APPROXIMATE DYNAMIC PROGRAMMING AND STOCHASTIC APPROXIMATION METHODS FOR INVENTORY CONTROL AND REVENUE MANAGEMENT

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Sumit Mathew Kunnumkal

August 2007

APPROXIMATE DYNAMIC PROGRAMMING AND STOCHASTIC APPROXIMATION METHODS FOR INVENTORY CONTROL AND REVENUE MANAGEMENT

Sumit Mathew Kunnumkal, Ph.D.

Cornell University 2007

In this thesis, we develop approximate dynamic programming and stochastic approximation methods for problems in inventory control and revenue management. A unifying feature of the methods we develop is that they exploit the underlying problem structure. By doing so, we are able to establish certain theoretical properties of our methods, make them more computationally efficient and obtain a faster rate of convergence.

In the stochastic approximation framework, we develop an algorithm for the monotone estimation problem that uses a projection operator with respect to the max norm onto the order simplex. We show the almost sure convergence of this algorithm and present applications to the Q-learning algorithm and the newsvendor problem with censored demands. Next, we consider a number of inventory control problems for which the so-called base-stock policies are known to be optimal. We propose stochastic approximation methods to compute the optimal base-stock levels. Existing methods in the literature have only local convergence guarantees. In contrast, we show that the iterates of our methods converge to base-stock levels that are globally optimal. Finally, we consider the revenue management problem of optimally allocating seats on a single flight leg to demands from multiple fare

classes that arrive sequentially. We propose a stochastic approximation algorithm to compute the optimal protection levels. The novel aspect of our method is that it works with the nonsmooth version of the problem where capacity can only be allocated in integer quantities. We show that the iterates of our algorithm converge to the globally optimal protection levels.

In the approximate dynamic programming framework, we use a Lagrangian relaxation strategy to make the inventory control decisions in a distribution system consisting of multiple retailers that face random demand and a warehouse that supplies the retailers. Our method is based on relaxing the constraints that ensure the nonnegativity of the shipments to the retailers by associating Lagrange multipliers to them. We show that our method naturally provides a lower bound on the optimal objective value. Furthermore, a good set of Lagrange multipliers can be obtained by solving a convex optimization problem.

BIOGRAPHICAL SKETCH

Sumit received a Bachelors degree in Civil Engineering from IIT Madras in 2000 and a Masters degree in Transportation Engineering from MIT in 2002. He joined the doctoral program in Operations Research at Cornell in 2002.

ACKNOWLEDGEMENTS

I would like to thank my advisor, Huseyin Topaloglu, for his encouragement, support and patience. I am grateful for his contagious enthusiasm and optimism, and his appetite for interesting research problems. I am thankful that he always had a "couple of minutes" for me to discuss my research.

I would like to thank Mike Todd and David Shmoys for agreeing to serve on my committee. I have greatly benefited from my interactions with all the faculty members. In particular, I would like to thank Robin Roundy and Jack Muckstadt for their valuable insights and Uma Prabhu for his interest and kind advice.

I owe many thanks to my fellow students for making the last five years very pleasant. Minbok, for tolerating my eccentric need for extreme silence at home. Serkan, for being a great co-TA and for his help. Peter, for answering my optimization questions. Chandra, for organizing every imaginable sports activity.

I would like to thank Donna Moore, Cindy Jay and Kathy King for their efficient and reliable administrative support. I would also like to thank the School of OR&IE for the financial support.

Last, but most important, I would like to thank my family for their love, support, understanding and prayers. Chachan and Amma for being there during the difficult times. Saj, for his unique sense of humour. Ammachi, Appachan and Achachan for constantly reminding me of how long I had been in the Ph.D. program.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Decision making under uncertainty is a recurring theme in many real world problems. Typically, one has to make a decision at a given point in time, subject to certain constraints, in order to optimize some performance measure, without complete knowledge of the consequences of the action. Such problems can be naturally modeled as dynamic programs. Given the state of the system at time period t, x_t, the optimal action can, at least conceptually, be obtained by solving the Bellman equation

$$v_t(x_t) = \min_{a_t \in \mathbf{A}_t} \mathbb{E}\left\{g_t(x_t, a_t, \omega_t) + v_{t+1}\big(X_{t+1}(x_t, a_t, \omega_t)\big)\right\},$$

where the value function $v_t(x_t)$ gives the optimal cost of operating the system from time period t to the end of the planning horizon when the state at time period t is x_t, $\mathbf{A}_t$ is the set of feasible actions, $g_t(\cdot, \cdot, \cdot)$ is the cost incurred in period t, ω_t represents the random component in the system and $X_{t+1}(\cdot, \cdot, \cdot)$ represents the state in time period $t + 1$. In practice, there are two difficulties with solving the Bellman equation. One difficulty comes from the fact that many real world problems have a multidimensional state vector, the elements of which represent the state of the various components of the system. As a result, the dynamic program has exponentially many states — the so-called "curse of dimensionality" — and solving the Bellman equation becomes computationally intractable. This difficulty is compounded when the state space is continuous, since we have to first discretize the state space and then solve the Bellman equation numerically. A second difficulty comes from the fact that computing the value functions requires computing expectations, which in turn requires access to the distribution functions of the random variables. This information may not be available for various reasons. For example, we may not have sufficient data to fit a parametric distribution. In some cases, we may not even be able to obtain unbiased samples of the random variables. For example, in many inventory control and revenue management settings, the demand is censored by the available inventory. So, we can observe only the amount of sales and not necessarily the demand. As a result, it becomes difficult to obtain unbiased samples of the demand random variables.

This thesis develops methods to address the above mentioned issues. Our target applications are in inventory control and revenue management. We develop stochastic approximation methods that work with only samples of the random variables and therefore do not require access to the distributional information. Moreover, our stochastic approximation methods are applicable when we only observe the amount of sales but not the demand. We apply approximate dynamic programming ideas to solve problems with large state spaces. A common feature of the methods we develop is that they exploit the underlying problem structure. By doing so, we are able to establish certain theoretical properties of our methods, make them more computationally efficient and obtain a faster rate of convergence.

The dynamic programming formulations of many inventory control and revenue management problems have a number of structural properties that can be exploited in different ways. One structural property that comes up often is the nondecreasing nature of the value functions in the state variable. That is, the value functions of such problems lie in the order simplex. If one wants to estimate the value functions through stochastic approximation, then we can exploit this information by projecting the iterates our algorithm onto the order simplex. The objective here is to speed up the rate of convergence of the algorithm. Another structural property which comes up in many inventory control settings is the convexity of the value functions in the state variable. This property leads to the so-called base-stock policies being optimal for such problems. Existing stochastic approximation methods compute the optimal base-stock levels by minimizing the expected total cost expressed as a function of the base-stock levels. However, it turns out that the expected total cost is not necessarily a convex function of the base-stock levels. Hence, the existing methods can only guarantee convergence to a stationary point and not the globally optimal solution. In contrast, by exploiting the convexity of the value functions in the dynamic program, we develop stochastic approximation methods that converge to the globally optimal base-stock levels. Finally, many multi-dimensional dynamic programs can be visualized as consisting of a number of one dimensional dynamic programs linked together by certain constraints. Hawkins (2003) and Adelman and Mersereau (2004) coin the term "weakly coupled dynamic programs" to describe this problem class. On relaxing the linking constraints by associating Lagrange multipliers with them, the resulting value function approximation can be obtained by solving a number of one dimensional dynamic programs. Furthermore, the one dimensional dynamic programs that we solve often have additional structural properties that make the solution procedure computationally efficient.

The organization of this thesis is as follows. In Chapter 2, we propose a stochastic approximation algorithm for the monotone estimation problem. The monotone estimation problem involves estimating the expectation of a random vector that is known to lie in the order simplex. We describe a stochastic approximation algorithm for the monotone estimation problem that exploits this property by projecting its iterates onto the order simplex. We use a projection operator with respect to the max norm. We resolve the difficulties arising from the fact that a projection operator with respect to the max norm is neither uniquely defined nor nonexpansive. We prove the almost sure convergence of the proposed stochastic approximation algorithm. To our knowledge, this is the only convergence result available for a projection operator with respect to the max norm. We present applications to the Q-learning algorithm and the newsvendor problem with censored demands.

In Chapter 3, we consider varaints of the multi-period newsvendor problem, where the so-called base-stock policies are optimal. We propose stochastic approximation methods to compute the globally optimal base-stock levels. The existing stochastic approximation methods in the literature guarantee that their iterates converge, but not necessarily to the globally optimal base-stock levels. In contrast, we prove that the iterates of our methods converge to the globally optimal values. Moreover, our methods 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.

In Chapter 4, we consider the revenue management problem of optimally allocating seats on a single flight leg to the demands from multiple fare classes that arrive sequentially. It is well-known that the optimal policy is characterized by a set of protection levels. We propose a stochastic approximation method to compute the optimal protection levels under the assumption that the demand distributions are not known and we only have access to the samples from the demand distributions. The novel aspect of our method is that it works with the nonsmooth version of the problem where capacity can only be allocated in integer quantities. We show that the iterates of our algorithm converge to the globally optimal protection levels. We also discuss applications to the case where the demand information is censored by the seat availability. While the stochastic approximation methods described in this chapter are similar to those in Chapter 4, the proof techniques to show convergence are considerably different.

In Chapter 5, we consider the inventory replenishment decisions in a distribution system consisting of multiple retailers that face random demand and a warehouse that supplies the retailers. The seminal work by Clark and Scarf (1960) shows that as long as the well-known balance assumption is satisfied, the optimal inventory replenishment policy for the whole distribution system can be found by focusing on one echelon at a time. We propose an approximate dynamic programming method to make the inventory replenishment decisions in a distribution system that also allows finding a "good" policy by focussing on one echelon at a time. Our method is based on formulating the problem as a dynamic program and relaxing the constraints that ensure the nonnegativity of the shipments to the retailers by associating Lagrange multipliers to them. Computational experiments indicate that our method can be effective when the balance assumption does not provide satisfactory results.

Chapter 2

A Stochastic Approximation Algorithm with Max-Norm Projection and its Applications to Monotone Estimation Problems

2.1 Introduction

We consider the problem of estimating the expectation of a random variable η taking values in $\mathbb{R}^n$. We do not have access to the probability distribution of η and we want to estimate the expectation in question only by using samples. Letting $\hat{\eta}$ be the expectation of η and $\hat{\eta}(j)$ be the j-th component of $\hat{\eta}$, we also know that $L \leq \hat{\eta}(1) \leq \ldots \leq \hat{\eta}(n) \leq U$ for finite scalars L and U, and we want to exploit this information when estimating $\hat{\eta}$.

Such problems arise in a variety of settings. For example, $\hat{\eta}$ may be the value function of a dynamic program defined over the state space $\{1, \ldots, n\}$. We may be able to show that the value function is increasing in the state, but we may not have access to the full transition probability matrices and costs to compute the value function. Therefore, we may want to estimate the value function by using sampled state and cost trajectories, and when doing so, we may want to exploit the information that the value function is increasing in the state. Another example arises in the context of the problem $\min_{j \in \{0,\ldots,n\}} \mathbb{E}\big\{F(j,\rho)\big\}$, where $F(j,\rho)$ is a convex function of j for almost all realizations of the random variable ρ. If we do not have access to the probability distribution of ρ and it is relatively easy to obtain samples of $F(j,\rho) - F(j-1,\rho)$, then we may want to estimate $\mathbb{E}\big\{F(j,\rho) - F(j-1,\rho)\big\}$ by using samples of $F(j,\rho) - F(j-1,\rho)$. Letting $\hat{\eta}(j) = \mathbb{E}\big\{F(j,\rho) - F(j-1,\rho)\big\}$ and defining L and U appropriately, since $\mathbb{E}\big\{F(j,\rho)\big\}$ is a convex function of j, we have $L \leq \hat{\eta}(1) \leq \ldots \leq \hat{\eta}(n) \leq U$ and we may want to exploit this information when estimating $\big\{\mathbb{E}\big\{F(j,\rho) - F(j-1,\rho)\big\} : j \in \{1,\ldots,n\}\big\}$.

We propose a stochastic approximation algorithm to address the estimation problem described above. Using $\|\cdot\|$ to denote the max norm, $\mathcal{V}(L,U)$ to denote the set $\big\{v \in \mathbb{R}^n : L \leq v(1) \leq \ldots \leq v(n) \leq U\big\}$ and e^j to denote the j-th unit vector in $\mathbb{R}^n$, this algorithm uses the iteration

$$v_{k+1} \in \operatorname*{argmin}_{v \in \mathcal{V}(L,U)} \|v_k + \alpha_k \left[\eta_k(J_k) - v_k(J_k)\right] e^{J_k} - v\|, \tag{2.1}$$

where α_k is a step size parameter, η_k is a random variable taking values in $\mathbb{R}^n$ and J_k is a random variable taking values in $\{1,\ldots,n\}$. Several comments are in order. The random variable η_k is used to obtain a sample of η. We assume that only one component of η_k is observable and this component is indicated by J_k. We further assume that J_k is independent of η_k. The role of the projection operator is to exploit the information that $\hat{\eta} \in \mathcal{V}(L,U)$. This projection operator is with respect to the max norm and its result is not always uniquely defined. We use a projection with respect to the max norm because this allows us to easily prove an "order preserving" property that roughly states that if we have $L_1 \leq L_2$, $U_1 \leq U_2$ and $v_1 \leq v_2$ with $v_1 \in \mathcal{V}(L_1,U_1)$ and $v_2 \in \mathcal{V}(L_2,U_2)$, then for all $\alpha \in [0,1]$, $\eta \in \mathbb{R}^n$

and $j \in \{1, \ldots, n\}$, there exist $w_1 \in \operatorname{argmin}_{v \in \mathcal{V}(L_1,U_1)} \|v_1 + \alpha\,[\eta(j) - v_1(j)]\, e^j - v\|$ and $w_2 \in \operatorname{argmin}_{v \in \mathcal{V}(L_2,U_2)} \|v_2 + \alpha\,[\eta(j) - v_2(j)]\, e^j - v\|$ such that $w_1 \leq w_2$. This "order preserving" property is important for the application areas that we describe below.

We show the almost sure convergence of the iteration in (2.1) to address the estimation problem described above and provides two application areas of this iteration. The first application area is to the Q-learning algorithm. We consider dynamic programs where the value functions are known to be increasing in the state. An example is a queue admission control problem, where the state of the system is the number of customers waiting to be served. The value function, which denotes the minimum cost of serving the customers, clearly increases with the number of customers. For dynamic programs where the value functions are known to be increasing in the state, we show how to exploit this information by embedding the iteration in (2.1) into the Q-learning algorithm. The second application area is the newsvendor problem with a discrete and bounded demand distribution and censored demands. The newsvendor model applies to single-period, stochastic demand problems where perishable goods are ordered in each period. Demand censorship refers to the situation where we only observe the amount of the product that is sold and not necessarily the demand for the product. We show how to use the iteration in (2.1) to construct a stable algorithm for the newsvendor problem with censored demands. Experiments indicate significant improvements in the empirical convergence behavior for both of these application areas.

Stochastic approximation algorithms date back to Robbins and Monro (1951). Kushner and Clark (1978), Bertsekas and Tsitsiklis (1996) and Kushner and Yin

(1997) provide comprehensive coverage of the topic. If the iterates of a stochastic approximation algorithm are constrained to lie in a closed and convex set, then the standard approach is to project the iterates onto this set. Traditionally, the projection operator is with respect to the Euclidean norm and the convergence of the stochastic approximation algorithm is established through the supermartingale convergence theorem and the nonexpansiveness of the projection operator. Specifically, if we let $\mathcal{W}$ be a closed, bounded and convex set such that $\hat{\eta} \in \mathcal{W}$, an algorithm to estimate $\hat{\eta}$ and that uses the Euclidean norm projection is of the form

$$\tilde{v}_{k+1} = \operatorname*{argmin}_{v \in \mathcal{W}} \|\tilde{v}_k + \alpha_k \left[\eta_k - \tilde{v}_k\right] - v\|_2.$$

The proof of convergence of the above algorithm uses the fact that

$$\begin{aligned} \|\tilde{v}_{k+1} - \hat{\eta}\|_2^2 &\leq \|\tilde{v}_k + \alpha_k \left[\eta_k - \tilde{v}_k\right] - \hat{\eta}\|_2^2 \\ &= \|\tilde{v}_k - \hat{\eta}\|_2^2 + 2\alpha_k \left\langle \tilde{v}_k - \hat{\eta}, \eta_k - \tilde{v}_k \right\rangle + \alpha_k^2 \|\eta_k - \tilde{v}_k\|_2^2 \end{aligned}$$

where $\|\cdot\|_2$ is the Euclidean norm, $\langle \cdot, \cdot \rangle$ denotes the inner product and the first line uses the nonexpansiveness of the Euclidean projection. Under certain assumptions on the samples η_k and step sizes α_k, we can use the supermartingale convergence theorem (see Neveu (1975)) to show that $\lim_{k \to \infty} \tilde{v}_k = \hat{\eta}$ with probability 1 (w.p.1).

We are interested in an algorithm that projects its iterates with respect to the max norm. As the projection with respect to the max norm is not always uniquely defined and nonexpansive, the algorithm may exhibit undesirable behavior if we let $\mathcal{W}$ to be an arbitrary convex set and choose an arbitrary element from the set of points that minimize the max norm distance to our current iterate. To illustrate, we let $\mathcal{W} = \{v \in \mathbb{R}^2 : v(1) = 1,\, 0 \leq v(2) \leq 1\}$ and $\hat{\eta} = (1, 0)$. We consider estimating $\hat{\eta}$ by sampling uniformly from $\{(0,0),(2,0)\}$. Therefore, η_k is a random vector taking values $(0,0)$ or $(2,0)$ with equal probabilities for all $k = 1, 2, \ldots$. If the updates are synchronous, that is, all components of η_k are observable and the updates are of the form

$$v_{k+1} \in \operatorname*{argmin}_{v \in \mathcal{W}} \|v_k + \alpha_k[\eta_k - v_k] - v\|,$$

where η_k takes the value $(0,0)$ and $\alpha_k \in [0,1]$, then, using the fact that $v_k \in \mathcal{W}$, we have $v_k + \alpha_k[\eta_k - v_k] = ((1-\alpha_k),(1-\alpha_k)\, v_k(2))$ and $\min_{v\in\mathcal{W}} \|v_k + \alpha_k[\eta_k - v_k] - v\| = \alpha_k$. Therefore, we have $v_k \in \operatorname{argmin}_{v\in\mathcal{W}} \|v_k + \alpha_k[\eta_k - v_k] - v\|$ and if we are allowed to choose an arbitrary element from the above set, then we can choose $v_{k+1} = v_k$. Similarly, if $\eta_k = (2,0)$, then, following a similar argument, we can also see that we can choose $v_{k+1} = v_k$. Consequently, the algorithm may generate the sequence $\{v_k\}$ such that $v_1 = v_2 = \ldots$ and the point it converges to depends on the initial conditions.

The above example indicates that we cannot hope for a convergent algorithm if $\mathcal{W}$ is an arbitrary set and the max norm projection is chosen in an arbitrary manner. For the case that $\mathcal{W} = \mathcal{V}(L,U)$, we provide a way of choosing a max norm projection that ensures convergence. Moreover, as the max norm projection operator does not necessarily have the nonexpansiveness property (with respect to the Euclidean norm), we need a new proof technique to show the convergence of a stochastic approximation algorithm that uses the iteration in (2.1). Our proof technique considers a fairly standard stochastic approximation algorithm that uses the iteration

$$w_{k+1} = w_k + \alpha_k \left[\eta_k(J_k) - w_k(J_k)\right] e^{J_k}. \tag{2.2}$$

There exist various convergence results for this iteration. We show that the difference between the sequences generated by the iterations in (2.1) and (2.2) becomes arbitrarily small as the iterations progress. This proof technique is quite simple, but we have not seen it being used in other settings and it is effective in the absence of the nonexpansiveness property.

The literature on approximate dynamic programming is also related to our work. There is recent research in this area indicating that the performance of approximate dynamic programming algorithms can be improved by exploiting the structural properties of the problem. Godfrey and Powell (2001), Topaloglu and Powell (2003), Papadaki and Powell (2003), Powell et al. (2004) and Topaloglu (2005) consider dynamic programs where the value functions are known to be convex in the state. They propose approximate dynamic programming algorithms that ensure that the estimates of the value functions obtained during the intermediate iterations are also convex in the state. Embedding the iteration in (2.1) into the Q-learning algorithm is closely related to this line of research.

Ding (2002) provides coverage of the literature on the newsvendor problem with censored demands. Demand censorship occurs when we can observe the amount of the product sold, but not the amount of demand for the product. In this case, if the amount of demand exceeds the product availability, then it is difficult to obtain unbiased samples of the demand random variable. A standard stochastic approximation algorithm naturally addresses the demand censorship because the only information it uses is whether the amount of demand exceeds the product availability. Nevertheless, its empirical performance may not be stable and it may take a large number of iterations to converge.

