{d_t \geq r_t^*\} + \mathbb{E}\{\dot{v}_{t+1}(r_t^* - d_t)\mathbf{1}(d_t < r_t^*)\} & \text{if } x_t < r_t^*. \end{cases} \quad (25)$$

In this case,

$$\dot{v}_t(x_t, d_t) = \begin{cases} h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t)\mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^* \\ h\mathbf{1}(d_t < r_t^*) - b\mathbf{1}(d_t \geq r_t^*) + \dot{v}_{t+1}(r_t^* - d_t)\mathbf{1}(d_t < r_t^*) & \text{if } x_t < r_t^* \end{cases} \quad (26)$$

clearly gives a stochastic gradient of $v_t(\cdot)$ at x_t .

To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$, we “mimic” the computation in (26) by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k : t = 1, \dots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k , we recursively define

$$\xi_t^k(x_t, d_t, \dots, d_\tau) = \begin{cases} h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) \\ \quad + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau)\mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^k \\ h\mathbf{1}(d_t < r_t^k) - b\mathbf{1}(d_t \geq r_t^k) \\ \quad + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \dots, d_\tau)\mathbf{1}(d_t < r_t^k) & \text{if } x_t < r_t^k, \end{cases} \quad (27)$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \dots, \cdot) = 0$. At iteration k , replacing $\dot{v}_{t+1}(x_t - d_t)$ in (23) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau)$, we use

$$s_t^k(x_t, d_t, \dots, d_\tau) = c + h\mathbf{1}(d_t < x_t) - b\mathbf{1}(d_t \geq x_t) + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau)\mathbf{1}(d_t < x_t) \quad (28)$$

to approximate the stochastic gradient of $f_t(\cdot)$ at x_t . Thus, we can use Algorithm 1 to search for the optimal base-stock levels. The only difference is that we need to use the step direction above in Step 2.

The proof of convergence for this algorithm follows from an argument similar to the one in Sections 3.1 and 3.2. In particular, we can follow the proof of Lemma 2 to derive bounds on $\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)$ and $s_t^k(\cdot, d_t^k, \dots, d_\tau^k)$. It is possible to establish an analogue of Lemma 3 to show that $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)\}$ is increasing. We define the error function as

$$e_t^k(x_t) = \dot{v}_t(x_t)\mathbf{1}(x_t > 0) - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\mathbf{1}(x_t > 0)\},$$

with $e_{\tau+1}^k(\cdot) = 0$. In this case, we can show that the same bound on the error function given in Lemma 4 holds. Once we have this bound on the error function, we can follow the same induction argument in Proposition 5, Lemmas 6-8 and Proposition 9 to show the final result. The details of the proof are given in the online supplement.

5 CENSORED DEMANDS

This section considers the multi-period newsvendor problem with lost sales and censored demands. Demand censorship refers to the situation where we only observe the amount of inventory sold, but not the amount of demand. In this case, our demand observations are “truncated” when the amount of demand exceeds the amount of available inventory. Our goal is to show that we can still compute the step direction in (28), which implies that the algorithm proposed in Section 4 remains applicable when the demand information is censored. We note if the unsatisfied demand is backlogged, then we can always observe the amount of demand and the censored demand information is not an issue.

If the unsatisfied demand is immediately lost and the demand information is censored, then we do not observe the random variables $\{d_t^k : t = 1, \dots, \tau\}$ in Step 2 of Algorithm 1. Instead, we simulate the behavior of the policy characterized by the base-stock levels $\{r_t^k : t = 1, \dots, \tau\}$ and Step 2 of Algorithm 1 is replaced by the following steps.

Step 2.a. Initialize the beginning inventory position x_1^k . Set $t = 1$.

Step 2.b. Place a replenishment order of $[r_t^k - x_t^k]^+$ units to raise the inventory position to $\max\{r_t^k, x_t^k\}$. Set the inventory position after the replenishment order as $y_t^k = \max\{r_t^k, x_t^k\}$.

Step 2.c. Compute the inventory position at the next time period as $x_{t+1}^k = y_t^k - \min\{y_t^k, d_t^k\}$. If $t < \tau$, then increase t by 1 and go to Step 2.b.

Step 2.d. Set $r_t^{k+1} = r_t^k - \alpha^k s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)$ for all $t = 1, \dots, \tau$.

Therefore, we only have access to $\{\min\{y_t^k, d_t^k\} : t = 1, \dots, \tau\}$, but not the demand random variables themselves. Proposition 10 shows that this information is adequate to compute the step direction.

Proposition 10 *Knowledge of $\{r_t^k : t = 1, \dots, \tau\}$, $\{y_t^k : t = 1, \dots, \tau\}$ and $\{\min\{y_t^k, d_t^k\} : t = 1, \dots, \tau\}$ is adequate to compute $s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)$ in (28) for all $t = 1, \dots, \tau$.*

Proof It is possible to show the result by induction over the time periods, but we use a constructive proof, which is more instructive and easier to follow. We begin with a chain of inequalities that directly follow from Steps 2.a-2.c above. For any time period s , we have $y_s^k \geq r_s^k$, $y_s^k \geq x_s^k$ and $x_{s+1}^k \geq y_s^k - d_s^k$, from which we obtain $r_s^k - d_s^k \leq y_s^k - d_s^k \leq x_{s+1}^k \leq y_{s+1}^k$, $y_{s+1}^k - d_{s+1}^k \leq y_{s+2}^k, \dots, y_{t-1}^k - d_{t-1}^k \leq y_t^k$ for all $t = s + 1, \dots, \tau$. Combining these inequalities, we have $r_s^k - d_s^k - d_{s+1}^k - \dots - d_{t-1}^k \leq y_t^k$ for all $t = s + 1, \dots, \tau$. Consequently, if we have $\min\{y_t^k, d_t^k\} = y_t^k$ for any time period t , then we must have $r_s^k \leq d_s^k + d_{s+1}^k + \dots + d_t^k$ for all $s = 1, \dots, t - 1$.

Assume that we want to compute $s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)$, where we have $s_t^k(r_t^k, d_t^k, \dots, d_\tau^k) = c + h \mathbf{1}(d_t^k < r_t^k) - b \mathbf{1}(d_t^k \geq r_t^k) + \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k)$. We consider two cases.

Case 1. Assume that $\min\{y_t^k, d_t^k\} = y_t^k$. In this case, we can deduce that $d_t^k \geq y_t^k \geq r_t^k$. Therefore, we have $s_t^k(r_t^k, d_t^k, \dots, d_\tau^k) = c - b$ and we are done.

Case 2. Assume that $\min\{y_t^k, d_t^k\} < y_t^k$. In this case, since we know the value of $\min\{y_t^k, d_t^k\}$, we can deduce the value of d_t^k as being equal to $\min\{y_t^k, d_t^k\}$. Thus, since we know the values of d_t^k and r_t^k , we can

compute $\mathbf{1}(d_t^k < r_t^k)$ and $\mathbf{1}(d_t^k \geq r_t^k)$. Therefore, it only remains to compute $\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k)$ for a known value of $r_t^k - d_t^k$. We consider two subcases.

Case 2.a. Assume that $\min\{y_{t+1}^k, d_{t+1}^k\} = y_{t+1}^k$. In this case, we can deduce that $d_{t+1}^k \geq y_{t+1}^k \geq r_{t+1}^k$. By the inequality we derive at the beginning of the proof, we have $r_t^k \leq d_t^k + d_{t+1}^k$. Using (27), we have

$$\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) = \begin{cases} h \mathbf{1}(d_{t+1}^k < r_t^k - d_t^k) - b \mathbf{1}(d_{t+1}^k \geq r_t^k - d_t^k) \\ \quad + \xi_{t+2}^k(r_t^k - d_t^k - d_{t+1}^k, d_{t+2}^k, \dots, d_\tau^k) \mathbf{1}(d_{t+1}^k < r_t^k - d_t^k) & \text{if } r_t^k - d_t^k \geq r_{t+1}^k \\ h \mathbf{1}(d_{t+1}^k < r_{t+1}^k) - b \mathbf{1}(d_{t+1}^k \geq r_{t+1}^k) \\ \quad + \xi_{t+2}^k(r_{t+1}^k - d_{t+1}^k, d_{t+2}^k, \dots, d_\tau^k) \mathbf{1}(d_{t+1}^k < r_{t+1}^k) & \text{if } r_t^k - d_t^k < r_{t+1}^k, \end{cases}$$

which is equal to $-b$ in either one of the cases and we are done.

Case 2.b. Assume that $\min\{y_{t+1}^k, d_{t+1}^k\} < y_{t+1}^k$. In this case, we can deduce the value of d_{t+1}^k as being equal to $\min\{y_{t+1}^k, d_{t+1}^k\}$. Thus, since we know the values of r_t^k , r_{t+1}^k , d_t^k and d_{t+1}^k , we can compute $\mathbf{1}(d_{t+1}^k < r_t^k - d_t^k)$, $\mathbf{1}(d_{t+1}^k \geq r_t^k - d_t^k)$, $\mathbf{1}(d_{t+1}^k < r_{t+1}^k)$ and $\mathbf{1}(d_{t+1}^k \geq r_{t+1}^k)$ in the expression above. Therefore, it only remains to compute $\xi_{t+2}^k(r_t^k - d_t^k - d_{t+1}^k, d_{t+2}^k, \dots, d_\tau^k)$ and $\xi_{t+2}^k(r_{t+1}^k - d_{t+1}^k, d_{t+2}^k, \dots, d_\tau^k)$ for known values of $r_t^k - d_t^k - d_{t+1}^k$ and $r_{t+1}^k - d_{t+1}^k$. The result follows by continuing in the same fashion for the subsequent time periods. \square

6 INVENTORY PURCHASING PROBLEM UNDER PRICE UNCERTAINTY

We want to make purchasing decisions for a product over the time periods $\{1, \dots, \tau\}$. The price of the product changes randomly over time and the goal is to satisfy the demand for the product at the end of the planning horizon with minimum total expected cost. We borrow this problem from Nascimento and Powell (2006). A possible application area for it is the situation where we need to lease storage space on an ocean liner. The price of storage space changes randomly over time and the amount of storage space that we actually need becomes known just before the departure time of the ocean liner.

We let p_t be the price at time period t , d be the demand and b be the penalty cost associated with not being able to satisfy a unit of demand. We assume that the random variables $\{p_t : t = 1, \dots, \tau\}$ and d are independent of each other, take positive values and have finite expectations. We assume that the cumulative distribution function of d is Lipschitz continuous and p_t has a finite support \mathcal{P}_t . When the distinction is crucial, we use \hat{p}_t to denote a particular realization of the random variable p_t . Letting x_t be the total amount of product purchased up to time period t , the optimal policy is characterized by the Bellman equations

$$v_t(x_t) = \mathbb{E} \left\{ \min_{y_t \geq x_t} p_t [y_t - x_t] + v_{t+1}(y_t) \right\}, \quad (29)$$

with $v_{\tau+1}(x_{\tau+1}) = b \mathbb{E} \{ [d - x_{\tau+1}]^+ \}$. Letting

$$f_t(r_t, p_t) = p_t r_t + v_{t+1}(r_t),$$

it can be shown that $f_t(\cdot, p_t)$ is a convex function with a finite unconstrained minimizer, say $r_t^*(p_t)$. In this case, it is easy to see that the optimal policy is a price-dependent base-stock policy characterized

by the base-stock levels $\{r_t^*(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$. That is, if the total amount of product purchased up to time period t is x_t and the price of the product is \hat{p}_t , then it is optimal to purchase $[r_t^*(\hat{p}_t) - x_t]^+$ units at time period t . It can be shown that $f_t(\cdot, p_t)$ and $v_t(\cdot)$ are positive, Lipschitz continuous, differentiable and convex functions. Since we have

$$\dot{f}_t(r_t, p_t) = p_t + \dot{v}_{t+1}(r_t), \quad (30)$$

we can compute the derivative of $f_t(\cdot, p_t)$ at x_t through

$$\Delta_t(x_t, p_t) = p_t + \dot{v}_{t+1}(x_t), \quad (31)$$

where $\dot{f}_t(\cdot, p_t)$ refers to the derivative with respect to the first argument. Since $r_t^*(\hat{p}_t)$ is the minimizer of $f_t(\cdot, \hat{p}_t)$, we can search for the optimal base-stock levels through

$$r_t^{k+1}(\hat{p}_t) = r_t^k(\hat{p}_t) - \alpha^k \Delta_t(r_t^k(\hat{p}_t), \hat{p}_t)$$

for all $\hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau$, where $\{r_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ are the estimates of the optimal base-stock levels at iteration k . Similar to Sections 3 and 4, we now approximate the derivatives of $\{f_t(\cdot, \hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ in a tractable manner.