We make the following research contributions. 1) We show the almost sure convergence of a stochastic approximation algorithm that uses the iteration in (2.1). The novel aspect of this algorithm is that it uses a projection operator with respect to the max norm. We use a novel proof technique that relies on showing that the difference between the sequences generated by the iterations in (2.1) and (2.2) becomes arbitrarily small as the iterations progress. 2) We consider dynamic programs where the value functions are known to be increasing in the state. We show how to exploit this information by embedding the iteration in (2.1) into the Q-learning algorithm. 3) We show how to use the iteration in (2.1) to construct a stable algorithm for the newsvendor problem with censored demands.

The rest of this chapter is organized as follows. In Section 2.2 we provide a convergence result for a stochastic approximation algorithm using a max norm projection that updates its components in a synchronous fashion. That is, in iteration k, we assume that all components of η_k are observable and use this information to update all components of v_k. While this assumption is rarely satisfied in our target applications, it illustrates and provides the intuition behind the proof technique that we use. In Section 2.3, we consider the more realistic case that only one component of η_k is observable. We describe a stochastic approximation algorithm that uses the iteration in (2.1) and show its almost sure convergence. Section 2.4 shows how to embed the iteration in (2.1) into the Q-learning algorithm. Section 2.5 shows how to use the iteration in (2.1) to solve the newsvendor problem with censored demands.

2.2 Convergence with Synchronous Updates

In this section, we show the convergence of a stochastic approximation algorithm that updates all its components in each iteration. Consider the algorithm described in Figure 2.1 Letting $\mathcal{F}_k$ be the σ-subalgebra generated by the random variables $\{\eta_1, \ldots, \eta_{k-1}\}$ in Algorithm 1, we assume that the following conditions hold. (For notational brevity, we assume that the initial iterates are always deterministic and they do not need to be included in σ-subalgebras.)

(A.1) We have $\delta_k(j) = 1$ for all $j \in \{1, \ldots, n\}$, $k = 1, 2, \ldots$.

(A.2) The random variable η_k satisfies $\mathbb{E}\{\eta_k \,|\, \mathcal{F}_k\} = \hat{\eta}$ and $\mathbb{E}\{\langle \eta_k, \eta_k \rangle \,|\, \mathcal{F}_k\} \leq A$ for all $k = 1, 2, \ldots$ with probability 1 for a finite vector $\hat{\eta} \in \mathcal{V}(L, U)$ and a finite scalar A.

(A.3) The step-size parameter α_k is positive and $\mathcal{F}_k$-measurable for all $k = 1, 2, \ldots$.

(A.4) We have $\sum_{k=1}^{\infty} \alpha_k = \infty$ w.p.1 and $\sum_{k=1}^{\infty} \mathbb{E}\big\{\alpha_k^2\big\} < \infty$.

(A.5) We have $\hat{\eta} \in \mathrm{int}(\mathcal{V}(L, U))$.

We note that (A.1) implies that in Step 2 of Algorithm 1, we have $z_k(j) = v_k(j) + \alpha_k[\eta_k(j) - v_k(j)]$ for all $j \in \{1, \ldots, n\}$. One way to satisfy (A.2) is to sample η_k from the probability distribution of η. (A.3) and (A.4) are standard assumptions. While (A.5) is somewhat restrictive, as mentioned earlier, our goal, in this section, is to provide the intuition behind the proof technique that we use. We show that if the sequence $\big\{v_k\big\}$ is generated by Algorithm 1 and (A.1)-(A.5) hold, then we have $\lim_{k \to \infty} v_k = \hat{\eta}$ w.p.1. One difficulty in Algorithm 1, however, is that the

Algorithm 1

(S.1) Choose $v_1 \in \mathcal{V}(L, U)$ and set $k = 1$.

(S.2) Letting $\delta_k(j) \in \{0, 1\}$ for all $j \in \{1, \ldots, n\}$ and η_k be a random variable taking values in $\mathbb{R}^n$ set

$$z_k(j) = v_k(j) + \alpha_k \delta_k(j) \left[\eta_k(j) - v_k(j)\right]$$

for all $j \in \{1, \ldots, n\}$.

(S.3) Set $v_{k+1} \in \operatorname{argmin}_{v \in \mathcal{V}(L,U)} \|z_k - v\|$. Increase k by 1 and go to Step 2.

Figure 2.1: Description of Algorithm 1.

projection operator in Step 3 is not uniquely defined. We resolve this difficulty by defining the set $\mathcal{P}_z^{L,U}$ and the vector $\Pi_z^{L,U}$ as

$$\begin{aligned} \mathcal{P}_z^{L,U} &= \operatorname*{argmin}_{v \in \mathcal{V}(L,U)} \|v - z\| \\ \Pi_z^{L,U} &\in \operatorname*{argmin}_{v \in \mathcal{P}_z^{L,U}} \|v - z\|_2. \end{aligned} \tag{2.3}$$

In Lemma 2 and Proposition 5 below, we show that $\Pi_z^{L,U}$ is uniquely defined and satisfies $\|\Pi_z^{L,U} - v\| \leq \|z - v\|$ for all $z \in \mathbb{R}^n$ and $v \in \mathcal{V}(L, U)$. Using these properties, we have the following convergence result for Algorithm 1. The result follows by comparing the distance between the iterates of Algorithm 1 and the stochastic approximation algorithm described in Figure 2.2. Algorithm 2 is a standard stochastic approximation algorithm and Proposition 4.1 in Bertsekas and Tsitsiklis (1996) shows that if (A.1)-(A.5) hold, then the iterates of Algorithm 2 converge to $\hat{\eta}$ w.p.1. The proof of Proposition 1 relies on showing that the distance between the iterates of Algorithms 1 and 2 gets arbitrarily small as the iterations progress.

Proposition 1 *Let* $\{v_k\}$ *be generated by Algorithm* 1. *Assume that* (A.1)-(A.5) *hold and* v_{k+1} *in Step* 3 *of Algorithm* 1 *is chosen as* $\Pi_{z_k}^{L,U}$. *We have* $\lim_{k\to\infty} v_k = \hat{\eta}$ *w.p.*1.

Proof All statements are in w.p.1 sense. By (A.4), there exists a finite iteration number K' such that $\alpha_k \le 1$ for all $k \ge K'$. We let $\{w_k\}$ be generated by Algorithm 2. By Proposition 4.1 in Bertsekas and Tsitsiklis (1996), we have $\lim_{k\to\infty} w_k = \hat{\eta}$. Since $\hat{\eta} \in \mathrm{int}(\mathcal{V}(L,U))$, there exists a finite iteration number K such that $K \ge K'$ and $w_k \in \mathcal{V}(L,U)$ for all $k \ge K$. We now show by induction that

$$\|v_k - w_k\| \le \prod_{i=K}^{k-1} (1-\alpha_i) \|v_K - w_K\| \tag{2.4}$$

for all $k \ge K$. The result holds for $k = K$. Assuming that the result holds for $k \ge K$, we have

$$\|v_{k+1} - w_{k+1}\| \le \|z_k - w_{k+1}\| = (1-\alpha_k)\|v_k - w_k\| \le \prod_{i=K}^{k} (1-\alpha_i) \|v_K - w_K\|,$$

where the first inequality follows from the fact that $w_{k+1} \in \mathcal{V}(L,U)$, the equality follows from (A.1) and the second inequality follows from the induction assumption. This completes the induction argument. By Lemma 3.3 in Bertsekas and Tsitsiklis (1996) and (A.4), we have $\lim_{k\to\infty} \Pi_{i=K}^{k}(1-\alpha_i) = 0$, and taking the limits in (2.4), we obtain $\lim_{k\to\infty} \|v_k - w_k\| = 0$. The result follows by noting that $\lim_{k\to\infty} w_k = \hat{\eta}$.
□

We now prove the claims made earlier regarding the uniqueness and nonexpansiveness of $\Pi_z^{L,U}$. The following lemma shows that $\mathcal{P}_z^{L,U}$ is closed and convex, and hence, $\Pi_z^{L,U}$ is uniquely defined.

Lemma 2 $\mathcal{P}_z^{L,U}$ *is closed and convex.*

Algorithm 2

(S.1) Choose $w_1 \in \mathcal{V}(L,U)$ and set $k = 1$.

(S.2) For all $j \in \{1, \ldots, n\}$, set

$$w_{k+1}(j) = w_k(j) + \alpha_k\, \delta_k(j)\, [\eta_k(j) - w_k(j)].$$

Increase k by 1 and go to Step 2.

Figure 2.2: Description of Algorithm 2.

Proof We consider a convergent sequence $\{p_k\} \subset \mathcal{P}_z^{L,U}$ with $\lim_{k\to\infty} p_k = \hat{p}$. We have

$$\|\hat{p} - z\| = \|\lim_{k\to\infty} p_k - z\| = \lim_{k\to\infty} \|p_k - z\| = \min_{v\in\mathcal{V}(L,U)} \|v - z\|$$

where the second equality follows from the continuity of $\|\cdot\|$. This establishes that $\hat{p} \in \mathcal{P}_z^{L,U}$ and thus, that $\mathcal{P}_z^{L,U}$ is closed.

To see that $\mathcal{P}_z^{L,U}$ is convex, consider p_1, $p_2 \in \mathcal{P}_z^{L,U}$ and $\lambda \in [0,1]$. We have

$$\begin{aligned}\|\lambda\, p_1 + (1-\lambda)\, p_2 - z\| &\le \lambda\|p_1 - z\| + (1-\lambda)\|p_2 - z\| \\ &= \min_{v\in\mathcal{V}(L,U)} \|v - z\|\end{aligned}$$

where the inequality uses the fact that $\|\cdot\|$ is a convex function and the equality holds since $\|p_1 - z\| = \|p_2 - z\| = \min_{v\in\mathcal{V}(L,U)} \|v - z\|$. On the other hand, since $\mathcal{V}(L,U)$ is a convex set, we have $\lambda\, p_1 + (1-\lambda)\, p_2 \in \mathcal{V}(L,U)$, which implies that $\|\lambda\, p_1 + (1-\lambda)\, p_2 - z\| \ge \min_{v\in\mathcal{V}(L,U)} \|v - z\|$. Therefore, we obtain $\|\lambda\, p_1 + (1 - \lambda)\, p_2 - z\| = \min_{v\in\mathcal{V}(L,U)} \|v - z\|$. Therefore, $\lambda\, p_1 + (1-\lambda)\, p_2 \in \mathcal{P}_z^{L,U}$ and $\mathcal{P}_z^{L,U}$ is convex. □

Proposition 5 below shows that $\|\Pi_z^{L,U} - v\| \le \|z - v\|$ for all $v \in \mathcal{V}(L,U)$. We begin with two preliminary results.

Lemma 3 *Let $z \in \mathbb{R}^n$ and $j^* \in \{1, \ldots, n\}$. We have the following results.*

1) *If $\Pi_z^{L,U}(j^*) < U$, then there exists $j' \in \{j^*, \ldots, n\}$ such that $z(j') \le U$.*

2) *If $\Pi_z^{L,U}(j^*) > L$, then there exists $j' \in \{1, \ldots, j^*\}$ such that $z(j') \ge L$.*

Proof We only show the first part; the second part of the lemma is symmetric and follows from analogous arguments. To get a contradiction, we assume that $\Pi_z^{L,U}(j^*) < U$ and $z(j) > U$ for all $j \in \{j^*, \ldots, n\}$. We let

$$v(j) = \begin{cases} \Pi_z^{L,U}(j) & \text{if } j \in \{1, \ldots, j^* - 1\} \\ U & \text{if } j \in \{j^*, \ldots, n\}. \end{cases}$$

Since $\Pi_z^{L,U} \in \mathcal{V}(L,U)$, we have $v \in \mathcal{V}(L,U)$. Furthermore, we have $|v(j) - z(j)| = |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{1, \ldots, j^* - 1\}$. We also have $|v(j) - z(j)| = z(j) - U \le z(j) - \Pi_z^{L,U}(j) = |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{j^*, \ldots, n\}$, where we use the fact that $\Pi_z^{L,U}(j) \le U < z(j)$ for all $j \in \{j^*, \ldots, n\}$. Therefore, we obtain $|v(j) - z(j)| \le |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{1, \ldots, n\}$, which implies that $\|v - z\| \le \|\Pi_z^{L,U} - z\|$, and hence, $v \in \mathcal{P}_z^{L,U}$. On the other hand, since $\Pi_z^{L,U}(j^*) < U$, we have $|v(j^*) - z(j^*)| = z(j^*) - U < z(j^*) - \Pi_z^{L,U}(j^*) = |\Pi_z^{L,U}(j^*) - z(j^*)|$, which, together with the fact that $|v(j) - z(j)| \le |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{1, \ldots, n\}$, implies that $\|v - z\|_2 < \|\Pi_z^{L,U} - z\|_2$. Since $v \in \mathcal{P}_z^{L,U}$, the last inequality contradicts (2.3). $\square$

Lemma 4 *Let $z \in \mathbb{R}^n$ and $j^* \in \{1, \ldots, n\}$. We have the following results.*

1) *If $\Pi_z^{L,U}(j^*) < z(j^*)$ and there exists $j' \in \{j^*, \ldots, n\}$ such that $z(j') \le U$, then we have $j^* \in \{1, \ldots, n-1\}$ and there exists $j'' \in \{j^* + 1, \ldots, n\}$ such that $\Pi_z^{L,U}(j^*) \ge z(j'')$.*

2) *If* $\Pi_z(j^*) > z(j^*)$ *and there exists* $j' \in \{1,\ldots,j^*\}$ *such that* $z(j') \geq L$, *then we have* $j^* \in \{2,\ldots,n\}$ *and there exists* $j'' \in \{1,\ldots,j^*-1\}$ *such that* $\Pi_z^{L,U}(j^*) \leq z(j'')$.

Proof We only show the first part. To get a contradiction, we assume that $\Pi_z^{L,U}(j^*) < z(j^*)$, $z(j') \leq U$ for some $j' \in \{j^*,\ldots,n\}$ and $j^* = n$. We let

$$v(j) = \begin{cases} \Pi_z^{L,U}(j) & \text{if } j \in \{1,\ldots,n-1\} \\ z(j) & \text{if } j = n. \end{cases}$$

One can easily check that $v \in \mathcal{V}(L,U)$. We have $|v(j) - z(j)| = |\pi_z(j) - z(j)|$ for all $j \in \{1,\ldots,n-1\}$ and $|v(n) - z(n)| = 0 < z(n) - \Pi_z^{L,U}(n) = |\Pi_z^{L,U}(n) - z(n)|$. Therefore, we have $\|v - z\| \leq \|\Pi_z^{L,U} - z\|$ and $\|v - z\|_2 < \|\Pi_z^{L,U} - z\|_2$, which contradicts (2.3). Therefore, we have $j^* \in \{1,\ldots,n-1\}$.

To get another contradiction, we now assume that $\Pi_z^{L,U}(j^*) < z(j^*)$, $z(j') \leq U$ for some $j' \in \{j^*,\ldots,n\}$, $j^* \in \{1,\ldots,n-1\}$ and $\Pi_z^{L,U}(j^*) < z(j)$ for all $j \in \{j^*+1,\ldots,n\}$. We let $\epsilon = \min_{j\in\{j^*,\ldots,n\}}\{z(j) - \Pi_z^{L,U}(j^*)\} > 0$ and

$$v(j) = \begin{cases} \Pi_z^{L,U}(j) & \text{if } j \in \{1,\ldots,j^*-1\} \\ \max\{\Pi_z^{L,U}(j), \Pi_z^{L,U}(j^*) + \epsilon\} & \text{if } j \in \{j^*,\ldots,n\}. \end{cases}$$

One can easily check that $v \in \mathcal{V}(L,U)$. We let $\mathcal{J} = \{j \in \{j^*,\ldots,n\} : \Pi_z^{L,U}(j) \leq \pi_z(j^*) + \epsilon\}$. Using the definition of ϵ, we have $\Pi_z^{L,U}(j) \leq \Pi_z^{L,U}(j^*) + \epsilon \leq z(j)$ for all $j \in \mathcal{J}$, which implies that $|v(j) - z(j)| = z(j) - v(j) \leq z(j) - \Pi_z^{L,U}(j) = |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \mathcal{J}$. Since we have $v(j) = \Pi_z^{L,U}(j)$ for all $j \notin \mathcal{J}$, we have $|v(j) - z(j)| = |\Pi_z^{L,U}(j) - z(j)|$ for all $j \notin \mathcal{J}$. Therefore, we obtain $|v(j) - z(j)| \leq |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{1,\ldots,n\}$, which implies that $\|v - z\| \leq$

$\|\Pi_z^{L,U} - z\|$, and hence, $v \in \mathcal{P}_z^{L,U}$. On the other hand, since $j^* \in \mathcal{J}$, we have $|v(j^*) - z(j^*)| = |\Pi_z^{L,U}(j^*) + \epsilon - z(j^*)| = z(j^*) - \Pi_z^{L,U}(j^*) - \epsilon < z(j^*) - \Pi_z^{L,U}(j^*) = |\Pi_z^{L,U}(j^*) - z(j^*)|$, which, together with the fact that $|v(j) - z(j)| \leq |\Pi_z^{L,U}(j) - z(j)|$ for all $j \in \{1, \ldots, n\}$, implies that $\|v - z\|_2 < \|\Pi_z^{L,U} - z\|_2$. Since $v \in \mathcal{P}_z^{L,U}$, the last inequality contradicts (2.3). □

We are now ready to show that $\Pi_z^{L,U}$ is nonexpansive.

Proposition 5 *For all $z \in \mathbb{R}^n$, we have $\|\Pi_z^{L,U} - v\| \leq \|z - v\|$ for all $v \in \mathcal{V}(L, U)$.*

Proof Letting $j^* \in \{1, \ldots, n\}$ be such that $|\Pi_z^{L,U}(j^*) - v(j^*)| = \|\Pi_z^{L,U} - v\|$, we consider three cases.

Case 1 Assume that $\Pi_z^{L,U}(j^*) < v(j^*)$. We consider two subcases.

Case 1.a Assume that $\Pi_z^{L,U}(j^*) < z(j^*)$. Since $v \in \mathcal{V}(L, U)$, we have $\Pi_z^{L,U}(j^*) < v(j^*) \leq U$ and there exists $j' \in \{j^*, \ldots, n\}$ such that $z(j') \leq U$ by Lemma 3. In this case, Lemma 4 implies that $j^* \in \{1, \ldots, n-1\}$ and there exists $j'' \in \{j^*+1, \ldots, n\}$ such that $\Pi_z^{L,U}(j^*) \geq z(j'')$. Since $j'' \in \{j^* + 1, \ldots, n\}$ and $v \in \mathcal{V}(L, U)$, we have $v(j'') \geq v(j^*) > \Pi_z^{L,U}(j^*) \geq z(j'')$, and hence, $\|z - v\| \geq |z(j'') - v(j'')| = v(j'') - z(j'') \geq v(j^*) - \Pi_z^{L,U}(j^*) = |\Pi_z^{L,U}(j^*) - v(j^*)| = \|\Pi_z^{L,U} - v\|$.

Case 1.b Assume that $\Pi_z^{L,U}(j^*) \geq z(j^*)$. We have $\|z - v\| \geq |z(j^*) - v(j^*)| = v(j^*) - z(j^*) \geq v(j^*) - \Pi_z^{L,U}(j^*) = |\Pi_z^{L,U}(j^*) - v(j^*)| = \|\Pi_z^{L,U} - v\|$.

The cases $\Pi_z^{L,U}(j^*) = v(j^*)$ and $\Pi_z^{L,U}(j^*) > v(j^*)$ can be handled by using similar arguments. □

In closing this section, we make a number of observations. First, Proposition 1 has a simple proof which relies on the nonexpansiveness of the particular max norm projection $\Pi_z^{L,U}$ and the fact that the iterates of Algorithm 2 converge to $\hat{\eta}$. The convergence result, in fact applies to any closed and convex set $\mathcal{W}$, provided there exists a max norm projection $\Pi_z^{\mathcal{W}}$ that is nonexpansive. The following example illustrates that this is not true in general. If we let $\mathcal{W} = \{v \in \mathbb{R}^2 : v(1) \leq 2v(2)\}$, $\tilde{v} = (4, 2)$ and $\tilde{z} = (1, -1)$, then we have $\tilde{v} \in \mathcal{W}$ and $\operatorname{argmin}_{v \in \mathcal{W}} \|v - \tilde{z}\| = \{(0, 0)\}$. Since the set $\operatorname{argmin}_{v \in \mathcal{W}} \|v - \tilde{z}\|$ is a singleton, the max norm projection is unique. Letting $\Pi_{\tilde{z}}^{\mathcal{W}} = (0, 0)$, we have $\|\Pi_{\tilde{z}}^{\mathcal{W}} - \tilde{v}\| = 4 > 3 = \|\tilde{z} - \tilde{v}\|$. Therefore, there does not exist a max norm projection that is nonexpansive for this choice of $\mathcal{W}$ and $\tilde{z}$. Even in the case that $\mathcal{W} = \mathcal{V}(L, U)$ and $\Pi_z^{L,U}$ is nonexpansive, computing $\Pi_z^{L,U}$ in (2.3) requires solving a quadratic program, which may not be computationally attractive. Second, (A.1) requires that every component of v_k is "updated" in Step 2 of Algorithm 1. This becomes problematic when n is large, which is the case when we are trying to approximate value function for a dynamic program with large state space. (A.3) does not allow $\hat{\eta}$ to lie on the boundary of $\mathcal{V}(L, U)$, which is again restrictive in the applications we are interested in. We overcome these limitations in the next section through a somewhat more complicated convergence result

2.3 Convergence with Asynchronous Updates

In this section, we consider the case that exactly one component of η_k is observable in each iteration. Letting J_k be a random variable taking values in the set $\{1, \ldots, n\}$ for all $k = 1, 2, \ldots$ and $\mathcal{F}_k$ be the σ-subalgebra generated by the random variables $\{\eta_1, \ldots, \eta_{k-1}, J_1, \ldots, J_{k-1}\}$ in Algorithm 1, we assume that the following conditions hold. (For notational brevity, we assume that the initial iterates are always deterministic and they do not need to be included in the σ-subalgebras.)

(B.1) We have $\delta_k(j) = \mathbf{1}(j = J_k)$ for all $j \in \{1, \dots, n\}$, $k = 1, 2, \dots$, where $\mathbf{1}(\cdot)$ denotes the indicator function.

(B.2) The random variable η_k satisfies $\mathbb{E}\big\{\eta_k \,|\, \mathcal{F}_k, J_k\big\} = \hat{\eta}$ and $\mathbb{E}\big\{\langle \eta_k, \eta_k \rangle \,|\, \mathcal{F}_k, J_k\big\} \leq A$ for all $k = 1, 2, \dots$ w.p.1 for a vector $\hat{\eta} \in \mathcal{V}(L, U)$ and a finite scalar A.

(B.3) The step size parameter α_k is positive and $\mathcal{F}_k$-measurable for all $k = 1, 2, \dots$.

(B.4) The random variable J_k and the step size parameter α_k satisfy $\sum_{k=1}^{\infty} [\mathbf{1}(j = J_k)\, \alpha_k] = \infty$ w.p.1 and $\sum_{k=1}^{\infty} \mathbb{E}\big\{\mathbf{1}(j = J_k)\, \alpha_k^2\big\} < \infty$ for all $j \in \big\{1, \dots, n\big\}$.

We note that (B.1) implies that we update exactly one component of v_k in iteration k. We also relax the requirement that $\hat{\eta}$ lies in the interior of $\mathcal{V}(L, U)$. (B.3) and (B.4) are standard assumptions. Our goal is to show that if the sequence $\big\{v_k\big\}$ is generated by Algorithm 1 and (B.1)-(B.4) hold, then we have $\lim_{k \to \infty} v_k = \hat{\eta}$ w.p.1. We begin with the next proposition, which shows that $\Pi_{z_k}^{L,U}$ can be computed by mere inspection. In this proposition, we omit the subscripts for the iteration number and write Step 2 of Algorithm 1 as

$$z(j) = \begin{cases} v(j) + \alpha\, [\eta(j) - v(j)] & \text{if } j = J \\ v(j) & \text{otherwise,} \end{cases} \tag{2.5}$$

where $v \in \mathcal{V}(L, U)$. Letting $z(0) = L$ and $z(n+1) = U$ for notational uniformity, since $v \in \mathcal{V}(L, U)$ and z differs from v only in the J-th component, we have either $z \in \mathcal{V}(L, U)$ or $z(J) > z(J+1)$ or $z(J-1) > z(J)$. We are now ready to show the result.

Proposition 6 *Let $v \in \mathcal{V}(L,U)$, z be as in (2.5) and*

$$M = \begin{cases} z(J) & \text{if } z \in \mathcal{V}(L,U) \\ \min\left\{\dfrac{z(J)+z(J+1)}{2}, U\right\} & \text{if } z(J) > z(J+1) \\ \max\left\{\dfrac{z(J-1)+z(J)}{2}, L\right\} & \text{if } z(J-1) > z(J) \end{cases} \tag{2.6}$$

$$p(j) = \begin{cases} \min\left\{z(j), M\right\} & \text{if } j \in \left\{1,\ldots,J-1\right\} \\ M & \text{if } j = J \\ \max\left\{z(j), M\right\} & \text{if } j \in \left\{J+1,\ldots,n\right\} \end{cases} \tag{2.7}$$

for all $j \in \left\{1,\ldots,n\right\}$. We have $\Pi_z^{L,U} = p$.

Proof We consider three cases.

Case 1 Assume that $z(J) > z(J+1)$. First, we show that $p \in \mathcal{V}(L,U)$. Since $v \in \mathcal{V}(L,U)$ and z differs from v only in the J-th component, we have

$$L \leq z(1) \leq z(2) \leq \ldots \leq z(J-1) \leq z(J+1) \leq \ldots \leq z(n) \leq U. \tag{2.8}$$

By (2.8), we have $z(J) > z(J+1) \geq L$, which implies that $[z(J)+z(J+1)]/2 > L$ and we obtain $L \leq M \leq U$ by (2.6). Therefore, (2.7) and (2.8) imply that $L \leq p(j) \leq U$ for all $j \in \left\{1,\ldots,n\right\}$. By (2.7), we have $p(J-1) \leq p(J) \leq p(J+1)$, and by (2.7) and (2.8), we have $p(1) \leq p(2) \leq \ldots \leq p(J-1)$ and $p(J+1) \leq p(J+2) \leq \ldots \leq p(n)$. Therefore, we obtain $p \in \mathcal{V}(L,U)$.

Second, we show that $\|p - z\| = z(J) - M$. We let $\mathcal{J}' = \left\{j \in \left\{J+1,\ldots,n\right\} : z(j) \leq M\right\}$ and $\mathcal{J}'' = \left\{j \in \left\{J+1,\ldots,n\right\} : z(j) > M\right\}$. By (2.7), we have $p(j) = M \geq z(j)$ for all $j \in \mathcal{J}'$, which noting (2.8), implies that $|p(j) - z(j)| = M - z(j) \leq M - z(J+1) \leq [z(J) - z(J+1)]/2$ for all $j \in \mathcal{J}'$, where the second inequality follows from (2.6). We have $|p(j) - z(j)| = 0$ for all $j \in \mathcal{J}''$ by (2.7). We have $p(J) = M \leq [z(J) + z(J+1)]/2 < z(J)$, which implies that $|p(J) - z(J)| = z(J) - M \geq [z(J) - z(J+1)]/2$. Finally, (2.8) implies that $[z(J) + z(J+1)]/2 > z(J+1) \geq z(J-1) \geq z(J-2) \geq \ldots \geq z(1)$. Since we also have $U \geq z(J-1) \geq z(J-2) \geq \ldots \geq z(1)$ by (2.8), we have $M \geq z(j)$ for all $j \in \{1, \ldots, J-1\}$, which, noting (2.7), implies that $|p(j) - z(j)| = 0$ for all $j \in \{1, \ldots, J-1\}$. Therefore, we obtain $\|p - z\| = z(J) - M$.

Third, we show that $\|v - z\| \geq z(J) - M$ for all $v \in \mathcal{V}(L, U)$. We consider two subcases.

Case 1.a Assume that $v(J) \leq M$. Since we have $z(J) > [z(J) + z(J+1)]/2$, (2.6) implies that $M < z(J)$ and we obtain $\|v - z\| \geq |v(J) - z(J)| = z(J) - v(J) \geq z(J) - M$.

Case 1.b Assume that $v(J) > M$. Letting $v(n+1) = U$, since $v \in \mathcal{V}(L, U)$, we have $U \geq v(J+1) \geq v(J) > M = \min\{[z(J) + z(J+1)]/2, U\}$, which implies that $M = [z(J) + z(J+1)]/2 > z(J+1)$ and we obtain $\|v - z\| \geq |v(J+1) - z(J+1)| = v(J+1) - z(J+1) > M - z(J+1) = z(J) - M$.

Therefore, we obtain $\|v - z\| \geq z(J) - M = \|p - z\|$ for all $v \in \mathcal{V}(L, U)$ and $p \in \operatorname{argmin}_{v \in \mathcal{V}(L,U)} \|z - v\|$. Finally, we observe that for any $\hat{p} \in \mathcal{P}_z^{L,U}$, we must have $\hat{p}(J) = M$. Since $\hat{p} \in \mathcal{V}(L, U)$, we also have $\hat{p}(j) \geq M$ for $j \in \mathcal{J}'$. It follows that for $j \in \mathcal{J}'$, we have $|\hat{p}(j) - z(j)| \geq M - z(j) = |p(j) - z(j)|$. It is easy to check that for all $j \notin \mathcal{J}'$ we also have $|\hat{p}(j) - z(j)| \geq |p(j) - z(j)|$, from which we conclude that $\|p - z\|_2 \leq \|\hat{p} - z\|_2$ for all $\hat{p} \in \mathcal{P}_z^{L,U}$. Therefore, we have $p = \Pi_z^{L,U}$. The cases $z \in \mathcal{V}(L, U)$ and $z(J-1) > z(J)$ can be handled by using similar arguments. $\square$

We are now ready to state the main result of this section. Its proof is divided between the next two subsections. The first subsection shows some preliminary results and the second one finishes the proof.

Proposition 7 *Let* $\{v_k\}$ *be generated by Algorithm* 1*. Assume that* (B.1)-(B.4) *hold and* v_{k+1} *in Step* 3 *of Algorithm* 1 *is chosen as* $\Pi^{L,U}_{z_k}$*. We have* $\lim_{k\to\infty} v_k = \hat{\eta}$ *w.p.*1*.*

2.3.1 Preliminary Results

Proposition 4.1 in Bertsekas and Tsitsiklis (1996) shows that if $\{w_k\}$ is generated by Algorithm 2 and (B.1)-(B.4) hold, then we have $\lim_{k\to\infty} w_k = \hat{\eta}$ w.p.1. The proof of Proposition 7 again relies on showing that the difference between the iterates of Algorithms 1 and 2 gets arbitrarily small as the iterations progress. We have the next three preliminary results.

Lemma 8 *Let* $\{w_k\}$ *be generated by Algorithm* 2 *and* $\epsilon > 0$*, and set* $w_k(0) = L$ *and* $w_k(n+1) = U$ *for all* $k = 1, 2, \ldots$ *for notational uniformity. Assume that* (B.1)-(B.4) *hold. There exists a finite iteration number* K *w.p.*1 *such that* $w_k(j) \leq w_k(j') + \epsilon$ *for all* $j \in \{0, \ldots, n\}$*,* $j' \in \{j+1, \ldots, n+1\}$*,* $k \geq K$*.*

Proof All statements are in w.p.1 sense. We let $\hat{\eta}(0) = L$ and $\hat{\eta}(n+1) = U$. Since we have $\lim_{k\to\infty} w_k = \hat{\eta}$, there exists a finite iteration number K such that $\|w_k - \hat{\eta}\| \leq \epsilon/2$ for all $k \geq K$. Noting that $\hat{\eta} \in \mathcal{V}(L, U)$, we obtain $w_k(j) \leq \hat{\eta}(j) + \epsilon/2 \leq \hat{\eta}(j') + \epsilon/2 \leq \omega_k(j') + \epsilon/2 + \epsilon/2$ for all $j \in \{0, \ldots, n+1\}$, $j' \in \{j, \ldots, n+1\}$, $k \geq K$. □

Lemma 9 *Let* $\{v_k\}$ *be generated by Algorithm* 1, $\{w_k\}$ *be generated by Algorithm* 2 *and* $\epsilon > 0$. *Assume that* (B.1)-(B.4) *hold and* v_{k+1} *in Step* 3 *of Algorithm* 1 *is chosen as* $\Pi_{z_k}^{L,U}$. *There exists a finite iteration number* K *w.p.*1 *such that*

$$\begin{aligned}
&\min\Big\{[1-\alpha_k\,\mathbf{1}(j=J_k)]\,[v_k(j)-w_k(j)],\\
&\qquad\min_{j'\in\{j+1,\ldots,n\}}\Big\{[1-\alpha_k\,\mathbf{1}(j'=J_k)]\,[v_k(j')-w_k(j')]\Big\}-\epsilon,-\epsilon\Big\}\\
&\le v_{k+1}(j)-w_{k+1}(j)\\
&\le \max\Big\{[1-\alpha_k\,\mathbf{1}(j=J_k)]\,[v_k(j)-w_k(j)],\\
&\qquad\max_{j'\in\{1,\ldots,j-1\}}\Big\{[1-\alpha_k\,\mathbf{1}(j'=J_k)]\,[v_k(j')-w_k(j')]\Big\}+\epsilon,\epsilon\Big\}
\end{aligned}$$

for all $j \in \{1,\ldots,n\}$, $k \ge K$.

Proof We only show the first inequality. All statements are in w.p.1 sense. We let K be as in Lemma 8, $k \ge K$, $z_k(n+1) = v_k(n+1) = w_k(n+1) = U$ and M be computed as in (2.6) but using (z_k, J_k) instead of (z, J). We consider three cases.

Case 1 Assume that $j \in \{J_k+1,\ldots,n\}$. Since $v_{k+1} = \Pi_{z_k}^{L,U}$, (2.7) implies that $v_{k+1}(j) = \max\{z_k(j), M\} \ge z_k(j)$. We have $v_{k+1}(j)-w_{k+1}(j) \ge z_k(j)-w_{k+1}(j) = [1-\alpha_k\,\mathbf{1}(j=J_k)]\,[v_k(j)-w_k(j)]$.

Case 2 Assume that $j \in \{1,\ldots,J_k-1\}$. We consider two subcases.

Case 2.a Assume that $z_k(j) \le M$. We have $v_{k+1}(j) = \min\{z_k(j), M\} = z_k(j)$ by (2.7), which implies that $v_{k+1}(j) - w_{k+1}(j) = z_k(j) - w_{k+1}(j) = [1-\alpha_k\,\mathbf{1}(j = J_k)]\,[v_k(j)-w_k(j)]$.

Case 2.b Assume that $z_k(j) > M$. We have either $z_k(J_k) > z_k(J_k+1)$ or $z_k(J_k - 1) > z_k(J_k)$ because, otherwise, we have $z_k \in \mathcal{V}(L,U)$ and $z_k(j) > M = z_k(J_k)$ by

(2.6), which contradict the fact that $z_k \in \mathcal{V}(L,U)$ and $j \in \{1,\ldots,J_k-1\}$. However, if $z_k(J_k) > z_k(J_k+1)$, then we have $M = \min\{[z_k(J_k)+z_k(J_k+1)]/2, U\} \geq \min\{z_k(J_k+1), U\}$. Since $v_k \in \mathcal{V}(L,U)$ and z_k differs from v_k only in the J_k-th component, we have $M \geq \min\{z_k(J_k+1),U\} = \min\{v_k(J_k+1),U\} = v_k(J_k+1) \geq v_k(J_k-1) = z_k(J_k-1) \geq v_k(J_k-2) = z_k(J_k-2) \geq \ldots \geq v_k(1) = z_k(1)$, which contradicts the fact that $j \in \{1,\ldots,J_k-1\}$ and $z_k(j) > M$. Therefore, we must have $z_k(J_k-1) > z_k(J_k)$ and $M = \max\{[z_k(J_k-1)+z_k(J_k)]/2, L\}$, which imply that $M = \max\{[z_k(J_k-1)+z_k(J_k)]/2, L\} \geq [z_k(J_k-1)+z_k(J_k)]/2 > z_k(J_k)$. We have $v_{k+1}(j) = \min\{z_k(j), M\} = M > z_k(J_k)$ by (2.7), which implies that

$$\begin{aligned} v_{k+1}(j) - w_{k+1}(j) &> z_k(J_k) - w_{k+1}(j) \geq z_k(J_k) - w_{k+1}(J_k) - \epsilon \\ &= [1-\alpha_k\,\mathbf{1}(J_k = J_k)]\,[v_k(J_k) - w_k(J_k)] - \epsilon \\ &\geq \min_{j' \in \{j+1,\ldots,n\}} \Big\{[1-\alpha_k\,\mathbf{1}(j' = J_k)]\,[v_k(j') - w_k(j')]\Big\} - \epsilon, \end{aligned}$$

where the second inequality follows from Lemma 8 and the third inequality follows from the fact that $J_k \in \{j+1,\ldots,n\}$.

Case 3 Assume that $j = J_k$. We have $v_{k+1}(J_k) = M$ by (2.7). We consider two subcases.

Case 3.a Assume that $z_k(J_k) \leq M$. We obtain $v_{k+1}(J_k) - w_{k+1}(J_k) = M - w_{k+1}(J_k) \geq z_k(J_k) - w_{k+1}(J_k) = [1-\alpha_k\,\mathbf{1}(J_k = J_k)]\,[v_k(J_k) - w_k(J_k)]$.

Case 3.b Assume that $z_k(J_k) > M$. By (2.6), we have either $z_k(J_k) > z_k(J_k+1)$ or $z_k(J_k-1) > z_k(J_k)$. However, if $z_k(J_k-1) > z_k(J_k)$, then we have $M = \max\{[z_k(J_k-1)+z_k(J_k)]/2, L\} \geq \max\{z_k(J_k), L\} \geq z_k(J_k)$. Therefore, we must have $z_k(J_k) > z_k(J_k+1)$ and $M = \min\{[z_k(J_k)+z_k(J_k+1)]/2, U\}$. Since $v_k \in \mathcal{V}(L,U)$ and $z_k(J_k+1) = v_k(J_k+1)$ by the definition of Algorithm 1, we have

$M = \min\{[z_k(J_k) + z_k(J_k+1)]/2, U\} \geq \min\{z_k(J_k+1), U\} = \min\{v_k(J_k+1), U\} = v_k(J_k+1) = z_k(J_k+1)$. Therefore, since $v_{k+1}(J_k) = M$ by (2.7), we obtain

$$\begin{aligned} v_{k+1}(J_k) - w_{k+1}(J_k) &= M - w_{k+1}(J_k) \geq z_k(J_k+1) - w_{k+1}(J_k) \\ &\geq z_k(J_k+1) - w_{k+1}(J_k+1) - \epsilon \\ &= [1 - \alpha_k \mathbf{1}(J_k + 1 = J_k)]\,[v_k(J_k+1) - w_k(J_k+1)] - \epsilon, \end{aligned}$$

where the second inequality follows from Lemma 8. We prefer not replacing $\mathbf{1}(J_k = J_k)$ with 1 or $\mathbf{1}(J_k + 1 = J_k)$ with 0 for notational uniformity. Since $v_k(J_k+1) - w_k(J_k+1) = 0$ when $J_k = n$, the result follows by merging Cases 1, 2 and 3. □

Lemma 10 *Let $\{v_k\}$ be generated by Algorithm 1, $\{w_k\}$ be generated by Algorithm 2, $j \in \{1, \ldots, n\}$ and $\epsilon > 0$. Assume that (B.1)-(B.4) hold and v_{k+1} in Step 3 of Algorithm 1 is chosen as $\Pi^{L,U}_{z_k}$. We have the following results.*

1) *If there exists a finite iteration number K w.p.1 such that $v_{k+1}(j) - w_{k+1}(j) \geq \min\{[1 - \alpha_k \mathbf{1}(j = J_k)]\,[v_k(j) - w_k(j)], -\epsilon\}$ for all $k \geq K$, then there exists a finite iteration number K' w.p.1 such that $v_{k+1}(j) - w_{k+1}(j) \geq -\epsilon$ for all $k \geq K'$.*

2) *If there exists a finite iteration number K w.p.1 such that $v_{k+1}(j) - w_{k+1}(j) \leq \max\{[1 - \alpha_k \mathbf{1}(j = J_k)]\,[v_k(j) - w_k(j)], \epsilon\}$ for all $k \geq K$, then there exists a finite iteration number K' w.p.1 such that $v_{k+1}(j) - w_{k+1}(j) \leq \epsilon$ for all $k \geq K'$.*

Proof We only show the first part. All statements are in w.p.1 sense. By (B.4), there exists a finite iteration number N such that $N \geq K$ and $0 \leq \alpha_k \leq 1$ for all

$k \geq N$. First, we show by induction that

$$v_{k+1}(j) - w_{k+1}(j) \geq \min \left\{ \prod_{i=N}^{k} [1 - \alpha_i \, \mathbf{1}(j = J_i)] \, [v_N(j) - w_N(j)], -\epsilon \right\} \tag{2.9}$$

for all $k \geq N$. By the assumption in the first part, (2.9) holds for $k = N$. Assuming that the result holds for $k \geq N$ and noting the assumption in the first part, we have

$$\begin{aligned} v_{k+2}(j) - w_{k+2}(j) &\geq \min \left\{ [1 - \alpha_{k+1} \, \mathbf{1}(j = J_{k+1})] \, [v_{k+1}(j) - w_{k+1}(j)], -\epsilon \right\} \\ &\geq \min \left\{ [1 - \alpha_{k+1} \, \mathbf{1}(j = J_{k+1})] \min \left\{ \prod_{i=N}^{k} [1 - \alpha_i \, \mathbf{1}(j = J_i)] \right. \right. \\ &\qquad \left. \left. [v_N(j) - w_N(j)], -\epsilon \right\}, -\epsilon \right\} \\ &= \min \left\{ \prod_{i=N}^{k+1} [1 - \alpha_i \, \mathbf{1}(j = J_i)] \, [v_N(j) - w_N(j)], -\epsilon \right\}. \end{aligned}$$

This completes the induction argument. By Lemma 3.3 in Bertsekas and Tsitsiklis (1996) and (B.4), we have $\lim_{k \to \infty} \prod_{i=N}^{k} [1 - \alpha_i \, \mathbf{1}(j = J_i)] = 0$. Therefore, (2.9) implies that there exists an iteration number K' such that $K' \geq N$ and $v_{k+1}(j) - w_{k+1}(j) \geq -\epsilon$ for all $k \geq K'$. □

2.3.2 Proof of Proposition 7

This subsection completes the proof of Proposition 7. We let $\{w_k\}$ be generated by Algorithm 2 and $\epsilon > 0$. By Proposition 4.1 in Bertsekas and Tsitsiklis (1996), we have $\lim_{k \to \infty} w_k = \hat{\eta}$ w.p.1. We now show that there exists a finite iteration number K w.p.1 such that $\|v_k - w_k\| \leq n\epsilon$ for all $k \geq K$. Since $\lim_{k \to \infty} w_k = \hat{\eta}$ w.p.1, we obtain $\lim_{k \to \infty} v_k = \hat{\eta}$ w.p.1. All statements below are in w.p.1 sense.