Using the optimal base-stock level $r_t^*(p_t)$, we write (29) as $v_t(x_t) = \mathbb{E}\{v_t(x_t, p_t)\}$, where

$$v_t(x_t, p_t) = \begin{cases} v_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ p_t [r_t^*(p_t) - x_t] + v_{t+1}(r_t^*(p_t)) & \text{if } x_t < r_t^*(p_t). \end{cases} \quad (32)$$

Therefore, a stochastic gradient of $v_t(\cdot)$ at x_t can be obtained through

$$\dot{v}_t(x_t, p_t) = \begin{cases} \dot{v}_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ -p_t & \text{if } x_t < r_t^*(p_t). \end{cases} \quad (33)$$

Since $r_t^*(p_t)$ is the minimizer of $f_t(\cdot, p_t)$, (30) implies that $-p_t = \dot{v}_{t+1}(r_t^*(p_t))$ and we obtain

$$\dot{v}_t(x_t, p_t) = \begin{cases} \dot{v}_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ \dot{v}_{t+1}(r_t^*(p_t)) & \text{if } x_t < r_t^*(p_t). \end{cases} \quad (34)$$

To construct tractable approximations to the derivatives of $\{f_t(\cdot, \hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$, we “mimic” the computation above by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k , we define

$$\xi_t^k(x_t, p_t, \dots, p_\tau, d) = \begin{cases} \xi_{t+1}^k(x_t, p_{t+1}, \dots, p_\tau, d) & \text{if } x_t \geq r_t^k(p_t) \\ \xi_{t+1}^k(r_t^k(p_t), p_{t+1}, \dots, p_\tau, d) & \text{if } x_t < r_t^k(p_t), \end{cases} \quad (35)$$

with $\xi_{\tau+1}^k(x_{\tau+1}, d) = -b \mathbf{1}(d \geq x_{\tau+1})$. At iteration k , replacing $\dot{v}_{t+1}(x_t)$ in (31) with $\xi_{t+1}^k(x_t, p_{t+1}, \dots, p_\tau, d)$, we use

$$s_t^k(x_t, p_t, \dots, p_\tau, d) = p_t + \xi_{t+1}^k(x_t, p_{t+1}, \dots, p_\tau, d)$$

to approximate the derivative of $f_t(\cdot, p_t)$ at x_t . Consequently, we propose the following algorithm to search for the optimal base-stock levels.

Algorithm 2

Step 1. Initialize the estimates of the optimal base-stock levels $\{r_t^1(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{p_t^k : t = 1, \dots, \tau\}$ be the price random variables and d^k be the demand random variable at iteration k , set

$$r_t^{k+1}(p_t^k) = r_t^k(p_t^k) - \alpha^k s_t^k(r_t^k, p_t^k, \dots, p_\tau^k, d^k)$$

for all $t = 1, \dots, \tau$. Furthermore, set $r_t^{k+1}(\hat{p}_t) = r_t^k(\hat{p}_t)$ for all $\hat{p}_t \in \mathcal{P}_t \setminus \{p_t^k\}, t = 1, \dots, \tau$.

Step 3. Increase k by 1 and go to Step 2.

We emphasize that only the base-stock levels $\{r_t^k(p_t^k) : t = 1, \dots, \tau\}$ are “updated” at iteration k in Step 2 of Algorithm 2. The other base-stock levels $\{r_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t \setminus \{p_t^k\}, t = 1, \dots, \tau\}$ remain the same. The proof of convergence for Algorithm 2 follows from an argument similar to the one in Sections 3.1 and 3.2. We can follow the proof of Lemma 2 to derive bounds on $\xi_t^k(\cdot, p_t^k, \dots, p_\tau^k, d^k)$ and $s_t^k(\cdot, p_t^k, \dots, p_\tau^k, d^k)$, and the proof of Lemma 3 to show that $\mathbb{E}_k\{\xi_t^k(\cdot, p_t^k, \dots, p_\tau^k, d^k)\}$ is increasing. We define the error function as

$$e_t^k(x_t, \hat{p}_t) = \dot{v}_t(x_t, \hat{p}_t) - \mathbb{E}_k\{\xi_t^k(x_t, \hat{p}_t, p_{t+1}^k, \dots, p_\tau^k, d^k)\},$$

with $e_{\tau+1}^k(\cdot, \cdot) = 0$. In this case, we can show that

$$|e_t^k(x_t, \hat{p}_t)| \leq \max \left\{ \left| \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) - \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\} \right|, \right. \\ \left. \mathbb{E}_k\{|e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)|\}, \mathbb{E}_k\{|e_{t+1}^k(x_t, p_{t+1}^k)|\} \right\}$$

w.p.1 for all $x_t \in \mathbb{R}, \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau, k = 1, 2, \dots$. Once we have this bound on the error function, we can follow the same induction argument in Proposition 5, Lemmas 6-8 and Proposition 9 to show the final result. In particular, we can show that $\lim_{k \rightarrow \infty} \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) = 0$ w.p.1 for all $\hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau$. The details of the proof are given in the online supplement.

7 NUMERICAL ILLUSTRATIONS

This section focuses on the problems described in Sections 3, 4 and 6, and numerically compares the performances of Algorithms 1 and 2 with standard stochastic approximation methods.

7.1 MULTI-PERIOD NEWSVENDOR PROBLEM WITH BACKLOGGED DEMANDS

We consider a policy characterized by the base-stock levels $\{r_t : t = 1, \dots, \tau\}$. That is, if the inventory position at time period t is x_t , then this policy orders $[r_t - x_t]^+$ units. If we follow this policy starting with the initial inventory position x_1 and the demands over the planning horizon turn out to be $\{d_t : t = 1, \dots, \tau\}$, then the inventory position at time period t is given by

$$x_t = \max \left\{ x_1 - \sum_{s=1}^{t-1} d_s, r_1 - \sum_{s=1}^{t-1} d_s, \dots, r_{t-1} - \sum_{s=t-1}^{t-1} d_s \right\}.$$

This is easy to see by noting that the inventory position at time period $t + 1$ is $\max\{x_t, r_t\} - d_t$ and using induction over the time periods. In this case, the holding cost that we incur at time period t is

$$\begin{aligned} H_t(x_1, D | r) &= h [\max\{x_t, r_t\} - d_t]^+ \\ &= h [\max\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \dots, r_{t-1} - D_{t-1}^{t-1}, r_t\} - d_t]^+ \\ &= h \max\{x_1 - D_1^t, r_1 - D_1^t, \dots, r_{t-1} - D_{t-1}^t, r_t - D_t^t, 0\}, \end{aligned} \quad (36)$$

where we let $D_s^t = d_s + \dots + d_t$ for notational brevity and use D to denote the cumulative demands $\{D_s^t : s = 1, \dots, \tau, t = s, \dots, \tau\}$ and r to denote the base-stock levels $\{r_t : t = 1, \dots, \tau\}$. Similarly, the backlogging cost that we incur at time period t is

$$\begin{aligned} B_t(x_1, D | r) &= b [d_t - \max\{x_t, r_t\}]^+ \\ &= b [d_t - \max\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \dots, r_{t-1} - D_{t-1}^{t-1}, r_t\}]^+ \\ &= b \max\{\min\{D_1^t - x_1, D_1^t - r_1, \dots, D_{t-1}^t - r_{t-1}, D_t^t - r_t\}, 0\}, \end{aligned} \quad (37)$$

whereas the replenishment cost that we incur at time period t is

$$\begin{aligned} C_t(x_1, D | r) &= c [r_t - x_t]^+ \\ &= c [r_t - \max\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \dots, r_{t-1} - D_{t-1}^{t-1}\}]^+ \\ &= c \max\{\min\{r_t - x_1 + D_1^{t-1}, r_t - r_1 + D_1^{t-1}, \dots, r_t - r_{t-1} + D_{t-1}^{t-1}\}, 0\}. \end{aligned} \quad (38)$$

Therefore, we can try to solve the problem $\min_r \mathbb{E}\{\sum_{t=1}^{\tau} [H_t(x_1, D | r) + B_t(x_1, D | r) + C_t(x_1, D | r)]\}$ to compute the optimal base-stock levels. However, it is easy to check that the objective function of this problem is not necessarily differentiable with respect to r . We overcome this technical difficulty by using an approach proposed by van Ryzin and Vulcano (2006). In particular, we let $\{\zeta_t : t = 1, \dots, \tau\}$ be uniformly distributed random variables over the small interval $[0, \epsilon]$ and perturb the base-stock levels by using these random variables. As a result, we solve the problem

$$\min_r \mathbb{E} \left\{ \sum_{t=1}^{\tau} [H_t(x_1, D | r + \zeta) + B_t(x_1, D | r + \zeta) + C_t(x_1, D | r + \zeta)] \right\}, \quad (39)$$

where we use $r + \zeta$ to denote the perturbed base-stock levels $\{r_t + \zeta_t : t = 1, \dots, \tau\}$. It is now possible to show that the objective function of problem (39) is differentiable with respect to r and its gradient is Lipschitz continuous. Therefore, we can use a standard stochastic approximation method to solve this problem. If ϵ is small, then solving problem (39) instead of the original problem should not cause too much error.

After straightforward algebraic manipulations on (36) and (37), it is easy to see that the r_s -th component in the gradient of $H_t(x_1, D | r)$ with respect to r is given by

$$\nabla_s H_t(x_1, D | r) = \begin{cases} h \mathbf{1}(r_s - D_s^t \geq 0) \times \mathbf{1}(r_s - D_s^t \geq x_1 - D_1^t) \\ \quad \times \mathbf{1}(r_s - D_s^t \geq r_1 - D_1^t) \times \dots \times \mathbf{1}(r_s - D_s^t \geq r_t - D_t^t) & \text{if } s \leq t \\ 0 & \text{if } s > t, \end{cases} \quad (40)$$

whereas the r_s -th component in the gradient of $B_t(x_1, D | r)$ with respect to r is given by

$$\nabla_s B_t(x_1, D | r) = \begin{cases} -b \mathbf{1}(D_s^t - r_s \geq 0) \times \mathbf{1}(D_s^t - r_s \leq D_1^t - x_1) \\ \quad \times \mathbf{1}(D_s^t - r_s \leq D_1^t - r_1) \times \dots \times \mathbf{1}(D_s^t - r_s \leq D_t^t - r_t) & \text{if } s \leq t \\ 0 & \text{if } s > t. \end{cases} \quad (41)$$

To be precise, the gradients of $H_t(x_1, D | r)$ or $B_t(x_1, D | r)$ do not exist everywhere. However, it is possible to check that the gradients of $H_t(x_1, D | r + \zeta)$ and $B_t(x_1, D | r + \zeta)$ exist everywhere w.p.1 and we can replace $\{r_t : t = 1, \dots, \tau\}$ with $\{r_t + \zeta_t : t = 1, \dots, \tau\}$ in the expressions above to compute the r_s -th components in the gradients of $H_t(x_1, D | r + \zeta)$ and $B_t(x_1, D | r + \zeta)$. Similarly, after straightforward algebraic manipulations on (38) and some simplifications, it is easy to see that the r_s -th component in the gradient of $C_t(x_1, D | r)$ with respect to r is given by

$$\nabla_s C_t(x_1, D | r) = \begin{cases} -c \mathbf{1}(r_t - r_s + D_s^{t-1} \geq 0) \times \mathbf{1}(D_s^{t-1} - r_s \leq D_1^{t-1} - x_1) \\ \quad \times \mathbf{1}(D_s^{t-1} - r_s \leq D_1^{t-1} - r_1) \times \dots \times \mathbf{1}(D_s^{t-1} - r_s \leq D_{t-1}^{t-1} - r_{t-1}) & \text{if } s < t \\ c \mathbf{1}(r_t - x_1 + D_1^{t-1} \geq 0) \\ \quad \times \mathbf{1}(r_t - r_1 + D_1^{t-1} \geq 0) \times \dots \times \mathbf{1}(r_t - r_{t-1} + D_{t-1}^{t-1} \geq 0) & \text{if } s = t \\ 0 & \text{if } s > t. \end{cases} \quad (42)$$

Consequently, the following algorithm is a standard stochastic approximation method for solving problem (39).

Algorithm 3

Step 1. Initialize the estimates of the optimal base-stock levels $\{r_t^1 : t = 1, \dots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{d_t^k : t = 1, \dots, \tau\}$ be the demand random variables and $\{\zeta_t^k : t = 1, \dots, \tau\}$ be the perturbation random variables at iteration k , set

$$r_t^{k+1} = r_t^k - \alpha^k \sum_{s=1}^{\tau} [\nabla_t H_s(x_1, D^k | r^k + \zeta^k) + \nabla_t B_s(x_1, D^k | r^k + \zeta^k) + \nabla_t C_s(x_1, D^k | r^k + \zeta^k)]$$

for all $t = 1, \dots, \tau$, we use D^k to denote the cumulative demands $\{d_s^k + \dots + d_t^k : s = 1, \dots, \tau, t = s, \dots, \tau\}$.

Step 3. Increase k by 1 and go to Step 2.

We can use Proposition 4.1 in Bertsekas and Tsitsiklis (1996) to show that the iterates of this algorithm converge w.p.1 to a stationary point of the objective function of problem (39).

Our test problems use four demand distributions labeled by NR, UN, EX and BT. Specifically, NR, UN, EX and BT respectively correspond to the cases where d_t is normally distributed with mean μ_t and standard deviation σ_t , uniformly distributed over the interval $[l_t, u_t]$, exponentially distributed with mean λ_t and beta distributed with shape parameters α_t^1 and α_t^2 . To choose values for $\{(\mu_t, \sigma_t) : t = 1, \dots, \tau\}$, $\{(l_t, u_t) : t = 1, \dots, \tau\}$, $\{\lambda_t : t = 1, \dots, \tau\}$ and $\{(\alpha_t^1, \alpha_t^2) : t = 1, \dots, \tau\}$, we sample $\mu_t, l_t, u_t, \lambda_t, \alpha_t^1$ and α_t^2 from the uniform distribution over the interval $[1, 20]$ and let $\sigma_t = \mu_t/3$. The per unit replenishment and backlogging costs are respectively 0.1 and 0.5.

We run Algorithms 1 and 3 for 10,000 iterations. We make 25 different runs for each algorithm to eliminate the effect of the randomness in the samples of the demand random variables. Each run starts from different initial base-stock levels and uses a different sequence of the samples of the demand random variables. To be fair, the s -th run for both algorithms starts from the same initial base-stock levels and uses the same sequence of the samples of the demand random variables. We choose the initial base-stock levels $\{r_t^1 : t = 1, \dots, \tau\}$ by sampling r_t^1 from the uniform distribution over the interval $[0, 40]$. We use the step size parameter $\alpha^k = 100/(40 + k)$ at iteration k . Letting $r^{A1}(s) = \{r_t^{A1}(s) : t = 1, \dots, \tau\}$ and $r^{A3}(s) = \{r_t^{A3}(s) : t = 1, \dots, \tau\}$ respectively be the base-stock levels obtained by Algorithms 1 and 3 after the 10,000-th iteration of the s -th run, we are interested in the performance measures

$$\begin{aligned} \text{AV}^A &= \frac{1}{25} \sum_{s=1}^{25} \mathbb{E} \left\{ \sum_{t=1}^{\tau} [H_t(x_1, D | r^A(s)) + B_t(x_1, D | r^A(s)) + C_t(x_1, D | r^A(s))] \right\} \\ \text{MX}^A &= \max_{s \in \{1, \dots, 25\}} \left\{ \mathbb{E} \left\{ \sum_{t=1}^{\tau} [H_t(x_1, D | r^A(s)) + B_t(x_1, D | r^A(s)) + C_t(x_1, D | r^A(s))] \right\} \right\} \\ \text{MI}^A &= \min_{s \in \{1, \dots, 25\}} \left\{ \mathbb{E} \left\{ \sum_{t=1}^{\tau} [H_t(x_1, D | r^A(s)) + B_t(x_1, D | r^A(s)) + C_t(x_1, D | r^A(s))] \right\} \right\} \end{aligned}$$

for $A \in \{A1, A3\}$. These performance measures capture the average, worst-case and best-case performances over all runs. We estimate the expectations in the expressions above by simulating the behavior of the policy characterized by the base-stock levels $\{r_t^A(s) : t = 1, \dots, \tau\}$ for all $s = 1, \dots, 25$.

Our first set of computational results are summarized in Table 1. The first column in this table shows the problem parameters by using the triplets $(\tau, h, d) \in \{5, 10\} \times \{0.1, 0.25\} \times \{\text{NR}, \text{UN}, \text{EX}, \text{BT}\}$, where τ is the number of time periods, h is the per unit holding cost and d is the demand distribution. The second column shows the total expected cost incurred by the optimal policy. We obtain the optimal policy by discretizing the demand distributions and solving the Bellman equations approximately. The third, fourth and fifth columns show AV, MX and MI for Algorithm 1, whereas the eighth, ninth and tenth columns show AV, MX and MI for Algorithm 3.

Our computational results show that even the worst-case performance of Algorithm 1 is always close to optimal. This result is expected since Algorithm 1 converges to the optimal base-stock levels w.p.1. Although the best-case performance of Algorithm 3 is always close to optimal, the average and worst-case performances of this algorithm can respectively be up to 9% and 55% worse than the performance of the optimal policy. Therefore, Algorithm 3 may converge to the optimal base-stock levels, but the performance of this algorithm depends on the initial solution and the sequence of the samples of the demand random variables.

Our second set of computational results explore how the performances of Algorithms 1 and 3 change when we choose the initial solutions carefully. We use test problem (10, 0.1, NR) as an example and perturb the mean demand at each time period in this test problem by $\mp \epsilon\%$ to obtain a perturbed test problem (10, 0.1, NR) $^\epsilon$. We compute the optimal base-stock levels for test problem (10, 0.1, NR) $^\epsilon$ and use these base-stock levels as the initial solution when computing the optimal-base stock levels for test problem (10, 0.1, NR). Figure 2 shows the performances of Algorithms 1 and 3 on test problem

Problem	OP	AV ^{A1}	MX ^{A1}	MI ^{A1}	$\frac{AV^{A1}}{OP}$	$\frac{MX^{A1}}{OP}$	AV ^{A3}	MX ^{A3}	MI ^{A3}	$\frac{AV^{A3}}{OP}$	$\frac{MX^{A3}}{OP}$
(5, 0.1, NR)	10.47	10.48	10.48	10.47	100.04	100.08	10.73	12.10	10.47	102.51	115.54
(5, 0.1, UN)	7.73	7.73	7.73	7.73	100.03	100.07	8.04	9.99	7.73	104.03	129.29
(5, 0.1, EX)	19.83	19.84	19.86	19.83	100.06	100.13	19.85	19.87	19.83	100.08	100.18
(5, 0.1, BT)	0.36	0.36	0.36	0.36	100.45	101.59	0.37	0.45	0.36	102.71	127.72
(5, 0.25, NR)	12.92	12.93	12.94	12.92	100.04	100.11	13.05	16.10	12.92	101.01	124.61
(5, 0.25, UN)	9.13	9.13	9.14	9.13	100.07	100.17	9.54	13.70	9.13	104.51	150.15
(5, 0.25, EX)	25.11	25.12	25.13	25.11	100.03	100.06	25.12	25.13	25.11	100.02	100.05
(5, 0.25, BT)	0.41	0.41	0.41	0.41	100.64	101.82	0.42	0.63	0.41	102.58	155.05
(10, 0.1, NR)	18.47	18.47	18.48	18.47	100.03	100.07	19.52	21.88	18.47	105.68	118.48
(10, 0.1, UN)	17.17	17.18	17.19	17.17	100.02	100.12	18.81	22.57	17.17	109.54	131.40
(10, 0.1, EX)	36.25	36.25	36.27	36.25	100.02	100.07	36.58	37.33	36.25	100.93	102.99
(10, 0.1, BT)	0.58	0.59	0.59	0.59	101.02	101.92	0.62	0.72	0.58	105.88	123.16
(10, 0.25, NR)	23.32	23.32	23.34	23.32	100.04	100.09	24.69	27.12	23.32	105.88	116.31
(10, 0.25, UN)	20.95	20.96	20.99	20.95	100.06	100.21	22.57	27.17	20.95	107.75	129.74
(10, 0.25, EX)	46.64	46.65	46.67	46.64	100.03	100.07	46.72	48.59	46.64	100.19	104.19
(10, 0.25, BT)	0.68	0.69	0.72	0.69	101.34	104.66	0.71	0.88	0.69	104.28	128.91

Table 1: Computational results for the multi-period newsvendor problem with backlogged demands.

(10, 0.1, NR) for 10 different runs starting from the initial solutions obtained by letting $\epsilon = 50$, $\epsilon = 75$ and $\epsilon = 100$. If we have $\epsilon = 50$, then both algorithms converge to the optimal base-stock levels for all runs and their performances are essentially identical. If we have $\epsilon = 75$, then both algorithms converge to the optimal base-stock levels for all runs, but the convergence behavior of Algorithm 3 is somewhat erratic. If we have $\epsilon = 100$, then Algorithm 1 converges to the optimal base-stock levels for all runs, but this is not the case for Algorithm 3. Therefore, if ϵ is small and the initial solution is close to the optimal base-stock levels, then the performance of Algorithm 3 may be quite good. This explains the success of the existing stochastic approximation methods in the literature, at least to a certain extent. On the other hand, if ϵ is large and the initial solution is far from the optimal base-stock levels, then Algorithm 3 may converge to different base-stock levels for different runs.

7.2 MULTI-PERIOD NEWSVENDOR PROBLEM WITH LOST SALES

This section assumes that the unsatisfied demand is immediately lost. If we follow a policy characterized by the base-stock levels $\{r_t : t = 1, \dots, \tau\}$ starting with the initial inventory position x_1 and the demands over the planning horizon turn out to be $\{d_t : t = 1, \dots, \tau\}$, then the inventory position at time period t is given by

$$x_t = \max \left\{ x_1 - \sum_{s=1}^{t-1} d_s, r_1 - \sum_{s=1}^{t-1} d_s, \dots, r_{t-1} - \sum_{s=t-1}^{t-1} d_s, 0 \right\}.$$

This is easy to see by noting that the inventory position at time period $t + 1$ is $[\max\{x_t, r_t\} - d_t]^+$ and using induction over the time periods. In this case, we can modify (36)-(38), (40)-(42) and Algorithm 3 in a straightforward manner to come up with a stochastic approximation method to solve problem (39) under the assumption that the unsatisfied demand is immediately lost.