First, we show by induction that there exists a finite iteration number K' such that $v_k(j) - w_k(j) \geq -n\epsilon$ for all $j \in \{1, \ldots, n\}$, $k \geq K'$. By Lemma 9, there exists a finite iteration number $K(n)$ such that

$$v_{k+1}(n) - w_{k+1}(n) \geq \min\Big\{[1 - \alpha_k\, \mathbf{1}(n = J_k)]\,[v_k(n) - w_k(n)], -\epsilon\Big\}$$

for all $k \geq K(n)$, in which case, Lemma 10 implies that there exists a finite iteration number $K'(n)$ such that $v_{k+1}(n) - w_{k+1}(n) \geq -\epsilon$ for all $k \geq K'(n)$. Assuming that there exists a finite iteration number $K'(j)$ such that $v_{k+1}(j') - w_{k+1}(j') \geq -(n - j + 1)\epsilon$ for all $j' \in \{j, \ldots, n\}$, $k \geq K'(j)$, we now show that there exists a finite iteration number $K'(j-1)$ such that $v_{k+1}(j') - w_{k+1}(j') \geq -(n - j + 2)\epsilon$ for all $j' \in \{j-1, \ldots, n\}$, $k \geq K'(j-1)$. By Lemma 9, there exists a finite iteration number $K(j)$ such that $K(j) \geq K'(j)$ and

$$\begin{aligned}
&v_{k+1}(j-1) - w_{k+1}(j-1)\\
&\geq \min\Big\{[1 - \alpha_k\, \mathbf{1}(j - 1 = J_k)]\,[v_k(j-1) - w_k(j-1)],\\
&\qquad\qquad \min_{j' \in \{j, \ldots, n\}} \Big\{[1 - \alpha_k\, \mathbf{1}(j' = J_k)]\,[v_k(j') - w_k(j')]\Big\} - \epsilon, -\epsilon\Big\}\\
&\geq \min\Big\{[1 - \alpha_k\, \mathbf{1}(j - 1 = J_k)]\,[v_k(j-1) - w_k(j-1)], -(n - j + 1)\epsilon - \epsilon, -\epsilon\Big\}
\end{aligned}$$

for all $k \geq K(j)$, where the second inequality assumes that $K(j)$ is large enough such that $\alpha_k \leq 1$ for all $k \geq K(j)$. In this case, since we have $v_{k+1}(j-1) - w_{k+1}(j-1) \geq \min\{[1 - \alpha_k\, \mathbf{1}(j - 1 = J_k)]\,[v_k(j-1) - w_k(j-1)], -(n - j + 2)\epsilon\}$ for all $k \geq K(j)$, Lemma 10 implies that there exists a finite iteration number $K'(j-1)$ such that $K'(j-1) \geq K(j)$ and $v_{k+1}(j-1) - w_{k+1}(j-1) \geq -(n-j+2)\epsilon$ for all $k \geq K'(j-1)$. Therefore, we have $v_{k+1}(j') - w_{k+1}(j') \geq -(n - j + 2)\epsilon$ for all $j' \in \{j-1, \ldots, n\}$, $k \geq K'(j-1)$. This completes the induction argument, and letting $K' = K'(1)$, we have $v_{k+1}(j) - w_{k+1}(j) \geq -n\epsilon$ for all $j \in \{1, \ldots, n\}$, $k \geq K'$. Using a similar argument, we can also show that there exists a finite iteration number K'' such that $v_{k+1}(j) - w_{k+1}(j) \leq n\epsilon$ for all $j \in \{1, \ldots, n\}$, $k \geq K''$. Letting $K = \max\{K', K''\} + 1$, we have $\|v_k - w_k\| \leq n\epsilon$ for all $k \geq K$ and this establishes Proposition 7.

2.4 Application to the Q-Learning Algorithm

We are interested in an infinite-horizon discounted-cost Markov decision problem with finite sets of states and actions denoted respectively by $\{1, \ldots, n\}$ and $\mathcal{A}$. If the system is in state j and we take action a, then the system moves to state s with probability $p_{js}(a)$ and we incur a finite cost of $g(j, a, s)$. The costs in future time periods are discounted by a factor $\lambda \in [0, 1)$. Letting $Q^a(j)$ be the so-called Q-factor for the state-action pair (j, a), Watkins and Dayan (1992) show that the optimal policy can be found by solving

$$Q^a(j) = \sum_{s=1}^{n} p_{js}(a) \left\{ g(j, a, s) + \lambda \min_{b \in \mathcal{A}} Q^b(s) \right\}, \tag{2.10}$$

in which case it is optimal to take an action in the set $\operatorname{argmin}_{a \in \mathcal{A}} Q^a(j)$ when the system is in state j.

The Q-learning algorithm solves (2.10) through stochastic approximation. Under certain conditions, it can be shown that the iterates of the Q-learning algorithm converge to the optimal Q-factors w.p.1. We refer the reader to Watkins and Dayan (1992), Tsitsiklis (1994) and Bertsekas and Tsitsiklis (1996) for a more detailed discussion. We focus on problems where the Q-factors are known to satisfy $Q^a(j) \leq Q^a(j+1)$ for all $j \in \{1, \ldots, n-1\}$, $a \in \mathcal{A}$ and consider the variant of the Q-learning algorithm described in Figure 2.3. Clearly, the Q-factor approximations $\{Q^a_k : a \in \mathcal{A}\}$ generated by Algorithm 3 satisfy $Q^a_k(j) \leq Q^a_k(j+1)$ for all

Algorithm 3

(S.1) Choose $Q_1^a \in \mathcal{V}(L,U)$ for all $a \in \mathcal{A}$ and set $k = 1$.

(S.2) Sample a state-action pair (J_k, A_k) and a subsequent state S_k.

(S.3) For all $j \in \{1, \dots, n\}$, $a \in \mathcal{A}$, set

$$R_k^a(j) = \begin{cases} Q_k^a(j) + \alpha_k \left[g(j,a,S_k) + \lambda \min_{b \in \mathcal{A}} Q_k^b(S_k) \right. & \\ \quad \left. - Q_k^a(j) \right] & \text{if } j = J_k \text{ and } a = A_k \\ Q_k^a(j) & \text{otherwise.} \end{cases}$$

(S.4) For all $a \in \mathcal{A}$, set $Q_{k+1}^a = \Pi_{R_k^a}^{L,U}$, where $\Pi_{R_k^a}^{L,U}$ is as in (2.7). Increase k by 1 and go to Step 2.

Figure 2.3: Description of Algorithm 3.

$j \in \{1, \dots, n-1\}$, $a \in \mathcal{A}$, $k = 1, 2, \dots$. Our goal is to improve the performance of the Q-learning algorithm by imposing the structural properties of the Q-factors on the Q-factor approximations. Kunnumkal and Topaloglu (2005) give various problems where the Q-factors satisfy $Q^a(j) \le Q^a(j+1)$ for all $j \in \{1, \dots, n-1\}$, $a \in \mathcal{A}$.

Letting $\mathcal{F}_k$ be the σ-subalgebra generated by the random variables $\{J_1, \dots, J_{k-1}, A_1, \dots, A_{k-1}, S_1, \dots, S_{k-1}\}$ in Algorithm 3, we assume that the following conditions hold.

(C.1) We have $\mathbb{P}\{S_k = s \,|\, \mathcal{F}_k, J_k, A_k\} = p_{J_k s}(A_k)$ for all $k = 1, 2, \dots$ w.p.1.

(C.2) We have $Q^a \in \mathcal{V}(L,U)$ for all $a \in \mathcal{A}$, where $\{Q^a : a \in \mathcal{A}\}$ is the solution to (2.10).

(C.3) The step size parameter α_k is positive and $\mathcal{F}_k$-measurable for all $k = 1, 2, \dots$.

(C.4) The random variables J_k and A_k, and the step size parameter α_k satisfy $\sum_{k=1}^{\infty}[\mathbf{1}((j,a) = (J_k, A_k))\,\alpha_k] = \infty$ w.p.1 and $\sum_{k=1}^{\infty}\mathbb{E}\big\{\mathbf{1}((j,a) = (J_k, A_k))\,\alpha_k^2\big\} < \infty$ for all $j \in \big\{1,\ldots,n\big\}$, $a \in \mathcal{A}$.

By (C.1), S_k is sampled according to the transition probabilities. It is easy to see that one can construct simple deterministic sequences $\big\{J_k\big\}$, $\big\{A_k\big\}$ and $\big\{\alpha_k\big\}$ that satisfy (C.4). We have the following convergence result.

Proposition 11 *Let* $\big\{Q_k^a : a \in \mathcal{A}\big\}$ *be generated by Algorithm* 3 *and assume that* (C.1)-(C.4) *hold. We have* $\lim_{k\to\infty} Q_k^a = Q^a$ *w.p.*1 *for all* $a \in \mathcal{A}$*, where* $\big\{Q^a : a \in \mathcal{A}\big\}$ *is the solution to* (2.10).

Proof We sketch the main ideas of the proof here and defer the details to the appendix. For $L_1 \le L_2$, $U_1 \le U_2$, $v_1 \le v_2$ with $v_1 \in \mathcal{V}(L_1, U_1)$ and $v_2 \in \mathcal{V}(L_2, U_2)$, we let z_1 and z_2 be computed as in (2.5) but respectively using v_1 and v_2 instead of v. If we have $\alpha \in [0,1]$, then $z_1 \le z_2$ and one can use Proposition 6 to easily show that $\Pi_{z_1}^{L_1,U_1} \le \Pi_{z_2}^{L_2,U_2}$. This is the "order preserving" property mentioned in the introduction. On the other hand, Step 3 of Algorithm 3 can be written as

$$R_k^{A_k}(J_k) = Q_k^{A_k}(J_k) + \alpha_k \left\{ \omega_k^{A_k}(J_k) + \sum_{s=1}^{n} p_{J_k s}(A_k)\Big\{ g(J_k, A_k, s) + \lambda \min_{b\in\mathcal{A}} Q_k^b(s)\Big\} - Q_k^{A_k}(J_k) \right\},$$

where we let $\omega_k^a(j) = g(j,a,S_k) + \lambda \min_{b\in\mathcal{A}} Q_k^b(S_k) - \sum_{s=1}^{n} p_{js}(a)\big\{g(j,a,s) + \lambda \min_{b\in\mathcal{A}} Q_k^b(s)\big\}$. By (C.1), we have $\mathbb{E}\big\{\omega_k^{A_k}(J_k)\,|\,\mathcal{F}_k, J_k, A_k\big\} = 0$ and $\omega_k^{A_k}(J_k)$ is essentially a noise term. The second term that involves the summation in the curly braces above corresponds to the so-called dynamic programming operator, which is known to be a contraction mapping with respect to the max norm. The rest of the proof uses these facts, the "order preserving" property and Proposition 7, and follows from an argument similar to the one used to show Proposition 3 in Kunnumkal and Topaloglu (2005). □

We illustrate the performance of Algorithm 3 on the following batch service problem. We have a service station with capacity κ to serve the products that arrive randomly over time. We use the number of waiting products as the state of the system. In each time period, we have to decide whether or not to run the service station. If we run the service station, then at most κ products receive service and we incur a cost of ρ. If we do not run the service station or the number of waiting products is greater than κ, then the products that are not served are held until the next time period. We incur a cost of h per waiting product per time period. We assume that $\kappa h/(1-\lambda) > \rho$ so that the cost of holding κ products for an infinite number of time periods is greater than the cost of operating the service station. This ensures that it is optimal to run the service station when there are κ or more products waiting. With the additional assumption that the number of product arrivals in a time period is bounded by κ, we can bound the number of waiting products by $2\kappa - 1$. Therefore, the state space is $\{0, \dots, 2\kappa - 1\}$ and the action space is $\{0, 1\}$, where 0 and 1 respectively correspond to not running and running the service station. It is easy to show that the Q-factors for this problem are increasing in the state.

Letting $\{Q_k^a : a \in \mathcal{A}\}$ and $\{\widetilde{Q}_k^a : a \in \mathcal{A}\}$ respectively be the Q-factor approximations obtained by Algorithm 3 and the standard version of the Q-learning algorithm, Figure 2.4 shows the percent gap between the total expected costs incurred by the optimal policy and the greedy policy characterized by the Q-factor approximations $\{Q_k^a : a \in \mathcal{A}\}$ or $\{\widetilde{Q}_k^a : a \in \mathcal{A}\}$ as a function of the iteration number k. Letting $\mathcal{N}_k(j, a)$ be the number of times the state-action pair (j, a) has been sampled up to iteration k, we set the step size parameter α_k to be $20/(40+\mathcal{N}_k(J_k, A_k))$. Setting the step size parameter in this manner seemed to produce good solutions within a reasonable amount of time. We note that the greedy policies take an action in the sets $\operatorname{argmin}_{a\in\mathcal{A}} Q_k^a(j)$ or $\operatorname{argmin}_{a\in\mathcal{A}} \widetilde{Q}_k^a(j)$ when the system is in state j. If we let $Q_{k+1}^a = R_k^a$ for all $a \in \mathcal{A}$ in Step 4 of Algorithm 3, then we obtain the standard version of the Q-learning algorithm. When generating Figure 2.4, we assume that the number of product arrivals has a truncated geometric distribution with parameter 0.1. The results indicate that as the size of the state space (2κ) or the discount factor (λ) increases, the greedy policies obtained by Algorithm 3 perform noticeably better. For the problem classes we consider, information about $Q^a(j)$ also gives us information about $\{Q^a(j') : j' \in \{1, \ldots, n\} \setminus \{j\}\}$ and the projection operator is instrumental in exploiting all of this information simultaneously.

2.5 Application to the Newsvendor Problem with Censored Demands

The newsvendor problem applies to controlling the inventory of a perishable product at minimum cost subject to stochastic demand. The vendor places a product order at the beginning of each time period, following which the demand for that product is realized. A holding or shortage cost is incurred depending on whether demand is less than or exceeds the order quantity. The newsvendor problem can therefore be written as $\min_{j\in\{0,\ldots,n\}} \mathbb{E}\{F(j, D)\}$, where $F(j, D) =$

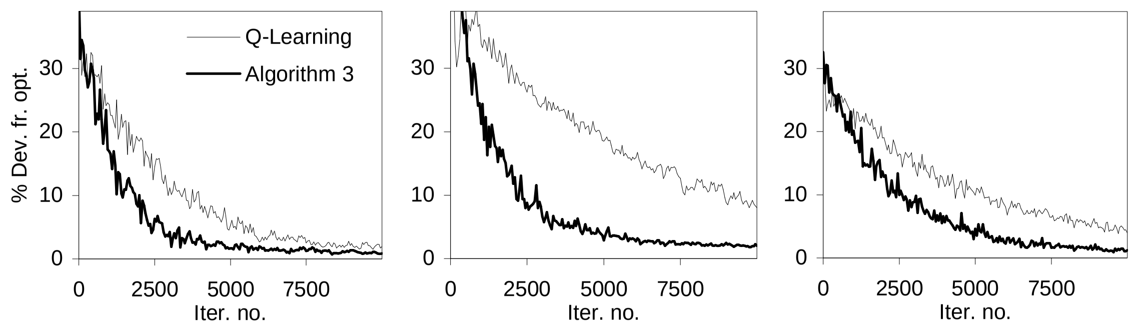

Figure 2.4: From left to right, the charts are for the cases where (κ, λ) is $(25, 0.90)$, $(50, 0.90)$ and $(25, 0.95)$.

$h \max\{j-D,0\} + b \max\{D-j,0\}$ and j is the order quantity, D is the demand random variable with support $\{0,\ldots,n\}$, h is the holding cost and b is the shortage cost. This problem is a special case of the minimization problem described in the introduction and we can use Algorithm 1 to estimate $\hat{\eta}(j) = \mathbb{E}\{F(j,D) - F(j-1,D)\}$ for all $j \in \{1,\ldots,n\}$. In particular, letting J_k be the order quantity and D_k be the demand random variable at iteration k, and assuming that D_k is independent of $\{D_1,\ldots,D_{k-1},J_1,\ldots,J_k\}$ and has the same probability distribution as D, we define

$$\eta_k(J_k) = F(J_k, D_k) - F(J_k - 1, D_k) = \begin{cases} h & \text{if } J_k > D_k \\ -b & \text{if } J_k \leq D_k \end{cases} \tag{2.11}$$

so that $\mathbb{E}\{\eta_k(J_k) \,|\, D_1,\ldots,D_{k-1},J_1,\ldots,J_k\} = \hat{\eta}(J_k)$. We can satisfy the other requirements of (B.1)-(B.4) by simply letting J_k be uniformly distributed over $\{1,\ldots,n\}$, in which case the sequence $\{v_k\}$ generated by Algorithm 1 converges to $\hat{\eta}$ w.p.1. We note that the computation of $\eta_k(J_k)$ in (2.11) only requires knowing whether D_k exceeds J_k and we can use Algorithm 1 when the demand information is censored by the product availability. We let p_k be the minimizer of the piecewise linear convex function characterized by the sequence of left slopes $\{v_k(j) : j \in \{1,\ldots,n\}\}$ at points $\{1,\ldots,n\}$. Therefore, p_k is an estimate of the optimal solution obtained by Algorithm 1 at iteration k. As a benchmark strategy, we use a standard stochastic subgradient algorithm that estimates the optimal solution by using the iteration $r_{k+1} = r_k - \alpha_k\, \eta_k(r_k)$, where $\eta_k(r_k)$ is as in (2.11). We note the slight abuse of notation here and emphasize that $\eta_k(r_k)$ may need to be computed at a noninteger point.

Using $[\cdot]^\phi$ to denote rounding to the nearest integer, Figure 2.5 shows the percent gap between the optimal objective value and the expected costs $\mathbb{E}\{F(p_k, D)\}$ or $\mathbb{E}\{F([r_k]^\phi, D)\}$ as a function of the iteration number k. The step size parameter α_k is set in the same manner as in the Q-learning application. When generating Figure 2.5, we assume that the shortage cost is 10 and the demand has a truncated and discretized normal distribution with mean 30 and variance σ^2. The results indicate that the performance of Algorithm 1 is noticeably better and more stable. However, more research is needed to establish the practical viability of Algorithm 1. For example, it may be wasteful to assume that J_k is uniformly distributed over $\{1, \ldots, n\}$. Empirically, choosing J_k as p_k seems to improve the performance of Algorithm 1 even further, but our convergence result does not hold in this case, since we cannot impose the assumption that $\sum_{k=1}^{\infty} [\mathbf{1}(j = J_k)\, \alpha_k] = \infty$ w.p.1 for all $j \in \{1, \ldots, n\}$.

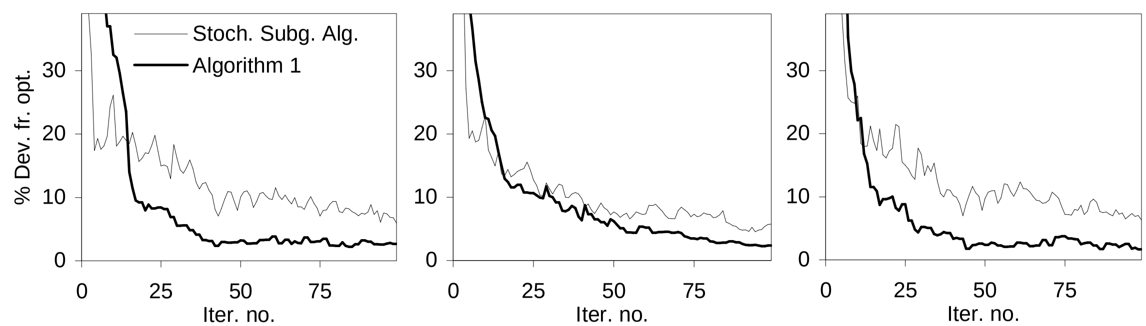

Figure 2.5: From left to right, the charts are for the cases where (h, σ) is $(5, 10)$, $(10, 10)$ and $(5, 5)$.