Our computational results are summarized in Table 2. The entries in this table have the same interpretations as the ones in Table 1. Similar to our computational results in Table 1, even the worst-case performance of Algorithm 1 is always close to optimal. Although the best-case performance of

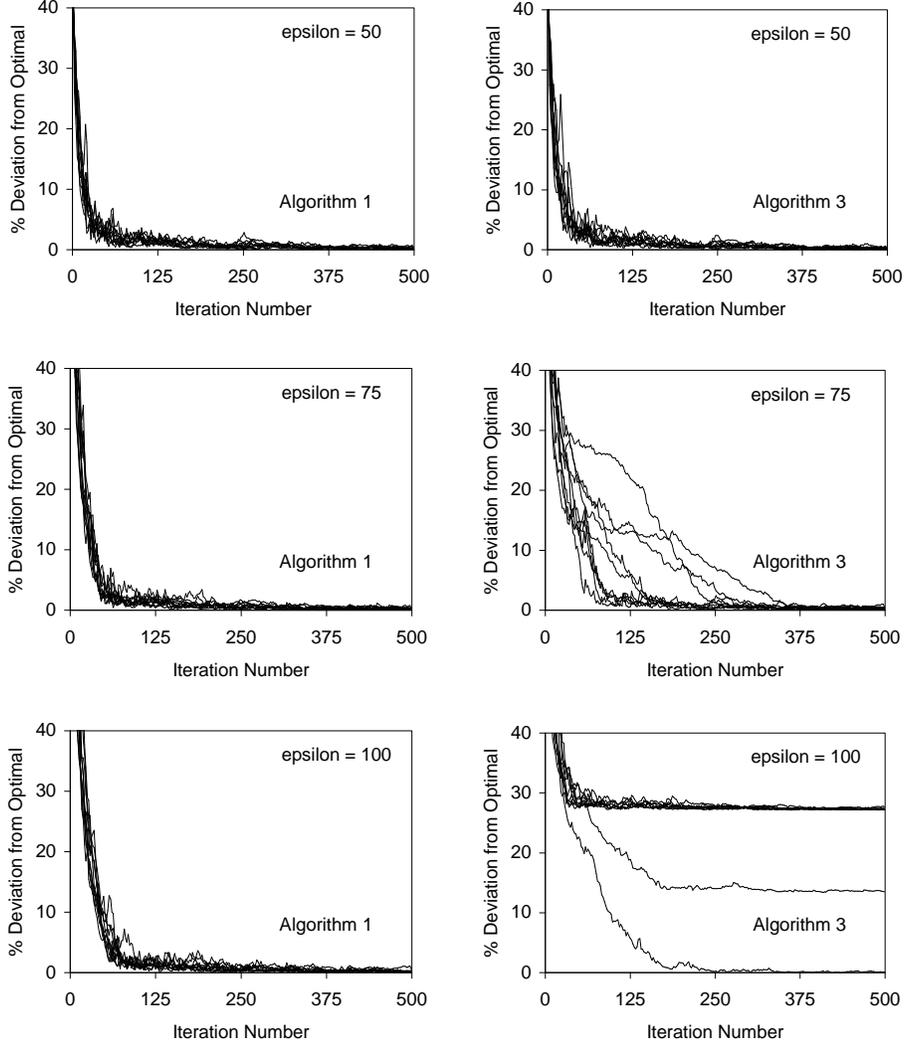


Figure 2: Performances of Algorithms 1 and 3 on test problem $(10, 0.1, \text{NR})$ for 10 different runs starting from the initial solutions that are chosen carefully.

Algorithm 3 is always close to optimal, the average and worst-case performances of this algorithm can respectively be up to 8% and 47% worse than the performance of the optimal policy.

7.3 INVENTORY PURCHASING PROBLEM UNDER PRICE UNCERTAINTY

We consider a policy characterized by the base-stock levels $\{r_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$. That is, if the total amount of product purchased up to time period t is x_t and the price of the product is p_t , then this policy purchases $[r_t(p_t) - x_t]^+$ units. If we follow this policy starting with the initial inventory position x_1 and the prices over the planning horizon turn out to be $\{p_t : t = 1, \dots, \tau\}$, then the inventory position at time period t is given by $x_t = \max\{x_1, r_1(p_1), \dots, r_{t-1}(p_{t-1})\}$. This can be seen by noting that the inventory position at time period $t + 1$ is $\max\{x_t, r_t(p_t)\}$ and using induction over the time

Problem	OP	AV ^{A1}	MX ^{A1}	MI ^{A1}	$\frac{AV^{A1}}{OP}$	$\frac{MX^{A1}}{OP}$	AV ^{A3}	MX ^{A3}	MI ^{A3}	$\frac{AV^{A3}}{OP}$	$\frac{MX^{A3}}{OP}$
(5, 0.1, NR)	10.30	10.30	10.31	10.30	100.00	100.04	10.57	11.96	10.30	102.57	116.13
(5, 0.1, UN)	7.66	7.66	7.66	7.66	100.01	100.04	7.96	9.83	7.66	103.95	128.30
(5, 0.1, EX)	18.82	18.83	18.84	18.82	100.02	100.07	18.83	18.87	18.82	100.05	100.25
(5, 0.1, BT)	0.35	0.35	0.36	0.35	100.18	101.03	0.36	0.45	0.35	101.64	127.09
(5, 0.25, NR)	12.52	12.52	12.52	12.52	100.02	100.05	12.65	15.78	12.52	101.06	126.05
(5, 0.25, UN)	8.93	8.93	8.94	8.93	100.02	100.13	9.32	13.21	8.93	104.37	147.87
(5, 0.25, EX)	23.23	23.23	23.24	23.23	100.01	100.05	23.23	23.24	23.23	100.00	100.04
(5, 0.25, BT)	0.40	0.40	0.40	0.40	100.04	100.94	0.40	0.40	0.40	100.00	100.80
(10, 0.1, NR)	18.06	18.07	18.07	18.06	100.01	100.06	19.06	21.33	18.06	105.53	118.07
(10, 0.1, UN)	16.97	16.97	16.97	16.97	100.01	100.02	18.42	22.12	16.97	108.54	130.35
(10, 0.1, EX)	33.60	33.60	33.61	33.60	100.02	100.05	33.94	34.68	33.60	101.02	103.23
(10, 0.1, BT)	0.57	0.58	0.58	0.58	100.87	101.37	0.60	0.66	0.58	104.26	115.46
(10, 0.25, NR)	22.36	22.36	22.37	22.36	100.00	100.03	23.64	25.93	22.36	105.74	115.99
(10, 0.25, UN)	20.36	20.36	20.38	20.36	100.03	100.10	21.06	26.19	20.36	103.46	128.66
(10, 0.25, EX)	42.07	42.08	42.09	42.07	100.01	100.04	42.08	42.09	42.07	100.01	100.04
(10, 0.25, BT)	0.68	0.68	0.68	0.68	100.00	100.49	0.68	0.81	0.68	100.17	120.24

Table 2: Computational results for the multi-period newsvendor problem with lost sales.

periods. In this case, the purchasing cost that we incur at time period t is

$$\begin{aligned}
C_t(x_1, p | r) &= p_t [r_t(p_t) - x_t]^+ = p_t \max \{r_t(p_t) - \max \{x_1, r_1(p_1), \dots, r_{t-1}(p_{t-1})\}, 0\} \\
&= p_t \max \{ \min \{r_t(p_t) - x_1, r_t(p_t) - r_1(p_1), \dots, r_t(p_t) - r_{t-1}(p_{t-1})\}, 0 \}, \quad (43)
\end{aligned}$$

where we use p to denote the prices $\{p_t : t = 1, \dots, \tau\}$ and r to denote the base-stock levels $\{r_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$. On the other hand, if the demand turns out to be d , then the penalty cost that we incur at the end of the planning horizon is

$$\begin{aligned}
B_{\tau+1}(x_1, p, d | r) &= b [d - x_{\tau+1}]^+ = b \max \{d - \max \{x_1, r_1(p_1), \dots, r_\tau(p_\tau)\}, 0\} \\
&= b \max \{ \min \{d - x_1, d - r_1(p_1), \dots, d - r_\tau(p_\tau)\}, 0 \}. \quad (44)
\end{aligned}$$

Therefore, we can try to solve the problem $\min_r \mathbb{E} \{ \sum_{t=1}^{\tau} C_t(x_1, p | r) + B_{\tau+1}(x_1, p, d | r) \}$ to compute the optimal base-stock levels. Similar to the situation in Section 7.1, it is easy to check that the objective function of this problem is not necessarily differentiable with respect to r and we perturb the base-stock levels by using the random variables $\{\zeta_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ that are uniformly distributed over the small interval $[0, \epsilon]$. Consequently, we solve the problem

$$\min_r \mathbb{E} \left\{ \sum_{t=1}^{\tau} C_t(x_1, p | r + \zeta) + B_{\tau+1}(x_1, p, d | r + \zeta) \right\}, \quad (45)$$

where we use $r + \zeta$ to denote the perturbed base-stock levels $\{r_t(\hat{p}_t) + \zeta_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$. It is now possible to show that the objective function of problem (45) is differentiable with respect to r and its gradient is Lipschitz continuous. This implies that we can use a standard stochastic approximation method to solve problem (45).

Similar to (40)-(42), after straightforward algebraic manipulations on (43) and (44), and some simplifications, it is easy to see that the $r_s(p_s)$ -th component in the gradient of $C_t(x_1, p | r)$ with respect

to r is given by

$$\nabla_s C_t(x_1, p | r) = \begin{cases} -p_t \mathbf{1}(r_t(p_t) - r_s(p_s) \geq 0) \times \mathbf{1}(r_s(p_s) - x_1 \geq 0) \\ \quad \times \mathbf{1}(r_s(p_s) - r_1(p_1) \geq 0) \times \dots \times \mathbf{1}(r_s(p_s) - r_{t-1}(p_{t-1}) \geq 0) & \text{if } s < t \\ p_t \mathbf{1}(r_t(p_t) - x_1 \geq 0) \\ \quad \times \mathbf{1}(r_t(p_t) - r_1(p_1) \geq 0) \times \dots \times \mathbf{1}(r_t(p_t) - r_{t-1}(p_{t-1}) \geq 0) & \text{if } s = t \\ 0 & \text{if } s > t, \end{cases}$$

whereas the $r_s(p_s)$ -th component in the gradient of $B_{\tau+1}(x_1, p, d | r)$ with respect to r is given by

$$\begin{aligned} \nabla_s B_{\tau+1}(x_1, p, d | r) = & -b \mathbf{1}(d - r_s(p_s) \geq 0) \times \mathbf{1}(r_s(p_s) - x_1 \geq 0) \\ & \times \mathbf{1}(r_s(p_s) - r_1(p_1) \geq 0) \times \dots \times \mathbf{1}(r_s(p_s) - r_\tau(p_\tau) \geq 0). \end{aligned}$$

The remarks for (40)-(42) also hold here. In particular, the gradients of $C_t(x_1, p | r)$ or $B_{\tau+1}(x_1, p, d | r)$ do not exist everywhere, but the gradients of $C_t(x_1, p | r + \zeta)$ and $B_{\tau+1}(x_1, p, d | r + \zeta)$ exist everywhere w.p.1. Consequently, the following algorithm is a standard stochastic approximation method for solving problem (45).