2.6 Appendix

In this section, we give a complete proof of Proposition 11. We start by showing that the max-norm projection $\Pi_z^{L,U}$ satisfies the following order preserving property.

Lemma 12 *For $L_1 \leq L_2$, $U_1 \leq U_2$, let $v_1 \in \mathcal{V}(L_1, U_1)$ and $v_2 \in \mathcal{V}(L_2, U_2)$ with $v_1 \leq v_2$. Also, let $J_1 = J_2 = J$ and let z_1 and z_2 be computed as in (2.5) but respectively using v_1 and v_2 instead of v. If we have $z_1 \leq z_2$, then we have $\Pi_{z_1}^{L_1,U_1} \leq \Pi_{z_2}^{L_2,U_2}$.*

Proof Let M_1 and M_2 be computed as in (2.6) but respectively using z_1 and z_2 instead of z. We show below that $M_1 \leq M_2$. Using this result and the fact that $z_1 \leq z_2$ in (2.7) completes the proof. We note that the maximum value of M_1 is $\min\{[z_1(J) + z_1(J+1)]/2, U_1\}$, which occurs when $z_1(J) > z_1(J+1)$. On the other hand, the minimum value of M_2 is $\max\{[z_2(J) + z_2(J-1)]/2, L_2\}$, which occurs when $z_2(J) < z_2(J-1)$. We have

$$M_1 \leq \min\left\{\frac{z_1(J) + z_1(J+1)}{2}, U_1\right\} \leq \frac{z_1(J) + z_1(J+1)}{2} \leq \frac{z_2(J) + z_2(J-1)}{2}$$
$$\leq \max\left\{\frac{z_2(J) + z_2(J-1)}{2}, L_2\right\} \leq M_2,$$

where the third inequality follows from the fact that $z_2(J-1) > z_2(J) \geq z_1(J) > z_1(J+1)$. $\square$

We now give a detailed proof for Proposition 11.

Proof of Proposition 11 The proof relies on analyzing Algorithm 4 described in Figure 2.6. Letting $\mathcal{F}_k$ be the σ-subalgebra generated by the random variables

$\{J_1, \ldots, J_{k-1}, A_1, \ldots, A_{k-1}, \omega_1, \ldots, \omega_{k-1}\}$ in Algorithm 4, we assume that the following conditions hold.

(D.1) The random variable ω_k^a satisfies $\mathbb{E}\{\omega_k^a \,|\, \mathcal{F}_k, J_k, A_k\} = 0$ and $\mathbb{E}\{\langle \omega_k^a, \omega_k^a \rangle \,|\, \mathcal{F}_k, J_k, A_k\} \leq A$ for all $k = 1, 2, \ldots$ w.p.1 for a finite scalar A.

(D.2) Using $\|\cdot\|^\phi$ to denote the norm on $\mathbb{R}^n \times \mathbb{R}^{|\mathcal{A}|}$ defined as $\|Q\|^\phi = \max_{a \in \mathcal{A}} \|Q^a\|$, the operator Γ satisfies $\|\Gamma Q_k - \widetilde{Q}\|^\phi \leq \lambda \, \|Q_k - \widetilde{Q}\|^\phi$ for all $k = 1, 2, \ldots$ w.p.1 for some $\lambda \in [0, 1)$ and some $\widetilde{Q} = \{\widetilde{Q}^a : a \in \mathcal{A}\}$ that satisfies $\widetilde{Q}^a \in \mathcal{V}(L, U)$ for all $a \in \mathcal{A}$.

(D.3) The step size parameter α_k is positive and $\mathcal{F}_k$-measurable for all $k = 1, 2, \ldots$.

(D.4) The random variables J_k and A_k, and the step size parameter α_k satisfy $\sum_{k=1}^{\infty} [\mathbf{1}((j, a) = (J_k, A_k)) \, \alpha_k] = \infty$ w.p.1 and $\sum_{k=1}^{\infty} \mathbb{E}\{\mathbf{1}((j, a) = (J_k, A_k)) \, \alpha_k^2\} < \infty$ for all $j \in \{1, \ldots, n\}$, $a \in \mathcal{A}$.

The first step is to show the almost sure convergence of the iterates of Algorithm 4 to $\{\widetilde{Q}^a : a \in \mathcal{A}\}$. All statements are in w.p.1 sense. Letting $C = \max\{|L|, |U|\}$, since $Q_k^a \in \mathcal{V}(L, U)$ and $\widetilde{Q}^a \in \mathcal{V}(L, U)$ for all $a \in \mathcal{A}$, we have $\|Q_k - \widetilde{Q}\|^\phi \leq \|Q_k\|^\phi + \|\widetilde{Q}\|^\phi \leq 2\,C$ for all $k = 1, 2 \ldots$. We choose $\epsilon > 0$ with $\lambda + \epsilon < 1$. Letting $D_1 = 2\,C$, we define the sequence $\{D_t\}$ as $D_{t+1} = (\lambda + \epsilon)\, D_t$. We have $\|Q_k - \widetilde{Q}\|^\phi \leq D_1$ for all $a \in \mathcal{A}$, $k = 1, 2, \ldots$. To show the result by induction, we assume that there exists a finite iteration number k_t such that $\|Q_k - \widetilde{Q}\|^\phi \leq D_t$ for all $k \geq k_t$. We show that this assumption implies that there exists a finite iteration number k_{t+1} such that $\|Q_k - \widetilde{Q}\|^\phi \leq D_{t+1}$ for all $k \geq k_{t+1}$. Since $\lim_{t \to \infty} D_t = 0$, we obtain $\lim_{k \to \infty} Q_k^a = \widetilde{Q}^a$ for all $a \in \mathcal{A}$.

We fix $a \in \mathcal{A}$ and let $e \in \mathbb{R}^n$ be the vector whose components are all ones. For $k \geq k_t$ and starting with $y^a_{k_t} = \widetilde{Q}^a - D_t\, e$ and $w^a_{k_t} = \widetilde{Q}^a + D_t\, e$, we define the sequences $\{y^a_k\}$ and $\{w^a_k\}$ through

$$\hat{y}^a_k(j) = y^a_k(j) + \alpha_k\, \mathbf{1}((j,a) = (J_k, A_k))\, [\omega^a_k(j) + \widetilde{Q}^a(j) - \lambda\, D_t - y^a_k(j)]$$

$$\text{for all } j \in \{1, \ldots, n\}$$

$$\hat{w}^a_k(j) = w^a_k(j) + \alpha_k\, \mathbf{1}((j,a) = (J_k, A_k))\, [\omega^a_k(j) + \widetilde{Q}^a(j) + \lambda\, D_t - w^a_k(j)]$$

$$\text{for all } j \in \{1, \ldots, n\}$$

$y^a_{k+1} = \Pi^{-3C,U}_{\hat{y}^a_k}$ and $w^a_{k+1} = \Pi^{L,3C}_{\hat{w}^a_k}$. Therefore, the sequences $\{y^a_k\}$ and $\{w^a_k\}$ are generated by an algorithm that is equivalent to Algorithm 1. The only difference is that we use projection operators onto $\mathcal{V}(-3C, U)$ and $\mathcal{V}(L, 3C)$ instead of $\mathcal{V}(L, U)$.

Since $\widetilde{Q}^a \in \mathcal{V}(L, U)$ by (D.2) and $D_t \leq D_1 = 2\,C$, we have

$$-3\,C \leq -C - \lambda\, D_t \leq \widetilde{Q}^a(j) - \lambda\, D_t \leq \widetilde{Q}^a(j) \leq U$$

$$L \leq \widetilde{Q}^a(j) \leq \widetilde{Q}^a(j) + \lambda\, D_t \leq C + \lambda\, D_t \leq 3\,C$$

for all $j \in \{1, \ldots, n\}$, which imply that $\widetilde{Q}^a - \lambda\, D_t\, e \in \mathcal{V}(-3C, U)$ and $\widetilde{Q}^a + \lambda\, D_t\, e \in \mathcal{V}(L, 3C)$. By (D.1) and (D.3)-(D.4), the other requirements of (B.1)-(B.4) that are needed for Proposition 7 to hold are satisfied. Therefore, since the sequences $\{y^a_k\}$ and $\{w^a_k\}$ are generated by an algorithm that is equivalent to Algorithm 1 and we have $\mathbb{E}\{\omega^a_k + \widetilde{Q}^a - \lambda\, D_t\, e \,|\, \mathcal{F}_k, J_k, A_k\} = \widetilde{Q}^a - \lambda\, D_t\, e$ and $\mathbb{E}\{\omega^a_k + \widetilde{Q}^a + \lambda\, D_t\, e \,|\, \mathcal{F}_k, J_k, A_k\} = \widetilde{Q}^a + \lambda\, D_t\, e$ by (D.1), Proposition 7 implies that $\lim_{k \to \infty} \|y^a_k - \widetilde{Q}^a + \lambda\, D_t\, e\| = 0$ and $\lim_{k \to \infty} \|w^a_k - \widetilde{Q}^a - \lambda\, D_t\, e\| = 0$.

We now show by induction that $y^a_k \leq Q^a_k \leq w^a_k$ for all $k \geq k_t$. The result holds for $k = k_t$ because we have $y^a_{k_t} = \widetilde{Q}^a - D_t\, e$, $w^a_{k_t} = \widetilde{Q}^a + D_t\, e$ and $\|Q_{k_t} - \widetilde{Q}\|^\phi \leq D_t$.

Assume that $y_k^a \leq Q_k^a \leq w_k^a$ holds for some $k \geq k_t$. We have

$$\begin{aligned}
R_k^a(j) &= [1 - \alpha_k \mathbf{1}((j,a) = (J_k, A_k))]\, Q_k^a(j) + \alpha_k \mathbf{1}((j,a) = (J_k, A_k))\big\{\omega_k^a(j) \\
&\qquad + [\Gamma Q_k]^a(j)\big\} \\
&\leq [1 - \alpha_k \mathbf{1}((j,a) = (J_k, A_k))]\, w_k^a(j) + \alpha_k \mathbf{1}((j,a) = (J_k, A_k))\big\{\omega_k^a(j) + \widetilde{Q}^a(j) \\
&\quad + \lambda \, \|Q_k - \widetilde{Q}\|^\phi\big\} \\
&\leq [1 - \alpha_k \mathbf{1}((j,a) = (J_k, A_k))]\, w_k^a(j) + \alpha_k \mathbf{1}((j,a) = (J_k, A_k))\big\{\omega_k^a(j) + \widetilde{Q}^a(j) \\
&\quad + \lambda \, D_t\big\} = \hat{w}_k^a(j)
\end{aligned}$$

for all $j \in \{1, \ldots, n\}$, where the first inequality uses the induction hypothesis and (D.2), and the second inequality uses the assumption that $\|Q_k - \widetilde{Q}\|^\phi \leq D_t$ for all $k \geq k_t$. Using a similar argument, we can also show that $\hat{y}_k^a(j) \leq R_k^a(j)$ for all $j \in \{1, \ldots, n\}$. Therefore, we have $\hat{y}_k^a \leq R_k^a \leq \hat{w}_k^a$. Since y_k^a and w_k^a are respectively the projections of $\hat{y}_{k-1}^a$ and $\hat{w}_{k-1}^a$ onto $\mathcal{V}(-3C, U)$ and $\mathcal{V}(L, 3C)$, we have $y_k^a \in \mathcal{V}(-3C, U)$ and $w_k^a \in \mathcal{V}(L, 3C)$. In this case, the "order preserving" property implies that $y_{k+1}^a = \Pi_{\hat{y}_k^a}^{-3C,U} \leq \Pi_{R_k^a}^{L,U} = Q_{k+1}^a \leq \Pi_{\hat{w}_k^a}^{L,3C} = w_{k+1}^a$. Therefore, we have $y_k^a \leq Q_k^a \leq w_k^a$ for all $k \geq k_t$.

Since we have $\lim_{k \to \infty} \|y_k^a - \widetilde{Q}^a + \lambda \, D_t \, e\| = 0$ and $\lim_{k \to \infty} \|w_k^a - \widetilde{Q}^a - \lambda \, D_t \, e\| = 0$, there exists a finite iteration number $\bar{k}_{t+1}^a$ such that $\bar{k}_{t+1}^a \geq k_t$, and $y_k^a - \widetilde{Q}^a + \lambda \, D_t \, e \geq -\epsilon \, D_t \, e$ and $w_k^a - \widetilde{Q}^a - \lambda \, D_t \, e \leq \epsilon \, D_t \, e$ for all $k \geq \bar{k}_{t+1}^a$. In this case, using the fact that $y_k^a \leq Q_k^a \leq w_k^a$ for all $k \geq k_t$, we have $-D_{t+1} \, e = -(\lambda + \epsilon) \, D_t \, e \leq y_k^a - \widetilde{Q}^a \leq Q_k^a - \widetilde{Q}^a \leq w_k^a - \widetilde{Q}^a \leq (\lambda + \epsilon) \, D_t \, e \leq D_{t+1} \, e$ for all $k \geq \bar{k}_{t+1}^a$. Letting $k_{t+1} = \max_{a \in \mathcal{A}} \bar{k}_{t+1}^a$, we have $\|Q_k - \widetilde{Q}\|^\phi \leq D_{t+1}$ for all $k \geq k_{t+1}$. This establishes that the iterates of Algorithm 4 converge to $\{\widetilde{Q}^a : a \in \mathcal{A}\}$, which, noting (D.2), is the fixed point of the operator Γ.

Algorithm 4

(S.1) Choose $Q_1^a \in \mathcal{V}(L,U)$ for all $a \in \mathcal{A}$ and set $k = 1$.

(S.2) Sample a state-action pair (J_k, A_k).

(S.3) For all $j \in \{1, \dots, n\}$, $a \in \mathcal{A}$, set

$$R_k^a(j) = Q_k^a(j) + \alpha_k \mathbf{1}((j,a) = (J_k, A_k)) \, [\omega_k^a(j) + (\Gamma Q_k)^a(j) - Q_k^a(j)],$$

where ω_k^a is a random variable taking values in $\mathbb{R}^n$ for all $a \in \mathcal{A}$, Γ is an operator on $\mathbb{R}^n \times \mathbb{R}^{|\mathcal{A}|}$ and $(\Gamma Q_k)^a(j)$ denotes the (j,a)-th component of ΓQ_k.

(S.4) For all $a \in \mathcal{A}$, set $Q_{k+1}^a = \Pi_{R_k^a}^{L,U}$, where $\Pi_{R_k^a}^{L,U}$ is as in (2.7). Increase k by 1 and go to Step 2.

Figure 2.6: Description of Algorithm 4.

Letting

$$\omega_k^a(j) = g(j,a,S_k) + \lambda \min_{b \in \mathcal{A}} Q_k^b(S_k) - \sum_{s=1}^{n} p_{js}(a) \big\{ g(j,a,s) + \lambda \min_{b \in \mathcal{A}} Q_k^b(s) \big\} \tag{2.12}$$

$$(\Gamma Q_k)^a(j) = \sum_{s=1}^{n} p_{js}(a) \Big\{ g(j,a,s) + \lambda \min_{b \in \mathcal{A}} Q_k^b(s) \Big\}, \tag{2.13}$$

we can identify Algorithm 3 with Algorithm 4. Furthermore, (C.1)-(C.4) imply that the requirements of (D.1)-(D.4) are satisfied. For example, if $\omega_k^a(j)$ is as in (2.12), then (C.1) implies that we have $\mathbb{E}\big\{\omega_k^{A_k}(J_k) \,|\, J_1, \dots, J_k, A_1, \dots, A_k, S_1, \dots, S_{k-1}\big\} = 0$. Since $Q_k^a \in \mathcal{V}(L,U)$ for all $a \in \mathcal{A}$, $\omega_k^{A_k}(J_k)$ is uniformly bounded for all $k = 1, 2, \dots$. Therefore, (D.1) holds for $\omega_k^{A_k}(J_k)$. We do not need to worry about the other random variables $\big\{\omega_k^a(j) : j \in \{1, \dots, n\} \setminus \{J_k\}, a \in \mathcal{A} \setminus \{A_k\}\big\}$ because they are not used in Algorithm 4. The operator Γ in (2.13) corresponds to the so-called dynamic programming operator, which is known to be a contraction mapping with respect to the max norm. Therefore, if we let $\left\{\widetilde{Q}^a : a \in \mathcal{A}\right\}$ be the solution to (2.10), then (D.2) immediately holds. (C.3)-(C.4) are identical to (D.3)-(D.4). Consequently, (C.1)-(C.4) imply that the requirements of (D.1)-(D.4) are satisfied and the iterates of Algorithm 3 converge to the fixed point of the dynamic programming operator. $\square$

Chapter 3

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

3.1 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 have widely been used in practice.

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 $\{r_1^*, \ldots, r_\tau^*\}$ such that it is optimal to keep the inventory position at time period t as close as possible to r_t^*. That is, 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. If we let the purchasing cost be zero and denote the initial inventory position as 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

$$\begin{aligned} g(x_1, r_1, r_2) = h\,\mathbb{E}\Big\{ &[(x_1 \vee r_1) - d_1]^+ + [\max\{(x_1 \vee r_1) - d_1, r_2\} - d_2]^+ \Big\} \\ &+ b\,\mathbb{E}\Big\{[d_1 - (x_1 \vee r_1)]^+ + [d_2 - \max\{(x_1 \vee r_1) - d_1, r_2\}]^+ \Big\}, \end{aligned}$$

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

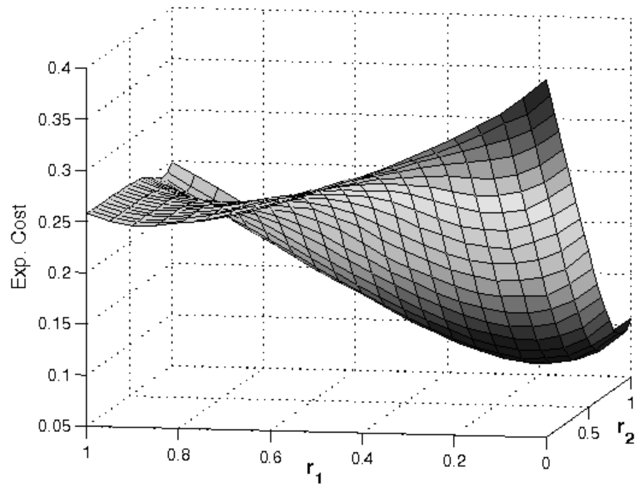

Figure 3.1: Total expected cost as a function of the base-stock levels for a two-period newsvendor problem. The problem parameters are $x_1 = 0$, $h = 0.25$, $b = 0.4$, $d_1 \sim \text{beta}(1, 5)$, $d_2 \sim \text{beta}(5, 1)$.

case, the optimal base-stock levels can be found by solving the problem

$$(r_1^*, r_2^*) = \underset{(r_1, r_2)}{\text{argmin}}\; g(x_1, r_1, r_2). \tag{3.1}$$

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 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 (3.1) and the solution obtained by a stochastic approximation method may not be very good. Figure 3.1 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-behaved" 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 (3.1), then we can only obtain a stationary point of $g(x_1, \cdot, \cdot)$.

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 somewhat nonstandard inventory purchasing problem where the price of the product changes randomly over time and we have to purchase a certain amount of the product to satisfy the random demand that occurs at the end of the planning horizon. Although the problems that we work on are well-studied, we make several substantial contributions. First, we offer a remedy for the aforementioned surprising observation by showing that it is 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. For example, we only require the ability to obtain samples of the demand random variables and we do not explicitly compute expectations. Second, it is difficult to solve the Bellman equations numerically when the demand distributions are continuous and our stochastic approximation methods provide alternatives for computing the optimal base-stock levels. Third, our stochastic approximation methods are applicable when we only observe the amount of inventory sold, but not the amount of demand. Therefore, we can still compute the optimal base-stock levels when the demand information is censored by the amount of available inventory. Finally, our computational experience indicates that our stochastic approximation methods can provide significantly better solutions than the existing stochastic approximation methods.

The remainder of this chapter is organized as follows. Section 3.2 briefly reviews the related literature. Sections 3.3 and 3.4 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 3.5 shows that the proposed stochastic approximation methods are applicable when the demand information is censored. Section 3.6 develops a stochastic approximation method for an inventory purchasing 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 3.7 presents numerical experiments.

3.2 Relevant Literature

We mainly consider the multi-period newsvendor problem with backlogged demands or lost sales. 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 et al. (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 a Bayesian framework to estimate the demand parameters and to adaptively update the replenishment quantities as the demand information becomes available. Levi et al. (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. Gallego and Moon (1993) address the uncertainty in the distribution of the demand by using the robust optimization framework and search for policies with the best worst-case performance.

Stochastic approximation methods can deal with the uncertainty in the distribution of the demand and the censored demand information, at least to a certain extent. They only require the ability to obtain samples from the demand distributions. Furthermore, they usually do not require to have 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 stochastic approximation methods that we propose also possess these features.

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 broad coverage of the theory 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 of the objective functions, 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 methods in the reinforcement learning literature that try to avoid this shortcoming by explicitly approximating the value functions in the dynamic programming formulation of the problem. Q-learning algorithm 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 et al. (2004) propose sampling-based methods to approximate piecewise-linear convex value functions and these methods are known to be convergent for certain stationary problems.

The stochastic approximation methods that we propose 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.

3.3 Multi-Period Newsvendor Problem with Backlogged Demands

We want to control the inventory of a product over the time periods $\{1, \ldots, \tau\}$. 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 $\{d_t : t = 1, \ldots, \tau\}$

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 $\{t, \ldots, \tau\}$ when the inventory position at time period t is x_t and the optimal policy is followed over the time periods $\{t, \ldots, \tau\}$. The functions $\{v_t(\cdot) : t = 1, \ldots, \tau\}$ satisfy the Bellman equations

$$v_t(x_t) = \min_{y_t \geq x_t} \; c\,[y_t - x_t] + \mathbb{E}\left\{h\,[y_t - d_t]^+ + b\,[d_t - y_t]^+ + v_{t+1}(y_t - d_t)\right\}, \tag{3.2}$$

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

$$f_t(r_t) = c\,r_t + \mathbb{E}\left\{h\,[r_t - d_t]^+ + b\,[d_t - r_t]^+ + v_{t+1}(r_t - d_t)\right\}, \tag{3.3}$$

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 $\{r_t^* : t = 1, \ldots, \tau\}$. 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 (3.2) as

$$\begin{aligned} v_t(x_t) &= \begin{cases} \mathbb{E}\left\{h\,[x_t - d_t]^+ + b\,[d_t - x_t]^+ + v_{t+1}(x_t - d_t)\right\} & \text{if } x_t \geq r_t^* \\ c\,[r_t^* - x_t] + \mathbb{E}\left\{h\,[r_t^* - d_t]^+ + b\,[d_t - r_t^*]^+ + v_{t+1}(r_t^* - d_t)\right\} & \text{if } x_t < r_t^* \end{cases} \\ &= \begin{cases} f_t(x_t) - c\,x_t & \text{if } x_t \geq r_t^* \\ f_t(r_t^*) - c\,x_t & \text{if } x_t < r_t^*. \end{cases} \end{aligned} \tag{3.4}$$

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 13 *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 $\hat{x}_t,\ \tilde{x}_t \in \mathbb{R}$, $t = 1, \ldots, \tau$.*

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

$$\dot{f}_t(x_t) = c + h\,\mathbb{P}\big\{d_t < x_t\big\} - b\,\mathbb{P}\big\{d_t \geq x_t\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(x_t - d_t)\big\}, \tag{3.5}$$

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 (3.4) to obtain

$$\dot{v}_t(x_t) = \begin{cases} \dot{f}_t(x_t) - c & \text{if } x_t \geq r_t^* \\ -c & \text{if } x_t < r_t^*. \end{cases} \tag{3.6}$$

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

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

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. □

We now consider computing the optimal base-stock levels $\{r_t^* : t = 1, \ldots, \tau\}$ through a stochastic approximation method. Noting (3.5) and using $\mathbf{1}(\cdot)$ to denote the indicator function, we can compute a stochastic gradient of $f_t(\cdot)$ at x_t through

$$\Delta_t(x_t, d_t) = c + h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t). \tag{3.7}$$

In this case, letting $\{r_t^k : t = 1, \ldots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k, $\{d_t^k : t = 1, \ldots, \tau\}$ 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

$$r_t^{k+1} = r_t^k - \alpha^k\, \Delta_t(r_t^k, d_t^k). \tag{3.8}$$

However, this approach is clearly not realistic because the computation in (3.7) requires the knowledge of $\{\dot{v}_t(\cdot) : t = 1, \ldots, \tau\}$. The stochastic approximation method we propose is based on constructing tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \ldots, \tau\}$.

Since r_t^* is the minimizer of $f_t(\cdot)$, (3.5) implies that $-c = \dot{f}_t(r_t^*) - c = h\,\mathbb{P}\big\{d_t < r_t^*\big\} - b\,\mathbb{P}\big\{d_t \geq r_t^*\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t^* - d_t)\big\}$. Therefore, using (3.5) and (3.6), we obtain

$$\dot{v}_t(x_t) = \begin{cases} h\,\mathbb{P}\big\{d_t < x_t\big\} - b\,\mathbb{P}\big\{d_t \geq x_t\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(x_t - d_t)\big\} & \text{if } x_t \geq r_t^* \\ h\,\mathbb{P}\big\{d_t < r_t^*\big\} - b\,\mathbb{P}\big\{d_t \geq r_t^*\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t^* - d_t)\big\} & \text{if } x_t < r_t^*. \end{cases} \tag{3.9}$$

From this expression, it is clear that

$$\dot{v}_t(x_t, d_t) = \begin{cases} h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t) & \text{if } x_t \geq r_t^* \\ h\,\mathbf{1}(d_t < r_t^*) - b\,\mathbf{1}(d_t \geq r_t^*) + \dot{v}_{t+1}(r_t^* - d_t) & \text{if } x_t < r_t^* \end{cases} \tag{3.10}$$

gives a stochastic gradient of $v_t(\cdot)$ at x_t, satisfying $\dot{v}_t(x_t) = \mathbb{E}\big\{\dot{v}_t(x_t, d_t)\big\}$. To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \ldots, \tau\}$, we "mimic" the computation in (3.10) by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k : t = 1, \ldots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k, we recursively define

$$\xi_t^k(x_t, d_t, \ldots, d_\tau) = \begin{cases} h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \geq x_t) \\ \qquad\qquad + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau) & \text{if } x_t \geq r_t^k \\ h\,\mathbf{1}(d_t < r_t^k) - b\,\mathbf{1}(d_t \geq r_t^k) \\ \qquad\qquad + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \ldots, d_\tau) & \text{if } x_t < r_t^k, \end{cases} \tag{3.11}$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \ldots, \cdot) = 0$. At iteration k, replacing $\dot{v}_{t+1}(x_t - d_t)$ in (3.7) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)$, we use

$$s_t^k(x_t, d_t, \ldots, d_\tau) = c + h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \geq x_t) + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau) \tag{3.12}$$

to approximate the stochastic gradient of $f_t(\cdot)$ at x_t. 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 $\{r_t^1 : t = 1, \ldots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

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

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

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 (3.7) and (3.12) 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 $\lim_{k\to\infty} \dot{f}_t(r_t^k) = 0$ 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 $k \to \infty$. We begin with several preliminary lemmas.

3.3.1 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 14 *There exists a constant M such that $|\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)| \leq M$ and $|s_t^k(x_t, d_t^k, \ldots, d_\tau^k)| \leq M$ w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \ldots, \tau$, $k = 1, 2, \ldots$.*

Proof We let $N = \max\{c, h, b\}$. We show by induction over the time periods that $|\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)| \leq 2\,[\tau - t + 1]\,N$ w.p.1 for all $x_t \in \mathbb{R}$, $t = 1, \ldots, \tau$, $k = 1, 2, \ldots$. The result holds for time period τ by (3.11). Assuming that the result holds for time period $t+1$, we have $|\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)| \leq h + b + 2\,[\tau - t]\,N \leq 2\,[\tau - t + 1]\,N$ w.p.1 and this establishes the result. Therefore, we have $|s_t^k(x_t, d_t^k, \ldots, d_\tau^k)| \leq c + h + b + 2\,[\tau - t]\,N \leq 2\,[\tau - t + 2]\,N$ w.p.1 by (3.12). 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 $\mathbb{E}_k\big\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)\big\}$ also satisfies this property.

Lemma 15 *If $\hat{x}_t$, $\tilde{x}_t$ satisfy $\hat{x}_t \leq \tilde{x}_t$, then we have $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, \ldots, d_\tau^k)\big\} \leq \mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, \ldots, d_\tau^k)\big\}$ w.p.1 for all $t = 1, \ldots, \tau$, $k = 1, 2, \ldots$.*

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 (3.11), we have $\mathbb{E}_k\big\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\big\} = h\,\mathbb{P}_k\big\{d_\tau^k < \hat{x}_\tau\big\} - b\,\mathbb{P}_k\big\{d_\tau^k \geq \hat{x}_\tau\big\} = [h + b]\,\mathbb{P}_k\big\{d_\tau^k < \hat{x}_\tau\big\} - b \leq [h + b]\,\mathbb{P}_k\big\{d_\tau^k < \tilde{x}_\tau\big\} - b = \mathbb{E}_k\big\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\big\}$. Second, we assume that $\hat{x}_\tau < r_\tau^k \leq \tilde{x}_\tau$. In this case, (3.11) and the argument in the previous sentence imply that $\mathbb{E}_k\big\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\big\} = \mathbb{E}_k\big\{\xi_\tau^k(r_\tau^k, d_\tau^k)\big\} \leq \mathbb{E}_k\big\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\big\}$. Third, we assume that $\hat{x}_\tau \leq \tilde{x}_\tau < r_\tau^k$. We have $\mathbb{E}_k\big\{\xi_\tau^k(\hat{x}_\tau, d_\tau^k)\big\} = \mathbb{E}_k\big\{\xi_\tau^k(r_\tau^k, d_\tau^k)\big\} = \mathbb{E}_k\big\{\xi_\tau^k(\tilde{x}_\tau, d_\tau^k)\big\}$. 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 $\hat{x}_t \leq \tilde{x}_t < r_t^k$. □

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

$$e_t^k(x_t) = \dot{v}_t(x_t) - \mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\}, \tag{3.13}$$

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 $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)\}$ is increasing.

Lemma 16 *We have*

$$|e_t^k(x_t)| \leq \max\Big\{\big|\dot{f}_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}\big|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}, \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\}\Big\} \tag{3.14}$$

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

Proof Using (3.5) and (3.11), we obtain

$$\mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\} = \begin{cases} h\,\mathbb{P}_k\{d_t^k < x_t\} - b\,\mathbb{P}_k\{d_t^k \geq x_t\} \\ \qquad + \mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\} & \text{if } x_t \geq r_t^k \\ h\,\mathbb{P}_k\{d_t^k < r_t^k\} - b\,\mathbb{P}_k\{d_t^k \geq r_t^k\} \\ \qquad + \mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\} & \text{if } x_t < r_t^k \end{cases}$$

$$= \begin{cases} \dot{f}_t(x_t) - c - \mathbb{E}_k\big\{\dot{v}_{t+1}(x_t - d_t^k)\big\} \\ \qquad\qquad + \mathbb{E}_k\big\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\big\} & \text{if } x_t \geq r_t^k \\ \dot{f}_t(r_t^k) - c - \mathbb{E}_k\big\{\dot{v}_{t+1}(r_t^k - d_t^k)\big\} \\ \qquad\qquad + \mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\big\} & \text{if } x_t < r_t^k. \end{cases} \tag{3.15}$$

We consider four cases. First, we assume that $x_t \geq r_t^k$ and $x_t \geq r_t^*$. Using (3.6) and (3.15), we have $e_t^k(x_t) = \mathbb{E}_k\big\{\dot{v}_{t+1}(x_t - d_t^k)\big\} - \mathbb{E}_k\big\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\big\} = \mathbb{E}_k\big\{e_{t+1}^k(x_t - d_t^k)\big\}$. Therefore, we obtain $|e_t^k(x_t)| \leq \mathbb{E}_k\big\{|e_{t+1}^k(x_t - d_t^k)|\big\}$ by Jensen's inequality.

Second, we assume that $x_t \geq r_t^k$ and $x_t < r_t^*$. We have $\mathbb{E}_k\big\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\big\} \geq \mathbb{E}_k\big\{\xi_t^k(r_t^k, d_t^k, \ldots, d_\tau^k)\big\}$ by Lemma 15. Using this inequality, (3.6) and (3.15), we obtain

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

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 (3.15), we also obtain

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

The last two chains of inequalities imply that

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

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 (3.6) and (3.15), we obtain $e_t^k(x_t) = \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k\big\{\dot{v}_{t+1}(r_t^k - d_t^k)\big\} - \mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\big\} = \dot{f}_t(x_t) - \dot{f}_t(r_t^k) + \mathbb{E}_k\big\{e_{t+1}^k(r_t^k - d_t^k)\big\}$, which implies that

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

Therefore, we obtain

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

Fourth, we assume that $x_t < r_t^k$ and $x_t < r_t^*$. In this case, (3.6) and (3.15) imply that $e_t^k(x_t) = -\dot{f}_t(r_t^k) + \mathbb{E}_k\big\{\dot{v}_{t+1}(r_t^k - d_t^k)\big\} - \mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\big\} = -\dot{f}_t(r_t^k) + \mathbb{E}_k\big\{e_{t+1}^k(r_t^k - d_t^k)\big\}$. Therefore, we obtain $|e_t^k(x_t)| = \big|\dot{f}_t(r_t^k) - \mathbb{E}_k\big\{e_{t+1}^k(r_t^k - d_t^k)\big\}\big|$. The result follows by combining the four cases. □

3.3.2 Convergence Proof

We have the following convergence result for Algorithm 1.

Proposition 17 *Assume that the sequence of step size parameters* $\{\alpha^k\}_k$ *satisfy* $\alpha_k \geq 0$ *for all* $k = 1, 2, \ldots,$ $\sum_{k=1}^\infty \alpha^k = \infty$ *and* $\sum_{k=1}^\infty [\alpha^k]^2 < \infty$ *w.p.1. If the sequence of base-stock levels* $\{r_t^k : t = 1, \ldots, \tau\}_k$ *are generated by Algorithm 1, then the sequence* $\{f_t(r_t^k)\}_k$ *converges w.p.1 for all* $t = 1, \ldots, \tau$ *and we have* $\lim_{k\to\infty} \dot{f}_t(r_t^k) = 0$ *w.p.1 for all* $t = 1, \ldots, \tau$*.*

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 $t = 1, \ldots, \tau$.

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

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

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

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

(E.5) There exists a constant $\mathcal{A}_t$ such that we have

$$|e_t^k(x_t)|^2 \leq \mathcal{A}_t \sum_{s=t}^{\tau} \big[\, |\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k)\ \mathbb{E}_k\big\{e_{s+1}^k(r_s^k - d_s^k)\big\}\big]^+ \tag{3.16}$$

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

We begin by showing that (E.1)-(E.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 13 and the Taylor series expansion of $f_\tau(\cdot)$ at r_τ^{k+1}, a standard argument yields

$$f_\tau(r_\tau^{k+1}) \leq f_\tau(r_\tau^k) - \alpha^k\, \dot{f}_\tau(r_\tau^k)\ s_\tau^k(r_\tau^k, d_\tau^k) + \frac{1}{2}\,[\alpha^k]^2 L\, |s_\tau^k(r_\tau^k, d_\tau^k)|^2; \tag{3.17}$$

see (3.39) in Bertsekas and Tsitsiklis (1996). Since we have $\mathbb{E}_k\big\{s_\tau^k(r_\tau^k, d_\tau^k)\big\} = c + h\,\mathbb{P}_k\big\{d_\tau^k < r_\tau^k\big\} - b\,\mathbb{P}_k\big\{d_\tau^k \geq r_\tau^k\big\} = \dot{f}_\tau(r_\tau^k)$, taking expectations in (3.17) and using Lemma 14 yield

$$\mathbb{E}_k\big\{f_\tau(r_\tau^{k+1})\big\} \leq f_\tau(r_\tau^k) - \alpha^k\, [\dot{f}_\tau(r_\tau^k)]^2 + \frac{1}{2}\,[\alpha^k]^2 L\, M^2. \tag{3.18}$$

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, (E.1) and (E.2) hold for time period τ. Since we have $\mathbb{E}_k\big\{s_\tau^k(r_\tau^k, d_\tau^k)\big\} = \dot{f}_\tau(r_\tau^k)$, 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 $\lim_{k\to\infty} \dot{f}_\tau(r_\tau^k) = 0$; see Proposition 4.1 in Bertsekas and Tsitsiklis (1996). Therefore, (E.3) holds for time period τ. Since

$e^k_{\tau+1}(\cdot) = 0$, Lemma 16 shows that (E.4) and (E.5) hold for time period τ. Therefore, (E.1)-(E.5) hold for time period τ.

Assuming that (E.1)-(E.5) hold for time periods $t+1, \dots, \tau$, Lemmas 18-20 below show that (E.1)-(E.5) also hold for time period t. This completes the induction argument, and the result follows by (E.1) and (E.3). $\square$

Lemmas 18-20 complete the induction argument given in the proof of Proposition 17. All statements in their proofs should be understood in w.p.1 sense.

Lemma 18 *If (E.1)-(E.5) hold w.p.1 for time periods $t+1, \dots, \tau$, then (E.1) and (E.2) hold w.p.1 for time period t.*

Proof Using (3.5) and (3.12), we have

$$\begin{aligned}\mathbb{E}_k\big\{s^k_t(r^k_t, d^k_t, \dots, d^k_\tau)\big\} &= c + h\,\mathbb{P}_k\big\{d^k_t < r^k_t\big\} - b\,\mathbb{P}_k\big\{d^k_t \geq r^k_t\big\} \\ &\qquad + \mathbb{E}_k\big\{\xi^k_{t+1}(r^k_t - d^k_t, d^k_{t+1}, \dots, d^k_\tau)\big\} \\ &= \dot{f}_t(r^k_t) - \mathbb{E}_k\{\dot{v}_{t+1}(r^k_t - d^k_t)\} \\ &\qquad + \mathbb{E}_k\big\{\xi^k_{t+1}(r^k_t - d^k_t, d^k_{t+1}, \dots, d^k_\tau)\big\}.\end{aligned}$$

Similar to (3.17) and (3.18), using the equality above, Lemma 13 and the Taylor series expansion of $f_t(\cdot)$ at r^{k+1}_t, we have

$$\begin{aligned}\mathbb{E}_k\big\{f_t(r^{k+1}_t)\big\} &\leq f_t(r^k_t) - \alpha^k\,\dot{f}_t(r^k_t)\;\mathbb{E}_k\big\{s^k_t(r^k_t, d^k_t, \dots, d^k_\tau)\big\} + \frac{1}{2}\,[\alpha^k]^2 L\,M^2 \\ &= f_t(r^k_t) - \alpha^k\,\dot{f}_t(r^k_t)\,\big[\dot{f}_t(r^k_t) - \mathbb{E}_k\big\{e^k_{t+1}(r^k_t - d^k_t)\big\}\big] + \frac{1}{2}\,[\alpha^k]^2 L\,M^2.\end{aligned}$$

Letting $X^k = \alpha^k\,\dot{f}_t(r^k_t)\,\big[\dot{f}_t(r^k_t) - \mathbb{E}_k\big\{e^k_{t+1}(r^k_t - d^k_t)\big\}\big]$, the expression above is of the form $\mathbb{E}_k\big\{f_t(r^{k+1}_t)\big\} \leq f_t(r^k_t) - [X^k]^+ + [-X^k]^+ + [\alpha^k]^2 L\,M^2/2$. 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

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

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

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

We note that the expression on the right side of (3.16) 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 (E.5), we obtain

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

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

Lemma 19 *If* $(E.1)$-$(E.5)$ *hold w.p.*1 *for time periods* $t+1,\ldots,\tau$*, then* $(E.3)$ *holds w.p.*1 *for time period* t*.*

Proof We first show that $\liminf_{k\to\infty}|\dot{f}_t(r_t^k)| = 0$. By the induction hypothesis (E.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 (E.3), we obtain $\lim_{k\to\infty} \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\} = 0$. 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 $k = K, K+1, \ldots$.

By Lemma 18, (E.2) holds for time period t. Since we have $\sum_{k=1}^{\infty} \alpha^k = \infty$, (E.2) implies that $\liminf_{k\to\infty}\big[\,|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k)\ \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}\big]^+ = 0$. In particular, we must have $\big[\,|\dot{f}_t(r_t^k)|^2 - \dot{f}_t(r_t^k)\ \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}\big]^+ \leq 3\,\epsilon^2$ 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 $\liminf_{k\to\infty}|\dot{f}_t(r_t^k)| = 0$.