Algorithm 4

Step 1. Initialize the estimates of the optimal base-stock levels $\{r_t^1(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{p_t^k : t = 1, \dots, \tau\}$ be the price random variables, d^k be the demand random variable and $\{\zeta_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t, t = 1, \dots, \tau\}$ be the perturbation random variables at iteration k , set

$$r_t^{k+1}(p_t^k) = r_t^k(p_t^k) - \alpha^k \left\{ \sum_{s=1}^{\tau} \nabla_t C_s(x_1, p^k | r^k + \zeta^k) + \nabla_t B_{\tau+1}(x_1, p^k, d^k | r^k + \zeta^k) \right\}$$

for all $t = 1, \dots, \tau$. Furthermore, set $r_t^{k+1}(\hat{p}_t) = r_t^k(\hat{p}_t)$ for all $\hat{p}_t \in \mathcal{P}_t \setminus \{p_t^k\}$, $t = 1, \dots, \tau$.

Step 3. Increase k by 1 and go to Step 2.

In our test problems, the penalty cost b is an integer. The price at each time period is uniformly distributed over the integers $\{1, \dots, b\}$ and the demand is uniformly distributed over the interval $[0, 1]$. Our experimental setup is the same as the one in Section 7.1 and our computational results are summarized in Table 3. The first column in this table shows the problem parameters by using the pairs $(\tau, b) \in \{1, 2, 4, 6\} \times \{2, 5, 10\}$, where τ is the number of time periods and b is the penalty cost. The other entries in this table have the same interpretations as the ones in Table 1.

Similar to our computational results in Tables 1 and 2, even the worst-case performance of Algorithm 2 is always close to optimal. However, it is interesting to note that even the best-case performance of Algorithm 4 can be significantly worse than the performance of the optimal policy.

8 CONCLUSIONS

We proposed three stochastic approximation methods to compute the optimal base-stock levels in three problem classes for which the base-stock policies are known to be optimal. The proposed methods

Problem	OP	AV ^{A2}	MX ^{A2}	MI ^{A2}	$\frac{AV^{A2}}{OP}$	$\frac{MX^{A2}}{OP}$	AV ^{A4}	MX ^{A4}	MI ^{A4}	$\frac{AV^{A4}}{OP}$	$\frac{MX^{A4}}{OP}$
(1, 2)	0.88	0.88	0.89	0.88	100.16	100.74	0.88	0.88	0.88	100.16	100.64
(1, 5)	1.92	1.93	1.94	1.92	100.39	101.10	1.93	1.95	1.92	100.42	101.44
(1, 10)	3.56	3.61	3.66	3.58	101.25	102.58	3.61	3.65	3.57	101.28	102.34
(2, 2)	0.79	0.79	0.80	0.79	100.35	101.73	0.84	0.85	0.84	106.92	107.58
(2, 5)	1.60	1.62	1.64	1.61	100.85	101.96	1.89	1.90	1.75	117.60	118.64
(2, 10)	2.90	2.96	3.04	2.92	101.74	104.74	3.48	3.65	3.24	119.95	125.64
(4, 2)	0.78	0.78	0.80	0.78	100.63	102.55	0.89	0.89	0.88	113.66	114.54
(4, 5)	1.34	1.36	1.37	1.34	100.94	101.99	1.84	1.90	1.60	137.07	141.13
(4, 10)	2.38	2.44	2.50	2.40	102.62	104.98	3.48	3.81	3.14	146.51	160.36
(6, 2)	0.76	0.77	0.77	0.76	100.65	101.86	0.89	0.89	0.88	116.50	116.99
(6, 5)	1.16	1.18	1.19	1.17	101.32	102.23	1.85	1.90	1.62	159.33	163.74
(6, 10)	2.02	2.07	2.11	2.04	102.74	104.72	3.21	3.55	2.88	158.99	176.20

Table 3: Computational results for the inventory purchasing problem under price uncertainty.

enjoy the well-known advantages of the stochastic approximation methods. They work with samples of the random variables, eliminating the need to compute expectations explicitly. Furthermore, they remain applicable when the demand information is censored by the amount of available inventory. The iterates of the proposed methods converge to the optimal base-stock levels, but this is not guaranteed for standard stochastic approximation methods, such as Algorithms 3 and 4.

One can unify the approaches in Sections 3, 4 and 6 to a certain extent. Equations (5), (22) and (30) characterize the first order conditions that must be satisfied by the optimal base-stock levels. However, finding a solution to these first order conditions through stochastic approximation methods requires knowing the function $\dot{v}_{t+1}(\cdot)$. In (10), (26) and (34), we come up with recursive expressions that can be used to compute the stochastic gradients of $\{v_t(\cdot) : t = 1, \dots, \tau\}$. At iteration k , we “mimic” these expressions by using the estimates of the optimal base-stock levels as in (11), (27) and (35). The convergence proofs are based on analyzing the error function.

Our results easily extend to other settings, such as the revenue management problems for which the booking-limit policies are known to be optimal. On the other hand, our analysis in this paper strictly exploits the fact that the base-stock policies are optimal for the problem classes that we consider. It is not yet clear what advantages our analysis can provide for the problem classes where the base-stock policies are not necessarily optimal, but we only look for a good set of base-stock levels.

REFERENCES

- Arrow, K., Karlin, S. and Scarf, H. (1958), *Studies in the Mathematical Theory of Inventory and Production*, Stanford University Press, Stanford, CA.
- Azoury, K. S. (1985), ‘Bayes solutions to dynamic inventory models under unknown demand distribution’, *Management Science* **31**(9), 1150–1160.
- Bashyam, S. and Fu, M. C. (1998), ‘Optimization of (s, S) inventory systems with random lead times and a service level constraint’, *Management Science* **44**(12), 243–256.
- Bertsekas, D. P. and Tsitsiklis, J. N. (1996), *Neuro-Dynamic Programming*, Athena Scientific, Belmont, MA.
- Braden, D. B. and Freimer, M. (1991), ‘Informational dynamics of censored observations’, *Management Science* **37**(11), 1390–1404.

- Conrad, S. A. (1976), ‘Sales data and the estimation of demand’, *Operations Research Quarterly* **27**(1), 123–127.
- Ding, X. (2002), Estimation and optimization in discrete inventory models, PhD thesis, The University of British Columbia, Vancouver, British Columbia.
- Fu, M. (1994), ‘Sample path derivatives for (s, S) inventory systems’, *Operations Research* **42**(2), 351–363.
- Glasserman, P. (1994), Perturbation analysis of production networks, in D. D. Yao, ed., ‘*Stochastic Modeling and Analysis of Manufacturing Systems*’, Springer, New York, NY.
- Glasserman, P. and Tayur, S. (1995), ‘Sensitivity analysis for base-stock levels in multiechelon production-inventory systems’, *Management Science* **41**(2), 263–281.
- Godfrey, G. A. and Powell, W. B. (2001), ‘An adaptive, distribution-free approximation for the newsvendor problem with censored demands, with applications to inventory and distribution problems’, *Management Science* **47**(8), 1101–1112.
- Iglehart, D. L. (1964), ‘The dynamic inventory problem with unknown demand distribution’, *Management Science* **10**(3), 429–440.
- Karaesmen, I. and van Ryzin, G. (2004), ‘Overbooking with substitutable inventory classes’, *Operations Research* **52**(1), 83–104.
- Kushner, H. J. and Clark, D. S. (1978), *Stochastic Approximation Methods for Constrained and Unconstrained Systems*, Springer-Verlag, Berlin.
- L’Ecuyer, P. and Glynn, P. (1994), ‘Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady state’, *Management Science* **40**, 1245–1261.
- Levi, R., Roundy, R. and Shmoys, D. M. (2005), Computing provably near-optimal sample-based policies for stochastic inventory control models, Technical report, Cornell University, School of Operations Research and Industrial Engineering.
- Mahajan, S. and van Ryzin, G. (2001), ‘Stocking retail assortments under dynamic customer substitution’, *Operations Research* **49**(3), 334–351.
- Nascimento, J. M. and Powell, W. B. (2006), An optimal approximate dynamic programming algorithm for the lagged asset acquisition problem, Technical report, Department of Operations Research and Financial Engineering, Princeton University.
- Neveu, J. (1975), *Discrete Parameter Martingales*, North Holland, Amsterdam.
- Porteus, E. L. (1990), *Handbooks in Operations Research and Management Science: Stochastic Models*, Vol. 2, North Holland, Amsterdam, chapter Stochastic Inventory Theory, pp. 605–652.
- Powell, W. B., Ruszczyński, A. and Topaloglu, H. (2004), ‘Learning algorithms for separable approximations of stochastic optimization problems’, *Mathematics of Operations Research* **29**(4), 814–836.
- Scarf, H. (1960), ‘Some remarks on Bayes solutions to the inventory problem’, *Naval Research Logistics Quarterly* **7**, 591–596.
- Sutton, R. S. (1988), ‘Learning to predict by the methods of temporal differences’, *Machine Learning* **3**, 9–44.
- Topaloglu, H. and Powell, W. B. (2003), ‘An algorithm for approximating piecewise linear functions from sample gradients’, *Operations Research Letters* **31**, 66–76.
- Tsitsiklis, J. N. (1994), ‘Asynchronous stochastic approximation and Q-learning’, *Machine Learning* **16**, 185–202.
- van Ryzin, G. and Vulcano, G. (2006), ‘Simulation-based optimization of virtual nesting controls for network revenue management’, *Operations Research* (to appear).
- Zipkin, P. (2000), *Foundations of Inventory Management*, McGraw-Hill, Boston, MA.

9 APPENDIX

This section completes the proof of Lemma 7 by showing that $\limsup_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| = 0$. For $\epsilon > 0$, we call $\{k', k' + 1, \dots, k''\}$ as an upcrossing interval from $\epsilon/2$ to ϵ , if we have $|\dot{f}_t(r_t^{k'})| < \epsilon/2$, $|\dot{f}_t(r_t^{k''})| > \epsilon$ and $\epsilon/2 \leq |\dot{f}_t(r_t^k)| \leq \epsilon$ for all $k = k' + 1, \dots, k'' - 1$. We show that there exist only a finite number of upcrossing intervals from $\epsilon/2$ to ϵ . Since we have $\liminf_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| = 0$, this implies that $\limsup_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| \leq \epsilon$.

To show the result by contradiction, we fix $\epsilon > 0$ and assume that the number of upcrossing intervals from $\epsilon/2$ to ϵ is infinite. We let $\{k'_n, k'_n + 1, \dots, k''_n\}$ be the n -th upcrossing interval. We have

$$\begin{aligned} |\dot{f}_t(r_t^{k'_n+1})| - |\dot{f}_t(r_t^{k'_n})| &\leq |\dot{f}_t(r_t^{k'_n+1}) - \dot{f}_t(r_t^{k'_n})| \\ &\leq L |r_t^{k'_n+1} - r_t^{k'_n}| = \alpha^{k'_n} L |s_t^{k'_n}(r_t^{k'_n}, d_t^{k'_n}, \dots, d_\tau^{k'_n})| \leq LM \alpha^{k'_n}, \end{aligned}$$

where we use Lemmas 1 and 2. Since $|\dot{f}_t(r_t^{k'_n+1})| \geq \epsilon/2$ and $\lim_{k \rightarrow \infty} \alpha^k = 0$, the chain of inequalities above imply that there exists a finite number \hat{N} such that we have $|\dot{f}_t(r_t^{k'_n})| \geq \epsilon/4$ for all $n = \hat{N}, \hat{N} + 1, \dots$. Since $\lim_{k \rightarrow \infty} \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} = 0$ by the argument at the beginning of the proof of Lemma 7, there exists a finite number \tilde{N} such that we have $\mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} \leq \epsilon/8$ for all $k = k'_{\tilde{N}}, k'_{\tilde{N}} + 1, \dots$. Therefore, letting $N = \max\{\hat{N}, \tilde{N}\}$, we have

$$|\dot{f}_t(r_t^{k'_n})| \geq \epsilon/4 \quad \text{and} \quad \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} \leq \epsilon/8 \quad (46)$$

for all $n = N, N + 1, \dots, k = k'_N, k'_N + 1, \dots$

On the other hand, using Lemmas 1 and 2, we have

$$\epsilon/2 \leq |\dot{f}_t(r_t^{k''_n})| - |\dot{f}_t(r_t^{k'_n})| \leq |\dot{f}_t(r_t^{k''_n}) - \dot{f}_t(r_t^{k'_n})| \leq L |r_t^{k''_n} - r_t^{k'_n}| \leq L \sum_{k=k'_n}^{k''_n-1} \alpha^k |s_t^k(r_t^k, d_t^k, \dots, d_\tau^k)|,$$

which implies that $\sum_{k=k'_n}^{k''_n-1} \alpha^k \geq \epsilon/[2LM]$. Therefore, using (46), we obtain

$$\sum_{n=N}^{\infty} \sum_{k=k'_n}^{k''_n-1} \alpha^k |\dot{f}_t(r_t^k)|^2 \geq \sum_{n=N}^{\infty} \frac{\epsilon^2}{16} \sum_{k=k'_n}^{k''_n-1} \alpha^k \geq \sum_{n=N}^{\infty} \frac{\epsilon^3}{32LM} = \infty. \quad (47)$$