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 $\limsup_{k\to\infty}|\dot{f}_t(r_t^k)| = 0$ and this establishes the result. We defer the proof of this part to the appendix. □

Lemma 20 *If* $(E.1)$-$(E.5)$ *hold w.p.*1 *for time periods* $t+1,\ldots,\tau$*, then* $(E.4)$ *and* $(E.5)$ *hold w.p.*1 *for time period* t*.*

Proof The induction hypothesis (E.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 (3.14), it is easy to see that (E.4) holds for time period t. For all $x_t \in \mathbb{R}$, squaring (3.14) also implies that

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

where we use the induction hypothesis (E.5) in the last inequality. Letting $\mathcal{A}_t = \max\{2, 3\,\mathcal{A}_{t+1}\}$, (E.5) holds for time period t. □

We close this section by investigating the performances of the policies characterized by the base-stock levels $\{r_t^k : t = 1, \ldots, \tau\}$. The policy characterized by the base-stock levels $\{r_t^k : t = 1, \ldots, \tau\}$ 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 $\{t, \ldots, \tau\}$ when the inventory position at time period t is x_t. The functions $\{V_t^k(\cdot) : t = 1, \ldots, \tau\}$ satisfy

$$V_t^k(x_t) = \begin{cases} \mathbb{E}_k\big\{h\,[x_t - d_t^k]^+ + b\,[d_t^k - x_t]^+ + V_{t+1}^k(x_t - d_t^k)\big\} & \text{if } x_t \ge r_t^k \\ c\,[r_t^k - x_t] + \mathbb{E}_k\big\{h\,[r_t^k - d_t^k]^+ + b\,[d_t^k - r_t^k]^+ & \\ \qquad\qquad + V_{t+1}^k(r_t^k - d_t^k)\big\} & \text{if } x_t < r_t^k. \end{cases} \tag{3.19}$$

In contrast, the function $v_1(\cdot)$ gives the minimum total expected cost incurred over the time periods $\{1, \ldots, \tau\}$. Proposition 21 shows that $\lim_{k\to\infty} |V_1^k(x_1) - v_1(x_1)| = 0$ w.p.1 for all $x_1 \in \mathbb{R}$ and establishes that the policies characterized by the base-stock levels $\{r_t^k : t = 1, \ldots, \tau\}$ are asymptotically optimal.

Proposition 21 *Assume that the sequence of step size parameters* $\{\alpha^k\}_k$ *satisfy* $\alpha_k \geq 0$ *for all* $k = 1, 2, \ldots,$ $\sum_{k=1}^{\infty} \alpha^k = \infty$ *and* $\sum_{k=1}^{\infty} [\alpha^k]^2 < \infty$ *w.p.1. If the sequence of base-stock levels* $\{r_t^k : t = 1, \ldots, \tau\}_k$ *are generated by Algorithm* 1*, then we have* $\lim_{k\to\infty} |V_1^k(x_1) - v_1(x_1)| = 0$ *w.p.1 for all* $x_1 \in \mathbb{R}$*.*

Proof All statements are in w.p.1 sense. We first show that $\lim_{k\to\infty} f_t(r_t^k) = f_t(r_t^*)$ for all $t = 1, \ldots, \tau$. By Proposition 17, 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 17, 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 (3.3), we can write (3.19) as

$$V_t^k(x_t) = \begin{cases} f_t(x_t) - c\,x_t + \mathbb{E}_k\big\{V_{t+1}^k(x_t - d_t^k) - v_{t+1}(x_t - d_t^k)\big\} & \text{if } x_t \geq r_t^k \\ f_t(r_t^k) - c\,x_t + \mathbb{E}_k\big\{V_{t+1}^k(r_t^k - d_t^k) - v_{t+1}(r_t^k - d_t^k)\big\} & \text{if } x_t < r_t^k. \end{cases} \tag{3.20}$$

We use induction over the time periods to show that $0 \leq V_t^k(x_t) - v_t(x_t) \leq \sum_{s=t}^{\tau} \big[f_s(r_s^k) - f_s(r_s^*)\big]$ for all $x_t \in \mathbb{R}$, $t = 1, \ldots, \tau$. 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 (3.4), (3.20) and the induction hypothesis, we obtain

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

The other three cases where we have $r_t^* \le x_t < r_t^k$, or $r_t^* \le x_t$ and $r_t^k \le x_t$, or $r_t^* \ge x_t$ and $r_t^k \ge x_t$ can be shown similarly. □

3.4 Multi-Period Newsvendor Problem with Lost Sales

This section shows how to extend the ideas in Section 3.3 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 $b > c \ge 0$, $h \ge 0$ 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 3.3, the functions $\{v_t(\cdot) : t = 1, \ldots, \tau\}$ satisfy the Bellman equations

$$v_t(x_t) = \min_{y_t \ge x_t} \; c\,[y_t - x_t] + \mathbb{E}\big\{h\,[y_t - d_t]^+ + b\,[d_t - y_t]^+ + v_{t+1}([y_t - d_t]^+)\big\}, \quad (3.21)$$

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

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

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 $\{r_t^* : t = 1, \ldots, \tau\}$ are the minimizers of the functions $\{f_t(\cdot) : t = 1, \ldots, \tau\}$.

Since we have

$$\dot{f}_t(r_t) = c + h\,\mathbb{P}\big\{d_t < r_t\big\} - b\,\mathbb{P}\big\{d_t \geq r_t\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t - d_t)\,\mathbf{1}(d_t < r_t)\big\}, \quad (3.22)$$

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

$$\Delta_t(x_t, d_t) = c + h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t)\,\mathbf{1}(d_t < x_t) \quad (3.23)$$

and iteratively search for the optimal base-stock levels through (3.8). However, this approach requires the knowledge of $\{v_t(\cdot) : t = 1, \ldots, \tau\}$. We now use ideas similar to those in Section 3.3 to approximate the stochastic gradients of $\{f_t(\cdot) : t = 1, \ldots, \tau\}$ in a tractable manner.

Using the optimal base-stock level r_t^*, we write (3.21) as

$$v_t(x_t) = \begin{cases} \mathbb{E}\big\{h\,[x_t - d_t]^+ + b\,[d_t - x_t]^+ + v_{t+1}([x_t - d_t]^+)\big\} & \text{if } x_t \geq r_t^* \\ c\,[r_t^* - x_t] + \mathbb{E}\big\{h\,[r_t^* - d_t]^+ + b\,[d_t - r_t^*]^+ + v_{t+1}([r_t^* - d_t]^+)\big\} & \text{if } x_t < r_t^*. \end{cases} \quad (3.24)$$

Since r_t^* is the minimizer of $f_t(\cdot)$, (3.22) implies that $-c = \dot{f}_t(r_t^*) - c = h\,\mathbb{P}\big\{d_t < r_t^*\big\} - b\,\mathbb{P}\big\{d_t \geq r_t^*\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t^* - d_t)\,\mathbf{1}(d_t < r_t^*)\big\}$. Therefore, using this expression in (3.24), we obtain

$$\dot{v}_t(x_t) = \begin{cases} h\,\mathbb{P}\big\{d_t < x_t\big\} - b\,\mathbb{P}\big\{d_t \geq x_t\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(x_t - d_t)\,\mathbf{1}(d_t < x_t)\big\} & \text{if } x_t \geq r_t^* \\ h\,\mathbb{P}\big\{d_t < r_t^*\big\} - b\,\mathbb{P}\big\{d_t \geq r_t^*\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t^* - d_t)\,\mathbf{1}(d_t < r_t^*)\big\} & \text{if } x_t < r_t^*. \end{cases} \quad (3.25)$$

In this case,

$$\dot{v}_t(x_t, d_t) = \begin{cases} h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \ge x_t) + \dot{v}_{t+1}(x_t - d_t)\,\mathbf{1}(d_t < x_t) & \text{if } x_t \ge r_t^* \\ h\,\mathbf{1}(d_t < r_t^*) - b\,\mathbf{1}(d_t \ge r_t^*) + \dot{v}_{t+1}(r_t^* - d_t)\,\mathbf{1}(d_t < r_t^*) & \text{if } x_t < r_t^* \end{cases} \tag{3.26}$$

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

To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \ldots, \tau\}$, we "mimic" the computation in (3.26) by using the estimates of the optimal base-stock levels. In particular, letting $\{r_t^k : t = 1, \ldots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k, we recursively define

$$\xi_t^k(x_t, d_t, \ldots, d_\tau) = \begin{cases} h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \ge x_t) \\ \qquad + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)\,\mathbf{1}(d_t < x_t) & \text{if } x_t \ge r_t^k \\ h\,\mathbf{1}(d_t < r_t^k) - b\,\mathbf{1}(d_t \ge r_t^k) \\ \qquad + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \ldots, d_\tau)\,\mathbf{1}(d_t < r_t^k) & \text{if } x_t < r_t^k, \end{cases} \tag{3.27}$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \ldots, \cdot) = 0$. At iteration k, replacing $\dot{v}_{t+1}(x_t - d_t)$ in (3.23) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)$, we use

$$\begin{aligned} s_t^k(x_t, d_t, \ldots, d_\tau) = c + h\,\mathbf{1}(d_t < x_t) - b\,\mathbf{1}(d_t \ge x_t) \\ + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)\,\mathbf{1}(d_t < x_t) \end{aligned} \tag{3.28}$$

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 3.3.1 and 3.3.2. In particular, we can follow the proof of Lemma 14 to derive bounds on $\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)$ and $s_t^k(\cdot, d_t^k, \ldots, d_\tau^k)$. Lemma 22 below shows that $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)\}$ is increasing and it is the analogue of Lemma 15. The proof of this lemma strictly requires the assumption that the cost parameters are stationary. We define the error function as

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

with $e_{\tau+1}^k(\cdot) = 0$. In this case, we can show that the same bound on the error function given in Lemma 16 holds. Once we have this bound on the error function, we can follow the same induction argument in Proposition 17, Lemmas 18-20 and Proposition 21 to show the final result. The only difference from the argument in Sections 3.3.1 and 3.3.2 occurs when showing that $\mathbb{E}_k\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)\}$ is increasing and the following lemma establishes this result.

Lemma 22 *If* $\hat{x}_t$, $\tilde{x}_t$ *satisfy* $\hat{x}_t \leq \tilde{x}_t$, *then we have* $\mathbb{E}_k\{\xi_t^k(\hat{x}_t, d_t^k, \ldots, d_\tau^k)\} \leq \mathbb{E}_k\{\xi_t^k(\tilde{x}_t, d_t^k, \ldots, d_\tau^k)\}$ *w.p.1 for all* $t = 1, \ldots, \tau,\ k = 1, 2, \ldots$.

Proof We show the result by induction over the time periods. Since (3.11) and (3.27) 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 15. Furthermore, we have $\mathbb{E}_k\{\xi_\tau^k(x_\tau, d_\tau^k)\} \geq -b$ for all $x_\tau \in \mathbb{R}$ by (3.27). Assuming that the result holds for time period $t+1$ and we have $\mathbb{E}_k\{\xi_{t+1}^k(x_{t+1}, d_{t+1}^k, \ldots, d_\tau^k)\} \geq -b$ for all $x_{t+1} \in \mathbb{R}$, we now show that the result holds for time period t and we have $\mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\} \geq -b$ 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

$\mathbb{E}_k\big\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\}$, 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 $\phi_t < \hat{x}_t \leq \tilde{x}_t$, (3.27) implies that $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h + \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$. 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 $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = h + \mathbb{E}_k\big\{\xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\}$ and $\mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = h + \mathbb{E}_k\big\{\xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\}$. Thus, the induction hypothesis implies that $\mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} \geq \mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} \geq -b$.

Case 1.b. Assume that $\hat{x}_t \leq \phi_t < \tilde{x}_t$. We have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = -b$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$ by (3.27). Taking expectations conditional on $d_t^k = \phi_t$, the induction hypothesis implies that $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = -b \leq h + \mathbb{E}_k\big\{\xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\} = \mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\}$.

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

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

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. □

3.5 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 (3.28), which implies that the algorithm proposed in Section 3.4 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 $\{d_t^k : t = 1, \ldots, \tau\}$ in Step 2 of Algorithm 1. Instead, we simulate the behavior of the policy characterized by the base-stock levels $\{r_t^k : t = 1, \ldots, \tau\}$ 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 $r_t^{k+1} = r_t^k - \alpha^k s_t^k(r_t^k, d_t^k, \ldots, d_\tau^k)$ for all $t = 1, \ldots, \tau$.

Therefore, we only have access to $\{\min\{y_t^k,\, d_t^k\} : t = 1, \ldots, \tau\}$, but not the demand random variables themselves. Proposition 23 shows that this information is adequate to compute the step direction.

Proposition 23 *Knowledge of* $\{r_t^k : t = 1, \ldots, \tau\}$, $\{y_t^k : t = 1, \ldots, \tau\}$ *and* $\{\min\{y_t^k,\, d_t^k\} : t = 1, \ldots, \tau\}$ *is adequate to compute* $s_t^k(r_t^k, d_t^k, \ldots, d_\tau^k)$ *in* (3.28) *for all* $t = 1, \ldots, \tau$.

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 $r_s^k - d_s^k \leq y_s^k - d_s^k \leq x_{s+1}^k \leq y_{s+1}^k$, $y_{s+1}^k - d_{s+1}^k \leq y_{s+2}^k, \ldots, y_{t-1}^k - d_{t-1}^k \leq y_t^k$ for all $t = s{+}1, \ldots, \tau$. Combining these inequalities, we have $r_s^k - d_s^k - d_{s+1}^k - \ldots - d_{t-1}^k \leq y_t^k$ for all $t = s+1, \ldots, \tau$. Consequently, if we have $\min\{y_t^k, d_t^k\} = y_t^k$ for any time period t, then we must have $r_s^k \leq d_s^k + d_{s+1}^k + \ldots + d_t^k$ for all $s = 1, \ldots, t-1$.

Assume that we want to compute $s_t^k(r_t^k, d_t^k, \ldots, d_\tau^k)$, where we have $s_t^k(r_t^k, d_t^k, \ldots, d_\tau^k) = c + h\,\mathbf{1}(d_t^k < r_t^k) - b\,\mathbf{1}(d_t^k \geq r_t^k) + \xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < r_t^k)$. 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 $s_t^k(r_t^k, d_t^k, \ldots, d_\tau^k) = c - b$ 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 $\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)$ 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 (3.27), we have

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

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 $\xi_{t+2}^k(r_t^k - d_t^k - d_{t+1}^k, d_{t+2}^k, \ldots, d_\tau^k)$ and $\xi_{t+2}^k(r_{t+1}^k - d_{t+1}^k, d_{t+2}^k, \ldots, d_\tau^k)$ 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. $\square$

3.6 Inventory Purchasing Problem under Price Uncertainty

We want to make purchasing decisions for a product over the time periods $\{1, \ldots, \tau\}$. 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 class from Nascimento and Powell (2006). Its one application area 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 $\{p_t : t = 1, \ldots, \tau\}$ 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 the product purchased up to time period t, the optimal policy is characterized by the Bellman equations

$$v_t(x_t) = \mathbb{E}\Big\{ \min_{y_t \geq x_t} \; p_t \, [y_t - x_t] + v_{t+1}(y_t) \Big\}, \tag{3.29}$$

with $v_{\tau+1}(x_{\tau+1}) = b \, \mathbb{E}\big\{[d - x_{\tau+1}]^+\big\}$. 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 $\{r_t^*(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$. That is, if the total amount of the 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

$$\dot{f}_t(r_t, p_t) = p_t + \dot{v}_{t+1}(r_t), \tag{3.30}$$

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

$$\Delta_t(x_t, p_t) = p_t + \dot{v}_{t+1}(x_t), \tag{3.31}$$

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 $\hat{p}_t \in \mathcal{P}_t$, $t = 1, \ldots, \tau$, where $\{r_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$ are the estimates of the optimal base-stock levels at iteration k. Similar to Sections 3.3 and 3.4, we now approximate the derivatives of $\{f_t(\cdot, \hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$ in a tractable manner.

Using the optimal base-stock level $r_t^*(p_t)$, we write (3.29) as $v_t(x_t) = \mathbb{E}\big\{v_t(x_t, p_t)\big\}$, where

$$v_t(x_t, p_t) = \begin{cases} v_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ p_t\,[r_t^*(p_t) - x_t] + v_{t+1}(r_t^*(p_t)) & \text{if } x_t < r_t^*(p_t). \end{cases} \tag{3.32}$$

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

$$\dot{v}_t(x_t, p_t) = \begin{cases} \dot{v}_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ -p_t & \text{if } x_t < r_t^*(p_t). \end{cases} \tag{3.33}$$

Since $r_t^*(p_t)$ is the minimizer of $f_t(\cdot, p_t)$, (3.30) implies that $-p_t = \dot{v}_{t+1}(r_t^*(p_t))$ and we obtain

$$\dot{v}_t(x_t, p_t) = \begin{cases} \dot{v}_{t+1}(x_t) & \text{if } x_t \geq r_t^*(p_t) \\ \dot{v}_{t+1}(r_t^*(p_t)) & \text{if } x_t < r_t^*(p_t). \end{cases} \tag{3.34}$$

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

$$\xi_t^k(x_t, p_t, \ldots, p_\tau, d) = \begin{cases} \xi_{t+1}^k(x_t, p_{t+1}, \ldots, p_\tau, d) & \text{if } x_t \geq r_t^k(p_t) \\ \xi_{t+1}^k(r_t^k(p_t), p_{t+1}, \ldots, p_\tau, d) & \text{if } x_t < r_t^k(p_t), \end{cases} \tag{3.35}$$

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 (3.31) with $\xi_{t+1}^k(x_t, p_{t+1}, \ldots, p_\tau, d)$, we use

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

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 $\{r_t^1(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

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

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

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

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

We emphasize that only the base-stock levels $\{r_t^k(p_t^k) : t = 1, \ldots, \tau\}$ are "updated" at iteration k in Step 2 of Algorithm 2. The other base-stock levels $\{r_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t \setminus \{p_t^k\},\ t = 1, \ldots, \tau\}$ remain the same. The proof of convergence for Algorithm 2 follows from an argument similar to the one in Sections 3.3.1 and 3.3.2. We can follow the proof of Lemma 14 to derive bounds on $\xi_t^k(\cdot, p_t^k, \ldots, p_\tau^k, d^k)$ and $s_t^k(\cdot, p_t^k, \ldots, p_\tau^k, d^k)$, and the proof of Lemma 15 to show that $\mathbb{E}_k\big\{\xi_t^k(\cdot, p_t^k, \ldots, p_\tau^k, d^k)\big\}$ is increasing. We define the error function as

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

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

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

w.p.1 for all $x_t \in \mathbb{R}$, $\hat{p}_t \in \mathcal{P}_t$, $t = 1, \ldots, \tau$, $k = 1, 2, \ldots$. Once we have this bound on the error function, we can follow the same induction argument in Proposition 17, Lemmas 18-20 and Proposition 21 to show the final result. In particular, we can show that $\lim_{k \to \infty} \dot{f}_t(r_t^k(\hat{p}_t), \hat{p}_t) = 0$ w.p.1 for all $\hat{p}_t \in \mathcal{P}_t$, $t = 1, \ldots, \tau$.

3.7 Numerical Illustrations

This section focuses on the problems described in Sections 3.3, 3.4 and 3.6, and numerically compares the performances of Algorithms 1 and 2 with standard stochastic approximation methods.

3.7.1 Multi-Period Newsvendor Problem with Backlogged Demands

We consider a policy characterized by the base-stock levels $\{r_t : t = 1, \ldots, \tau\}$. 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 $\{d_t : t = 1, \ldots, \tau\}$, then the inventory position at time period t is given by

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

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

$$\begin{aligned} H_t(x_1, D \,|\, r) &= h\,[\max\{x_t, r_t\} - d_t]^+ \\ &= h\left[\max\left\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \ldots, r_{t-1} - D_{t-1}^{t-1}, r_t\right\} - d_t\right]^+ \\ &= h\,\max\left\{x_1 - D_1^t, r_1 - D_1^t, \ldots, r_{t-1} - D_{t-1}^t, r_t - D_t^t, 0\right\}, \quad (3.36) \end{aligned}$$

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

$$\begin{aligned} B_t(x_1, D \,|\, r) &= b\,[d_t - \max\{x_t, r_t\}]^+ \\ &= b\,\big[\, d_t - \max\big\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \ldots, r_{t-1} - D_{t-1}^{t-1}, r_t\big\}\big]^+ \\ &= b\,\max\big\{\min\big\{D_1^t - x_1, D_1^t - r_1, \ldots, D_{t-1}^t - r_{t-1}, D_t^t - r_t\big\}, 0\big\}, \end{aligned} \quad (3.37)$$

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

$$\begin{aligned} C_t(x_1, D \,|\, r) &= c\,[r_t - x_t]^+ \\ &= c\,\big[\, r_t - \max\big\{x_1 - D_1^{t-1}, r_1 - D_1^{t-1}, \ldots, r_{t-1} - D_{t-1}^{t-1}\big\}\big]^+ \\ &= c\,\max\big\{\min\big\{r_t - x_1 + D_1^{t-1}, r_t - r_1 + D_1^{t-1}, \ldots, \\ &\qquad\qquad r_t - r_{t-1} + D_{t-1}^{t-1}\big\}, 0\big\}. \end{aligned} \quad (3.38)$$

Therefore, we can try to solve the problem $\min_r \mathbb{E}\big\{\sum_{t=1}^{\tau}\big[H_t(x_1, D \,|\, r) + B_t(x_1, D \,|\, r) + C_t(x_1, D \,|\, r)\big]\big\}$ 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 $\{\zeta_t : t = 1, \ldots, \tau\}$ 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

$$\min_r \; \mathbb{E}\left\{\sum_{t=1}^{\tau}\big[\, H_t(x_1, D \,|\, r+\zeta) + B_t(x_1, D \,|\, r+\zeta) + C_t(x_1, D \,|\, r+\zeta)\big]\right\}, \quad (3.39)$$

where we use $r+\zeta$ to denote the perturbed base-stock levels $\{r_t+\zeta_t : t = 1, \ldots, \tau\}$. It is now possible to show that the objective function of problem (3.39) 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 (3.39) instead of the original problem should not cause too much error.