We have $|\dot{f}_t(r_t^k)|/2 \geq \epsilon/8 \geq \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}$ for all $n = N, N + 1, \dots, k = k'_n, k'_n + 1, \dots, k''_n - 1$ by (46) and the definition of an upcrossing interval. This implies that

$$|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} \geq |\dot{f}_t(r_t^k)|^2 - |\dot{f}_t(r_t^k)| \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} \geq |\dot{f}_t(r_t^k)|^2 - |\dot{f}_t(r_t^k)|^2/2$$

for all $n = N, N + 1, \dots, k = k'_n, k'_n + 1, \dots, k''_n - 1$. Using this chain of inequalities, since (A.2) holds for time period t by Lemma 6, we obtain

$$\begin{aligned} \sum_{n=N}^{\infty} \sum_{k=k'_n}^{k''_n-1} \alpha^k |\dot{f}_t(r_t^k)|^2/2 &\leq \sum_{n=N}^{\infty} \sum_{k=k'_n}^{k''_n-1} \alpha^k [|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}] \\ &\leq \sum_{k=1}^{\infty} \alpha^k [|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k) \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}]^+ < \infty. \end{aligned}$$

This expression contradicts (47). Therefore, we must have a finite number of upcrossing intervals from $\epsilon/2$ to ϵ , which implies that $\limsup_{k \rightarrow \infty} |\dot{f}_t(r_t^k)| \leq \epsilon$. Since ϵ is arbitrary, we obtain the desired result. \square

This section shows that the stochastic approximation methods that we propose for the multi-period newsvendor problem with lost sales and the inventory purchasing problem under price uncertainty converge to the optimal base-stock levels w.p.1.

10.1 MULTI-PERIOD NEWSVENDOR PROBLEM WITH LOST SALES AND STATIONARY COST PARAMETERS

We consider the setting described in Section 4 and show that the error function $e_t^k(x_t) = \dot{v}_t(x_t) \mathbf{1}(x_t > 0) - \mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \mathbf{1}(x_t > 0) \}$ satisfies the same bound given in Lemma 4. Once we have this bound, we can follow the same induction argument in Proposition 5, Lemmas 6-8 and Proposition 9 to show that the stochastic approximation method that we propose for the multi-period newsvendor problem with lost sales converges to the optimal base-stock levels w.p.1. The following lemma is analogous to Lemma 3.

Lemma 11 *If \hat{x}_t, \tilde{x}_t satisfy $\hat{x}_t \leq \tilde{x}_t$, then we have $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, \dots, d_\tau^k) \} \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, \dots, d_\tau^k) \}$ w.p.1 for all $t = 1, \dots, \tau, k = 1, 2, \dots$*

Proof We show the result by induction over the time periods. Since (11) and (27) reduce to the same expression for time period τ , we can show that the result holds for time period τ by following the argument in the proof of Lemma 3. Furthermore, we have $\mathbb{E}_k \{ \xi_\tau^k(x_\tau, d_\tau^k) \} \geq -b$ for all $x_\tau \in \mathbb{R}$ by (27). Assuming that the result holds for time period $t+1$ and we have $\mathbb{E}_k \{ \xi_{t+1}^k(x_{t+1}, d_{t+1}^k, \dots, d_\tau^k) \} \geq -b$ for all $x_{t+1} \in \mathbb{R}$, we now show that the result holds for time period t and we have $\mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \} \geq -b$ for all $x_t \in \mathbb{R}$. We consider three cases. First, we assume that $r_t^k \leq \hat{x}_t \leq \tilde{x}_t$. We investigate the conditional expectation $\mathbb{E}_k \{ \xi_t^k(\cdot, d_t^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \}$, where ϕ_t is a known constant, by examining the following three subcases.

Case 1.a. Assume that $\phi_t < \hat{x}_t$. Since we have $\phi_t < \hat{x}_t \leq \tilde{x}_t$, (27) implies that $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h + \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$. Taking expectations conditional on $d_t^k = \phi_t$ and noting the fact that the demand random variables at different time periods are independent, we obtain $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} = h + \mathbb{E}_k \{ \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \}$ and $\mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} = h + \mathbb{E}_k \{ \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \}$. Thus, the induction hypothesis implies that $\mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} \geq \mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} \geq -b$.

Case 1.b. Assume that $\hat{x}_t \leq \phi_t < \tilde{x}_t$. We have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = -b$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$ by (27). Taking expectations conditional on $d_t^k = \phi_t$, the induction hypothesis implies that $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} = -b \leq h + \mathbb{E}_k \{ \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \} = \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \}$.

Case 1.c. Assume that $\phi_t \geq \tilde{x}_t$. In this case, we have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = -b = \xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k)$. Taking expectations conditional on $d_t^k = \phi_t$, we obtain $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \} = -b = \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mid d_t^k = \phi_t \}$.

The three subcases above show that if $r_t^k \leq \hat{x}_t \leq \tilde{x}_t$, then we have $\mathbb{E}_k\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k\} \geq \mathbb{E}_k\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k\} \geq -b$. Taking expectations, we obtain $\mathbb{E}_k\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} \geq \mathbb{E}_k\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k)\} \geq -b$.

It can be shown that the result holds for time period t by considering the other two cases where we have $\hat{x}_t < r_t^k \leq \tilde{x}_t$ or $\hat{x}_t \leq \tilde{x}_t < r_t^k$. This completes the induction argument. \square

We now show that the error function $e_t^k(x_t) = \dot{v}_t(x_t) \mathbf{1}(x_t > 0) - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \mathbf{1}(x_t > 0)\}$ satisfies the same bound given in Lemma 4. If $x_t \leq 0$, then we have $e_t^k(x_t) = 0$ and the bound immediately holds. Therefore, we assume that $x_t > 0$ for the rest of the discussion. Using (22) and (27), we obtain

$$\begin{aligned} \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} &= \begin{cases} h \mathbb{P}_k\{d_t^k < x_t\} - b \mathbb{P}_k\{d_t^k \geq x_t\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t)\} & \text{if } x_t \geq r_t^k \\ h \mathbb{P}_k\{d_t^k < r_t^k\} - b \mathbb{P}_k\{d_t^k \geq r_t^k\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k)\} & \text{if } x_t < r_t^k \end{cases} \\ &= \begin{cases} \dot{f}_t(x_t) - c - \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k) \mathbf{1}(d_t^k < x_t)\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t)\} & \text{if } x_t \geq r_t^k \\ \dot{f}_t(r_t^k) - c - \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k)\} \\ \quad + \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k)\} & \text{if } x_t < r_t^k. \end{cases} \end{aligned} \quad (48)$$

Since r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $\dot{f}_t(r_t^*) = 0$. Using (22) and (25), we obtain

$$\dot{v}_t(x_t) = \begin{cases} \dot{f}_t(x_t) - c & \text{if } x_t \geq r_t^* \\ -c & \text{if } x_t < r_t^*. \end{cases} \quad (49)$$

We consider four cases. First, we assume that $x_t \geq r_t^k$ and $x_t \geq r_t^*$. Using (48) and (49), we have $e_t^k(x_t) = \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k) \mathbf{1}(d_t^k < x_t)\} - \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t)\} = \mathbb{E}_k\{e_{t+1}^k(x_t - d_t^k)\}$. Therefore, we obtain $|e_t^k(x_t)| \leq \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\}$ by Jensen's inequality.

Second, we assume that $x_t \geq r_t^k$ and $x_t < r_t^*$. We have $\mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \geq \mathbb{E}_k\{\xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\}$ by Lemma 11. Using this inequality, (48) and (49), we obtain

$$\begin{aligned} e_t^k(x_t) &= -c - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \leq -c - \mathbb{E}_k\{\xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k)\} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k\{\dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k)\} - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k)\} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}. \end{aligned}$$

Since $x_t < r_t^*$ and r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $\dot{f}_t(x_t) \leq 0$. Using (48), we also obtain

$$\begin{aligned} e_t^k(x_t) &= -c - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k)\} \\ &= -\dot{f}_t(x_t) + \mathbb{E}_k\{\dot{v}_{t+1}(x_t - d_t^k) \mathbf{1}(d_t^k < x_t)\} - \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t)\} \\ &\geq \mathbb{E}_k\{e_{t+1}^k(x_t - d_t^k)\}. \end{aligned}$$

The last two chains of inequalities imply that

$$|e_t^k(x_t)| \leq \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|, \mathbb{E}_k \{ |e_{t+1}^k(x_t - d_t^k)| \} \right\}.$$

Third, we assume that $x_t < r_t^k$ and $x_t \geq r_t^*$. Since $f_t(\cdot)$ is convex, we have $\dot{f}_t(r_t^k) \geq \dot{f}_t(x_t) \geq \dot{f}_t(r_t^*) = 0$. Using (48) and (49), we obtain

$$\begin{aligned} e_t^k(x_t) &= \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k \{ \dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k) \} - \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k) \} \\ &= \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}, \end{aligned}$$

which implies that $-\dot{f}_t(r_t^k) + \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \leq e_t^k(x_t) \leq \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}$. Therefore, we have

$$|e_t^k(x_t)| \leq \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|, \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \} \right\}.$$

Fourth, we assume that $x_t < r_t^k$ and $x_t < r_t^*$. In this case, (48) and (49) imply that

$$\begin{aligned} e_t^k(x_t) &= -\dot{f}_t(r_t^k) + \mathbb{E}_k \{ \dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k) \} - \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k) \} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}. \end{aligned}$$

Therefore, we have $|e_t^k(x_t)| = \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|$. The result follows by combining the four cases.

10.2 MULTI-PERIOD NEWSVENDOR PROBLEM WITH LOST SALES AND NONSTATIONARY COST PARAMETERS

This section shows how to extend the ideas in Section 4 to the case where the cost parameters are nonstationary. We use c_t , h_t and b_t to respectively denote the per unit replenishment, holding and penalty costs at time period t . In addition to the earlier assumptions for the demand random variables and the lead times, we also assume that the cost parameters satisfy $b_t > c_{t+1} \geq 0$ and $h_t \geq 0$ for all $t = 1, \dots, \tau$, with $c_{\tau+1} = 0$. This assumption is standard and holds in many applications since the per unit penalty cost is usually much higher than the per unit replenishment cost. Intuitively, this assumption ensures that it is never optimal to hold inventory to satisfy the future demand while leaving the demand in the current time period unsatisfied. Under this assumption, it is possible to show that the functions $\{v_t(\cdot) : t = 1, \dots, \tau\}$ and $\{f_t(\cdot) : t = 1, \dots, \tau\}$ defined in Section 4 are convex.

The functions $\{v_t(\cdot) : t = 1, \dots, \tau\}$ satisfy the Bellman equations

$$v_t(x_t) = \min_{y_t \geq x_t} c_t [y_t - x_t] + \mathbb{E} \{ h_t [y_t - d_t]^+ + b_t [d_t - y_t]^+ + v_{t+1}([y_t - d_t]^+) \},$$

with $v_{\tau+1}(\cdot) = 0$. We also let

$$f_t(r_t) = c_t r_t + \mathbb{E} \{ h_t [r_t - d_t]^+ + b_t [d_t - r_t]^+ + v_{t+1}([r_t - d_t]^+) \}.$$

It can be shown that $v_t(\cdot)$ and $f_t(\cdot)$ are positive, Lipschitz continuous, differentiable and convex functions, and $f_t(\cdot)$ has a finite unconstrained minimizer. In this case, the optimal base-stock levels $\{r_t^* : t = 1, \dots, \tau\}$ are the minimizers of the functions $\{f_t(\cdot) : t = 1, \dots, \tau\}$.

Our approach is based on constructing tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$. Since we have

$$\dot{f}_t(r_t) = c_t + h_t \mathbb{P}\{d_t < r_t\} - b_t \mathbb{P}\{d_t \geq r_t\} + \mathbb{E}\{\dot{v}_{t+1}(r_t - d_t) \mathbf{1}(d_t < r_t)\}, \quad (50)$$

we can compute a stochastic gradient of $f_t(\cdot)$ at x_t through

$$\Delta_t(x_t, d_t) = c_t + h_t \mathbf{1}(d_t < x_t) - b_t \mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t) \mathbf{1}(d_t < x_t). \quad (51)$$

On the other hand, (25) implies that

$$\begin{aligned} \dot{v}_t(x_t) &= \begin{cases} \dot{f}_t(x_t) - c_t & \text{if } x_t \geq r_t^* \\ -c_t & \text{if } x_t < r_t^*, \end{cases} \\ &= \begin{cases} h_t \mathbb{P}\{d_t < x_t\} - b_t \mathbb{P}\{d_t \geq x_t\} + \mathbb{E}\{\dot{v}_{t+1}(x_t - d_t) \mathbf{1}(d_t < x_t)\} & \text{if } x_t \geq r_t^* \\ h_t \mathbb{P}\{d_t < r_t^*\} - b_t \mathbb{P}\{d_t \geq r_t^*\} + \mathbb{E}\{\dot{v}_{t+1}(r_t^* - d_t) \mathbf{1}(d_t < r_t^*)\} & \text{if } x_t < r_t^*, \end{cases} \end{aligned} \quad (52)$$

in which case

$$\dot{v}_t(x_t, d_t) = \begin{cases} h_t \mathbf{1}(d_t < x_t) - b_t \mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t) \mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^* \\ h_t \mathbf{1}(d_t < r_t^*) - b_t \mathbf{1}(d_t \geq r_t^*) + \dot{v}_{t+1}(r_t^* - d_t) \mathbf{1}(d_t < r_t^*) & \text{if } x_t < r_t^*, \end{cases} \quad (53)$$

gives a stochastic gradient of $v_t(\cdot)$ at x_t . To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \dots, \tau\}$, we “mimic” the computation in (53) by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k : t = 1, \dots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k , we recursively define

$$\xi_t^k(x_t, d_t, \dots, d_\tau) = \begin{cases} h_t \mathbf{1}(d_t < x_t) - b_t \mathbf{1}(d_t \geq x_t) \\ \quad + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau) \mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^k \\ h_t \mathbf{1}(d_t < r_t^k) - b_t \mathbf{1}(d_t \geq r_t^k) \\ \quad + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \dots, d_\tau) \mathbf{1}(d_t < r_t^k) & \text{if } x_t < r_t^k, \end{cases} \quad (54)$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \dots, \cdot) = 0$. At iteration k , replacing $\dot{v}_{t+1}(x_t - d_t)$ in (51) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau)$, we use

$$s_t^k(x_t, d_t, \dots, d_\tau) = c_t + h_t \mathbf{1}(d_t < x_t) - b_t \mathbf{1}(d_t \geq x_t) + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \dots, d_\tau) \mathbf{1}(d_t < x_t)$$

to approximate the stochastic gradient of $f_t(\cdot)$ at x_t . Thus, we can use Algorithm 1 to search for the optimal base-stock levels. The only difference is that we need to use the step direction above in Step 2.

To establish the convergence of Algorithm 1 to the optimal base-stock levels for the multi-period newsvendor problem with lost sales and nonstationary cost parameters, we analyze the error function defined as

$$e_t^k(x_t) = \dot{v}_t(x_t) \mathbf{1}(x_t > 0) - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \mathbf{1}(x_t > 0)\},$$

with $e_{\tau+1}^k(\cdot) = 0$. We can easily follow the argument in the proof of Lemma 2 to derive bounds on $\xi_t^k(\cdot, d_t^k, \dots, d_\tau^k)$ and $s_t^k(\cdot, d_t^k, \dots, d_\tau^k)$. The following lemma is analogous to Lemma 3.

Lemma 12 *If \hat{x}_t, \tilde{x}_t satisfy $\hat{x}_t \leq \tilde{x}_t$, then we have*

$$\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, \dots, d_\tau^k) \} \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, \dots, d_\tau^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(\tilde{x}_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \quad (55)$$

w.p.1 for all $t = 1, \dots, \tau$, $k = 1, 2, \dots$

Proof We show the result by induction over the time periods. It is easy to show that the result holds for time period τ by following the corresponding argument in the proof of Lemma 3. Assuming that the result holds for time period $t + 1$, we now show that the result holds for time period t . We consider three cases. First, we assume that $r_t^k \leq \hat{x}_t \leq \tilde{x}_t$. We investigate the conditional expectation $\mathbb{E}_k \{ \xi_t^k(\cdot, d_t^k, \dots, d_\tau^k) | d_t^k = \phi_t \}$, where ϕ_t is a known constant, by examining the following three subcases.

Case 1.a. Assume that $\phi_t < \hat{x}_t$. Since we have $\phi_t < \hat{x}_t \leq \tilde{x}_t$, (54) implies that $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h_t + \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h_t + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$. Taking expectations conditional on $d_t^k = \phi_t$ and noting the fact that the demand random variables at different time periods are independent, we obtain $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = h_t + \mathbb{E}_k \{ \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \}$ and $\mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = h_t + \mathbb{E}_k \{ \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \}$. Thus, the induction hypothesis implies that

$$\begin{aligned} & \mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} \\ & \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} + \sum_{s=t+2}^{\tau} \mathbb{E}_k \{ |e_s^k(\tilde{x}_t - \phi_t - \sum_{s'=t+1}^{s-1} d_{s'}^k)| \}. \end{aligned}$$

Case 1.b. Assume that $\hat{x}_t \leq \phi_t < \tilde{x}_t$. We have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = -b_t$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = h_t + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k)$ by (54). Taking expectations conditional on $d_t^k = \phi_t$ and noting that $\tilde{x}_t - \phi_t > 0$, we have $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = -b_t$ and $\mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = h_t + \mathbb{E}_k \{ \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \dots, d_\tau^k) \} = h_t + \dot{v}_{t+1}(\tilde{x}_t - \phi_t) - e_{t+1}^k(\tilde{x}_t - \phi_t)$. Since r_t^* is the minimizer of the convex function $f_t(\cdot)$, (52) implies that we have $\dot{v}_t(\cdot) \geq -c_t$ for all $t = 1, \dots, \tau$. Therefore, we obtain

$$\begin{aligned} & \mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = -b_t \leq h_t - c_{t+1} \\ & \leq h_t + \dot{v}_{t+1}(\tilde{x}_t - \phi_t) \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} + |e_{t+1}^k(\tilde{x}_t - \phi_t)|, \end{aligned}$$

where the first inequality follows from the assumption that $b_t \geq c_{t+1}$.

Case 1.c. Assume that $\phi_t \geq \tilde{x}_t$. In this case, we have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k) = -b_t = \xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \dots, d_\tau^k)$. Taking expectations conditional on $d_t^k = \phi_t$, we obtain $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \} = -b_t = \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k = \phi_t \}$.

The three subcases above show that if $r_t^k \leq \hat{x}_t \leq \tilde{x}_t$, then we have $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k \} \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) | d_t^k \} + |e_{t+1}^k(\tilde{x}_t - d_t^k)| + \sum_{s=t+2}^{\tau} \mathbb{E}_k \{ |e_s^k(\tilde{x}_t - d_t^k - \sum_{s'=t+1}^{s-1} d_{s'}^k)| | d_t^k \}$. Taking expectations, we obtain $\mathbb{E}_k \{ \xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \} \leq \mathbb{E}_k \{ \xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \dots, d_\tau^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(\tilde{x}_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \}$.

It can be shown that the result holds for time period t by considering the other two cases where we have $\hat{x}_t < r_t^k \leq \tilde{x}_t$ or $\hat{x}_t \leq \tilde{x}_t < r_t^k$. This completes the induction argument. \square

The following lemma is analogous to Lemma 4.

Lemma 13 *We have*

$$|e_t^k(x_t)| \leq 2 \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|, \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \}, \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \right\}$$

w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, $k = 1, 2, \dots$

Proof Using (50) and (54), we have

$$\begin{aligned} \mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \} &= \begin{cases} h_t \mathbb{P}_k \{ d_t^k < x_t \} - b_t \mathbb{P}_k \{ d_t^k \geq x_t \} \\ \quad + \mathbb{E}_k \{ \xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t) \} & \text{if } x_t \geq r_t^k \\ h_t \mathbb{P}_k \{ d_t^k < r_t^k \} - b_t \mathbb{P}_k \{ d_t^k \geq r_t^k \} \\ \quad + \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k) \} & \text{if } x_t < r_t^k \end{cases} \\ &= \begin{cases} \dot{f}_t(x_t) - c_t - \mathbb{E}_k \{ \dot{v}_{t+1}(x_t - d_t^k) \mathbf{1}(d_t^k < x_t) \} \\ \quad + \mathbb{E}_k \{ \xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t) \} & \text{if } x_t \geq r_t^k \\ \dot{f}_t(r_t^k) - c_t - \mathbb{E}_k \{ \dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k) \} \\ \quad + \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k) \} & \text{if } x_t < r_t^k. \end{cases} \end{aligned} \quad (56)$$

As before, we consider four cases. It is easy to show that

$$|e_t^k(x_t)| \leq \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|, \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \}, \mathbb{E}_k \{ |e_{t+1}^k(x_t - d_t^k)| \} \right\}$$

for the cases where $x_t \geq r_t^k$ and $x_t \geq r_t^*$, or $x_t < r_t^k$ and $x_t \geq r_t^*$, or $x_t < r_t^k$ and $x_t < r_t^*$ by following the argument in the proof of Lemma 4. This implies that the result holds for these three cases. We only consider the remaining case where $x_t \geq r_t^k$ and $x_t < r_t^*$. If $x_t \leq 0$, then we have $e_t^k(x_t) = 0$ and the result immediately holds. Therefore, we assume that $x_t > 0$ for the rest of the discussion. We have

$$\mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \geq \mathbb{E}_k \{ \xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k) \}$$

by Lemma 12. Using this inequality, (52) and (56), we obtain

$$\begin{aligned} e_t^k(x_t) &= -c_t - \mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \} \\ &\leq -c_t - \mathbb{E}_k \{ \xi_t^k(r_t^k, d_t^k, \dots, d_\tau^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k \{ \dot{v}_{t+1}(r_t^k - d_t^k) \mathbf{1}(d_t^k < r_t^k) \} \\ &\quad - \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < r_t^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \\ &= -\dot{f}_t(r_t^k) + \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \}. \end{aligned}$$

Since $x_t < r_t^*$ and r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $\dot{f}_t(x_t) \leq 0$. Using (56), we also obtain

$$\begin{aligned} e_t^k(x_t) &= -c_t - \mathbb{E}_k \{ \xi_t^k(x_t, d_t^k, \dots, d_\tau^k) \} \\ &= -\dot{f}_t(x_t) + \mathbb{E}_k \{ \dot{v}_{t+1}(x_t - d_t^k) \mathbf{1}(d_t^k < x_t) \} - \mathbb{E}_k \{ \xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \dots, d_\tau^k) \mathbf{1}(d_t^k < x_t) \} \\ &\geq \mathbb{E}_k \{ e_{t+1}^k(x_t - d_t^k) \}. \end{aligned}$$

The last two chains of inequalities imply that

$$\begin{aligned} |e_t^k(x_t)| &\leq \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right| + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \}, \mathbb{E}_k \{ |e_{t+1}^k(x_t - d_t^k)| \} \right\} \\ &\leq 2 \max \left\{ \left| \dot{f}_t(r_t^k) - \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} \right|, \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \right\}. \end{aligned}$$

The result follows by combining the four cases. \square

Finally, we have the following lemma, which is analogous to (A.4) and (A.5) in Proposition 5.

Lemma 14 *There exist constants A_t and B_t such that we have*

$$|e_t^k(x_t)| \leq A_t \sum_{s=t}^{\tau} |\dot{f}_s(r_s^k)| \tag{57}$$

$$|e_t^k(x_t)|^2 \leq B_t \sum_{s=t}^{\tau} \left[|\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k) \mathbb{E}_k \{ e_{s+1}^k(r_s^k - d_s^k) \} \right]^+ \tag{58}$$

w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \dots, \tau$, $k = 1, 2, \dots$

Proof We show the result by induction over the time periods. Since $e_{\tau+1}^k(\cdot) = 0$, Lemma 13 shows that (57) and (58) hold for time period τ with $A_\tau = 2$ and $B_\tau = 4$. Assuming that the result holds for time periods $t+1, \dots, \tau$, Lemma 13 and the induction hypothesis imply that

$$\begin{aligned} |e_t^k(x_t)| &\leq 2 \left\{ |\dot{f}_t(r_t^k)| + \mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \} + \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \right\} \\ &\leq 2 \left\{ |\dot{f}_t(r_t^k)| + A_{t+1} \sum_{s=t+1}^{\tau} |\dot{f}_s(r_s^k)| + \sum_{s=t+1}^{\tau} \left\{ A_s \sum_{s'=s}^{\tau} |\dot{f}_{s'}(r_{s'}^k)| \right\} \right\}. \end{aligned}$$

If we let $A_t = 2(1 + A_{t+1} + \sum_{s=t+1}^{\tau} A_s)$, then (57) holds for time period t . For all $x_t \in \mathbb{R}$, squaring the bound in Lemma 13 also implies that

$$\begin{aligned} |e_t^k(x_t)|^2 &\leq 4 \left\{ [\dot{f}_t(r_t^k)]^2 - 2 \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \} + [\mathbb{E}_k \{ e_{t+1}^k(r_t^k - d_t^k) \}]^2 \right. \\ &\quad \left. + [\mathbb{E}_k \{ |e_{t+1}^k(r_t^k - d_t^k)| \}]^2 + \left\{ \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k(x_t - \sum_{s'=t}^{s-1} d_{s'}^k)| \} \right\}^2 \right\} \end{aligned}$$

$$\begin{aligned}
&\leq 4 \left\{ 2 \left[[\dot{f}_t(r_t^k)]^2 - \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k (r_t^k - d_t^k) \} \right]^+ \right. \\
&\quad \left. + 2 \mathbb{E}_k \{ |e_{t+1}^k (r_t^k - d_t^k)|^2 \} + [\tau - t] \sum_{s=t+1}^{\tau} \mathbb{E}_k \{ |e_s^k (x_t - \sum_{s'=t}^{s-1} d_{s'}^k)|^2 \} \right\} \\
&\leq 4 \left\{ 2 \left[[\dot{f}_t(r_t^k)]^2 - \dot{f}_t(r_t^k) \mathbb{E}_k \{ e_{t+1}^k (r_t^k - d_t^k) \} \right]^+ \right. \\
&\quad \left. + 2 B_{t+1} \sum_{s=t+1}^{\tau} \left[|\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k) \mathbb{E}_k \{ e_{s+1}^k (r_s^k - d_s^k) \} \right]^+ \right. \\
&\quad \left. + [\tau - t] \sum_{s=t+1}^{\tau} \left\{ B_s \sum_{s'=s}^{\tau} \left[|\dot{f}_{s'}(r_{s'}^k)|^2 - \dot{f}_{s'}(r_{s'}^k) \mathbb{E}_k \{ e_{s'+1}^k (r_{s'}^k - d_{s'}^k) \} \right]^+ \right\} \right\},
\end{aligned}$$

where the second inequality uses the fact that $[\sum_{i=1}^n a_i]^2 \leq n \sum_{i=1}^n a_i^2$ and Jensen's inequality, and the third inequality uses the induction hypothesis. If we let $B_t = 4[2 + 2B_{t+1} + (\tau - t) \sum_{s=t+1}^{\tau} B_s]$, then (58) holds for time period t . \square

Once we have these preliminary results, we can follow the same induction argument in Proposition 5, Lemmas 6-8 and Proposition 9 to show that the stochastic approximation method that we propose for the multi-period newsvendor problem with lost sales and nonstationary cost parameters converges to the optimal base-stock levels w.p.1. We also note that the lemma above already shows that results analogous to (A.4) and (A.5) in Proposition 5 are satisfied.

10.3 INVENTORY PURCHASING PROBLEM UNDER PRICE UNCERTAINTY

We consider the setting described in Section 6 and show that the error function $e_t^k(x_t, \hat{p}_t) = \dot{v}_t(x_t, \hat{p}_t) - \mathbb{E}_k \{ \xi_t^k(x_t, \hat{p}_t, p_{t+1}^k, \dots, p_{\tau}^k, d^k) \}$ satisfies a bound similar to the one given in Lemma 4. Once we have this bound, we can follow the same induction argument in Proposition 5, Lemmas 6-8 and Proposition 9 to show that the stochastic approximation method that we propose for the inventory purchasing problem under price uncertainty converges to the optimal base-stock levels w.p.1.

Using (35), we obtain

$$\mathbb{E}_k \{ \xi_t^k(x_t, \hat{p}_t, p_{t+1}^k, \dots, p_{\tau}^k, d^k) \} = \begin{cases} \mathbb{E}_k \{ \xi_{t+1}^k(x_t, p_{t+1}^k, \dots, p_{\tau}^k, d^k) \} & \text{if } x_t \geq r_t^k(\hat{p}_t) \\ \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_{\tau}^k, d^k) \} & \text{if } x_t < r_t^k(\hat{p}_t). \end{cases} \quad (59)$$

We consider four cases. First, we assume that $x_t \geq r_t^k(\hat{p}_t)$ and $x_t \geq r_t^*(\hat{p}_t)$. Using (34) and (59), we have $e_t^k(x_t, \hat{p}_t) = \dot{v}_{t+1}(x_t) - \mathbb{E}_k \{ \xi_{t+1}^k(x_t, p_{t+1}^k, \dots, p_{\tau}^k, d) \} = \mathbb{E}_k \{ e_{t+1}^k(x_t, p_{t+1}^k) \}$. Therefore, we obtain $|e_t^k(x_t, \hat{p}_t)| \leq \mathbb{E}_k \{ |e_{t+1}^k(x_t, p_{t+1}^k)| \}$ by Jensen's inequality.

Second, we assume that $x_t \geq r_t^k(\hat{p}_t)$ and $x_t < r_t^*(\hat{p}_t)$. Following the argument in the proof of Lemma 3, it is easy to show that $\mathbb{E}_k \{ \xi_t^k(\cdot, \hat{p}_t, p_{t+1}^k, \dots, p_{\tau}^k, d^k) \}$ is increasing. In this case, using (33) and (59), we obtain $e_t^k(x_t, \hat{p}_t) = -\hat{p}_t - \mathbb{E}_k \{ \xi_{t+1}^k(x_t, p_{t+1}^k, \dots, p_{\tau}^k, d^k) \} \leq -\hat{p}_t - \mathbb{E}_k \{ \xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_{\tau}^k, d^k) \}$.

Therefore, (30) implies that

$$\begin{aligned} e_t^k(x_t, \hat{p}_t) &\leq -\hat{p}_t - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= -\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \dot{v}_{t+1}(r_t^k(\hat{p}_t)) - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= -\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}. \end{aligned}$$

Since $x_t < r_t^*(\hat{p}_t)$ and $r_t^*(\hat{p}_t)$ is the minimizer of the convex function $f_t(\cdot, \hat{p}_t)$, we have $\dot{f}_t(x_t, \hat{p}_t) \leq 0$. Using (30), we also obtain

$$\begin{aligned} e_t^k(x_t, \hat{p}_t) &= -\hat{p}_t - \mathbb{E}_k\{\xi_{t+1}^k(x_t, p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= -\dot{f}_t(x_t, \hat{p}_t) + \dot{v}_{t+1}(x_t) - \mathbb{E}_k\{\xi_{t+1}^k(x_t, p_{t+1}^k, \dots, p_\tau^k, d^k)\} \geq \mathbb{E}_k\{e_{t+1}^k(x_t, p_{t+1}^k)\}. \end{aligned}$$

The last two chains of inequalities imply that

$$|e_t^k(x_t, \hat{p}_t)| \leq \max\left\{|\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) - \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(x_t, p_{t+1}^k)|\}\right\}.$$

Third, we assume that $x_t < r_t^k(\hat{p}_t)$ and $x_t \geq r_t^*(\hat{p}_t)$. Since $f_t(\cdot, \hat{p}_t)$ is convex, we have $\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) \geq \dot{f}_t(x_t, \hat{p}_t) \geq \dot{f}_t(r_t^*(\hat{p}_t), \hat{p}_t) = 0$. On the other hand, (30) implies that $\dot{v}_{t+1}(x_t) = \dot{f}_t(x_t, \hat{p}_t) - \hat{p}_t = \dot{f}_t(x_t, \hat{p}_t) - \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \dot{v}_{t+1}(r_t^k(\hat{p}_t))$. In this case, using (34) and (59), we obtain

$$\begin{aligned} e_t^k(x_t, \hat{p}_t) &= \dot{f}_t(x_t, \hat{p}_t) - \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \dot{v}_{t+1}(r_t^k(\hat{p}_t)) - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= \dot{f}_t(x_t, \hat{p}_t) - \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}, \end{aligned}$$

which implies that $-\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\} \leq e_t^k(x_t, \hat{p}_t) \leq \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}$. Therefore, we have

$$|e_t^k(x_t, \hat{p}_t)| \leq \max\left\{|\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) - \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)|\}\right\}.$$

Fourth, we assume that $x_t < r_t^k(\hat{p}_t)$ and $x_t < r_t^*(\hat{p}_t)$. In this case, (30), (33) and (59) imply that

$$\begin{aligned} e_t^k(x_t, \hat{p}_t) &= -p_t - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= -\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \dot{v}_{t+1}(r_t^k(\hat{p}_t)) - \mathbb{E}_k\{\xi_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k, \dots, p_\tau^k, d^k)\} \\ &= -\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) + \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}. \end{aligned}$$

Therefore, we have $|e_t^k(x_t, \hat{p}_t)| = |\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) - \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}|$. Combining the four cases, we obtain

$$\begin{aligned} |e_t^k(x_t, \hat{p}_t)| &\leq \max\left\{|\dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) - \mathbb{E}_k\{e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}|, \right. \\ &\quad \left. \mathbb{E}_k\{|e_{t+1}^k(r_t^k(\hat{p}_t), p_{t+1}^k)\}|, \mathbb{E}_k\{|e_{t+1}^k(x_t, p_{t+1}^k)|\}\right\}. \end{aligned}$$