After straightforward algebraic manipulations on (3.36) and (3.37), 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

$$\nabla_s H_t(x_1, D \,|\, r) = \begin{cases} h\,\mathbf{1}(r_s - D_s^t \geq 0) \times \mathbf{1}(r_s - D_s^t \geq x_1 - D_1^t) \\ \quad \times \mathbf{1}(r_s - D_s^t \geq r_1 - D_1^t) \times \ldots \\ \qquad \ldots \times \mathbf{1}(r_s - D_s^t \geq r_t - D_t^t) & \text{if } s \leq t \\ 0 & \text{if } s > t, \end{cases} \tag{3.40}$$

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

$$\nabla_s B_t(x_1, D \,|\, r) = \begin{cases} -b\,\mathbf{1}(D_s^t - r_s \geq 0) \times \mathbf{1}(D_s^t - r_s \leq D_1^t - x_1) \\ \quad \times \mathbf{1}(D_s^t - r_s \leq D_1^t - r_1) \times \ldots \\ \qquad \ldots \times \mathbf{1}(D_s^t - r_s \leq D_t^t - r_t) & \text{if } s \leq t \\ 0 & \text{if } s > t. \end{cases} \tag{3.41}$$

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 $\{r_t : t = 1, \ldots, \tau\}$ with $\{r_t + \zeta_t : t = 1, \ldots, \tau\}$ 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 (3.38) 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

$$\nabla_s C_t(x_1, D \,|\, r) = \begin{cases} -c\,\mathbf{1}(r_t - r_s + D_s^{t-1} \geq 0) \\ \quad \times\,\mathbf{1}(D_s^{t-1} - r_s \leq D_1^{t-1} - x_1) \\ \qquad \times\,\mathbf{1}(D_s^{t-1} - r_s \leq D_1^{t-1} - r_1) \times \ldots \\ \qquad\quad \ldots \times \mathbf{1}(D_s^{t-1} - r_s \leq D_{t-1}^{t-1} - r_{t-1}) & \text{if } s < t \\ c\,\mathbf{1}(r_t - x_1 + D_1^{t-1} \geq 0) \\ \quad \times\,\mathbf{1}(r_t - r_1 + D_1^{t-1} \geq 0) \times \ldots \\ \qquad\quad \ldots \times \mathbf{1}(r_t - r_{t-1} + D_{t-1}^{t-1} \geq 0) & \text{if } s = t \\ 0 & \text{if } s > t. \end{cases} \tag{3.42}$$

Consequently, the following algorithm is a standard stochastic approximation method for solving problem (3.39).

Algorithm 3

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

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

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

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

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 (3.39).

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 $\{(\mu_t, \sigma_t) : t = 1, \ldots, \tau\}$, $\{(l_t, u_t) : t = 1, \ldots, \tau\}$, $\{\lambda_t : t = 1, \ldots, \tau\}$ and $\{(\alpha_t^1, \alpha_t^2) : t = 1, \ldots, \tau\}$, we sample μ_t, l_t, u_t, λ_t, α_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 and 25 sample paths. Each sample path starts from a different initial solution and uses a different sequence of the samples of the demand random variables. To be fair, the s-th sample path for both algorithms starts from the same initial solution and uses the same sequence of the samples of the demand random variables. We choose an initial solution $\{r_t^1 : t = 1, \ldots, \tau\}$ 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 $r^{\mathrm{A1}}(s) = \{r_t^{\mathrm{A1}}(s) : t = 1, \ldots, \tau\}$ and $r^{\mathrm{A3}}(s) = \{r_t^{\mathrm{A3}}(s) : t = 1, \ldots, \tau\}$ respectively be the base-stock levels obtained by Algorithms 1 and 3 after the 10,000-th iteration of the s-th sample path, we are interested in the performance measures

$$\mathrm{AV}^{\mathrm{A}} = \frac{1}{25} \sum_{s=1}^{25} \mathbb{E} \left\{ \sum_{t=1}^{\tau} \left[H_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + B_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + C_t(x_1, D \,|\, r^{\mathrm{A}}(s)) \right] \right\}$$

$$\mathrm{MX}^{\mathrm{A}} = \max_{s \in \{1,\ldots,25\}} \left\{ \mathbb{E} \left\{ \sum_{t=1}^{\tau} \left[H_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + B_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + C_t(x_1, D \,|\, r^{\mathrm{A}}(s)) \right] \right\} \right\}$$

$$\mathrm{MI}^{\mathrm{A}} = \min_{s \in \{1,\ldots,25\}} \left\{ \mathbb{E} \left\{ \sum_{t=1}^{\tau} \left[H_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + B_t(x_1, D \,|\, r^{\mathrm{A}}(s)) + C_t(x_1, D \,|\, r^{\mathrm{A}}(s)) \right] \right\} \right\}$$

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

Our first set of computational results are summarized in Table 3.1. The first column in this table shows the problem parameters by using the triplets $(\tau, h, d) \in \{5, 10\} \times \{0.1, 0.25\} \times \{\mathrm{NR}, \mathrm{UN}, \mathrm{EX}, \mathrm{BT}\}$, 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

Table 3.1: Computational results for the multi-period newsvendor problem with backlogged demands.

Problem	OP	$\mathrm{AV}^{\mathrm{A1}}$	$\mathrm{MX}^{\mathrm{A1}}$	$\mathrm{MI}^{\mathrm{A1}}$	$\frac{\mathrm{AV}^{\mathrm{A1}}}{\mathrm{OP}}$	$\frac{\mathrm{MX}^{\mathrm{A1}}}{\mathrm{OP}}$	$\mathrm{AV}^{\mathrm{A3}}$	$\mathrm{MX}^{\mathrm{A3}}$	$\mathrm{MI}^{\mathrm{A3}}$	$\frac{\mathrm{AV}^{\mathrm{A3}}}{\mathrm{OP}}$	$\frac{\mathrm{MX}^{\mathrm{A3}}}{\mathrm{OP}}$
(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

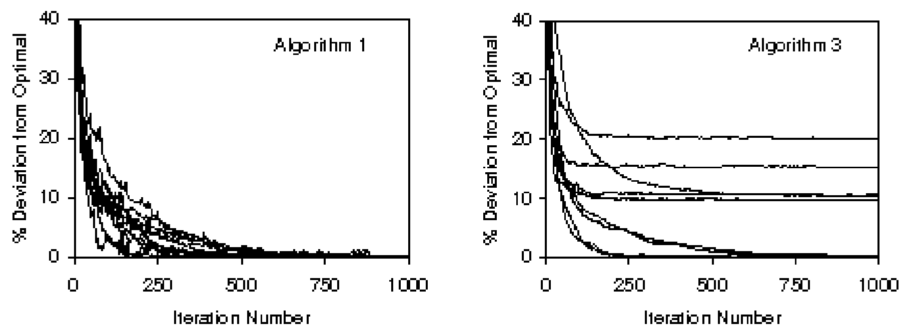

Figure 3.2: Performances of Algorithms 1 and 3 on test problem (10, 0.1, NR) for 10 sample paths.

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. Figure 3.2 shows the performances of Algorithms 1 and 3 on test problem (10, 0.1, NR) for 10 sample paths. Similar to Table 3.1, this figure shows that the performance Algorithm 1 is always close to optimal, but the performance of Algorithm 3 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 $(10, 0.1, \text{NR})^{\epsilon}$. We compute the optimal base-stock levels for test problem $(10, 0.1, \text{NR})^{\epsilon}$ and use these base-stock levels as the initial solution when computing the optimal-base stock levels for test problem $(10, 0.1, \mathrm{NR})$. Figure 3.3 shows the performances of Algorithms 1 and 3 on test problem $(10, 0.1, \mathrm{NR})$ for 10 sample paths 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 sample paths and their performances are essentially identical. If we have $\epsilon = 75$, then both algorithms converge to the optimal base-stock levels for all sample paths, 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 sample paths, 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 sample paths.

3.7.2 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 $\{r_t : t = 1, \ldots, \tau\}$ starting with the initial inventory position x_1 and the demands over the planning horizon turn out to be $\{d_t : t = 1, \ldots, \tau\}$, then the inventory position at time period t is given by

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

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

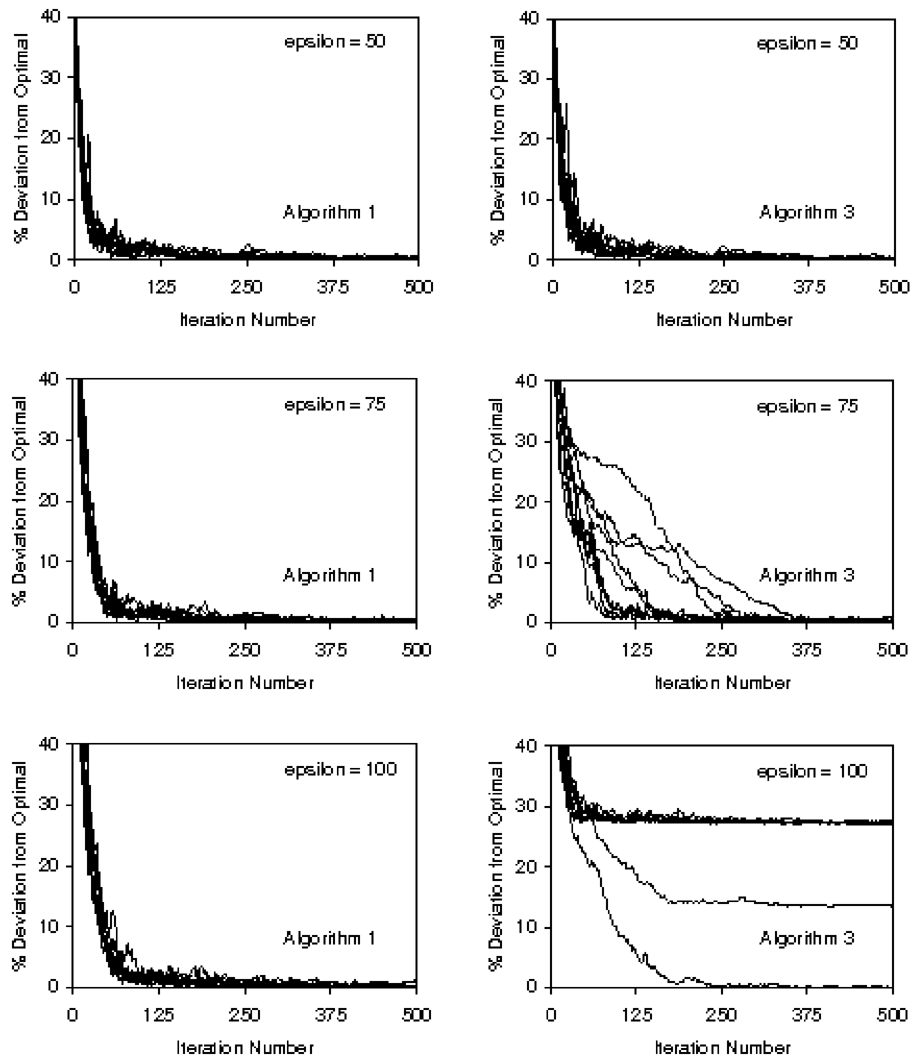

Figure 3.3: Performances of Algorithms 1 and 3 on test problem (10, 0.1, NR) for 10 sample paths starting from the initial solutions that are chosen carefully.

modify (3.36)-(3.38), (3.40)-(3.42) and Algorithm 3 in a straightforward manner to come up with a stochastic approximation method to solve problem (3.39) under the assumption that the unsatisfied demand is immediately lost.

Our computational results are summarized in Table 3.2. The entries in this table have the same interpretations as the ones in Table 3.1. Similar to our computational results in Table 3.1, even the worst-case performance of Algorithm 1 is always close to optimal. 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 8% and 47% worse than the performance of the optimal policy.

3.7.3 Inventory Purchasing Problem under Price Uncertainty

We consider a policy characterized by the base-stock levels $\{r_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$. That is, if the total amount of the 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 $\{p_t : t = 1, \ldots, \tau\}$, then the inventory position at time period t is given by $x_t = \max\big\{x_1, r_1(p_1), \ldots, r_{t-1}(p_{t-1})\big\}$. 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 periods. In this case, the purchasing cost that

Table 3.2: Computational results for the multi-period newsvendor problem with lost sales.

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

we incur at time period t is

$$
\begin{aligned}
C_t(x_1, p \,|\, r) &= p_t \,[r_t(p_t) - x_t]^+ \\
&= p_t \max\big\{ r_t(p_t) - \max\big\{ x_1, r_1(p_1), \ldots, r_{t-1}(p_{t-1}) \big\}, 0 \big\} \\
&= p_t \max\big\{ \min\big\{ r_t(p_t) - x_1, r_t(p_t) - r_1(p_1), \ldots, \\
&\qquad\qquad r_t(p_t) - r_{t-1}(p_{t-1}) \big\}, 0 \big\}, \qquad (3.43)
\end{aligned}
$$

where we use p to denote the prices $\{p_t : t = 1, \ldots, \tau\}$ and r to denote the base-stock levels $\{r_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$. 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

$$
\begin{aligned}
B_{\tau+1}(x_1, p, d \,|\, r) &= b\,[d - x_{\tau+1}]^+ = b \max\big\{ d - \max\big\{ x_1, r_1(p_1), \ldots, r_\tau(p_\tau) \big\}, 0 \big\} \\
&= b \max\big\{ \min\big\{ d - x_1, d - r_1(p_1), \ldots, d - r_\tau(p_\tau) \big\}, 0 \big\}. \qquad (3.44)
\end{aligned}
$$

Therefore, we can try to solve the problem $\min_r \mathbb{E}\big\{ \sum_{t=1}^{\tau} C_t(x_1, p \,|\, r) + B_{\tau+1}(x_1, p, d \,|\, r) \big\}$ to compute the optimal base-stock levels. Similar to the situation in Section 3.7.1, 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 $\{\zeta_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$ that are uniformly distributed over the small interval $[0, \epsilon]$. Consequently, we solve the problem

$$
\min_r \ \mathbb{E}\left\{ \sum_{t=1}^{\tau} C_t(x_1, p \,|\, r + \zeta) + B_{\tau+1}(x_1, p, d \,|\, r + \zeta) \right\}, \qquad (3.45)
$$

where we use $r + \zeta$ to denote the perturbed base-stock levels $\{r_t(\hat{p}_t) + \zeta_t(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$. It is now possible to show that the objective function of problem (3.45) 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 (3.45).

Similar to (3.40)-(3.42), after straightforward algebraic manipulations on (3.43) and (3.44), 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

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

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

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

The remarks for (3.40)-(3.42) 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 (3.45).

Algorithm 4

Step 1. Initialize the estimates of the optimal base-stock levels $\{r_t^1(\hat{p}_t) : \hat{p}_t \in$

$\mathcal{P}_t,\ t = 1, \ldots, \tau\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{p_t^k : t = 1, \ldots, \tau\}$ be the price random variables, d^k be the demand random variable and $\{\zeta_t^k(\hat{p}_t) : \hat{p}_t \in \mathcal{P}_t,\ t = 1, \ldots, \tau\}$ 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 $t = 1, \ldots, \tau$. Furthermore, set $r_t^{k+1}(\hat{p}_t) = r_t^k(\hat{p}_t)$ for all $\hat{p}_t \in \mathcal{P}_t \setminus \{p_t^k\}$, $t = 1, \ldots, \tau$.

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 $\{1, \ldots, b\}$ and the demand is uniformly distributed over the interval $[0, 1]$. Our experimental setup is the same as the one in Section 3.7.1 and our computational results are summarized in Table 3.3. 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 3.1.

Similar to our computational results in Tables 3.1 and 3.2, 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.

Table 3.3: Computational results for the inventory purchasing problem under price uncertainty.

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

3.8 Conclusions

We proposed three stochastic approximation methods to compute the optimal base-stock levels in three problem classes for which the so-called base-stock policies are known to be optimal. The proposed methods enjoy the well-known advantages of the stochastic approximation methods. They work with samples of the random variables and 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 3.3, 3.4 and 3.6 to a certain extent. Equations (3.5), (3.22) and (3.30) characterize the first order conditions that must be satisfied by the optimal base-stock levels. However, finding a solution to these equations through stochastic approximation methods requires knowing the function $\dot{v}_{t+1}(\cdot)$. In (3.10), (3.26) and (3.34), we come up with recursive expressions that can be used to compute the stochastic gradients of $\{v_t(\cdot) : t = 1, \ldots, \tau\}$. At iteration k, we "mimic" these expressions by using the estimates of the optimal base-stock levels as in (3.11), (3.27) and (3.35). The convergence proofs are based on analyzing the error function.

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

3.9 Appendix

3.9.1 Proof of Lemma 19

This section completes the proof of Lemma 19 by showing that $\text{limsup}_{k\to\infty}|\dot{f}_t(r_t^k)| = 0$.

For $\epsilon > 0$, we call $\{k', k'+1, \ldots, k''\}$ 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 $\epsilon/2 \le |\dot{f}_t(r_t^k)| \le \epsilon$ for all $k = k'+1, \ldots, k''-1$. Our proof shows that there exist only a finite number of upcrossing intervals from $\epsilon/2$ to ϵ. Since we have $\text{liminf}_{k\to\infty}|\dot{f}_t(r_t^k)| = 0$, this implies that $\text{limsup}_{k\to\infty}|\dot{f}_t(r_t^k)| \le \epsilon$.

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 $\{k'_n, k'_n+1, \ldots, k''_n\}$ be the n-th upcrossing interval. We have

$$|\dot{f}_t(r_t^{k'_n+1})| - |\dot{f}_t(r_t^{k'_n})| \le |\dot{f}_t(r_t^{k'_n+1}) - \dot{f}_t(r_t^{k'_n})|$$
$$\le L\,|r_t^{k'_n+1} - r_t^{k'_n}| = \alpha^{k'_n} L\,|s_t^{k'_n}(r_t^{k'_n}, d_t^{k'_n}, \ldots, d_\tau^{k'_n})| \le L\,M\,\alpha^{k'_n},$$

where we use Lemmas 13 and 14. Since $|\dot{f}_t(r_t^{k'_n+1})| \ge \epsilon/2$ and $\lim_{k\to\infty}\alpha^k = 0$, the chain o f inequalities above imply that there exists a finite number $\hat{N}$ such that we have $|\dot{f}_t(r_t^{k'_n})| \ge \epsilon/4$ for all $n = \hat{N}, \hat{N}+1, \ldots$. Since $\lim_{k\to\infty}\mathbb{E}_k\big\{|e_{t+1}^k(r_t^k - d_t^k)|\big\} = 0$ by the argument at the beginning of the proof of Lemma 19, there exists a finite number $\tilde{N}$ such that we have $\mathbb{E}_k\big\{|e_{t+1}^k(r_t^k - d_t^k)|\big\} \le \epsilon/8$ for all $k = k'_{\tilde{N}}, k'_{\tilde{N}}+1, \ldots$. Therefore, letting $N = \max\{\hat{N}, \tilde{N}\}$, we have

$$|\dot{f}_t(r_t^{k'_n})| \ge \epsilon/4 \quad \text{and} \quad \mathbb{E}_k\big\{|e_{t+1}^k(r_t^k - d_t^k)|\big\} \le \epsilon/8 \tag{3.46}$$

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

On the other hand, using Lemmas 13 and 14, we have

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

which implies that $\sum_{k=k'_n}^{k''_n-1} \alpha^k \ge \epsilon/[2\,L\,M]$. Therefore, using (3.46), we obtain

$$\sum_{n=N}^{\infty} \sum_{k=k'_n}^{k''_n-1} \alpha^k\,|\dot{f}_t(r_t^k)|^2 \ge \sum_{n=N}^{\infty} \frac{\epsilon^2}{16} \sum_{k=k'_n}^{k''_n-1} \alpha^k \ge \sum_{n=N}^{\infty} \frac{\epsilon^3}{32\,L\,M} = \infty. \tag{3.47}$$

We have $|\dot{f}_t(r_t^k)|/2 \ge \epsilon/8 \ge \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}$ for all $n = N, N+1, \ldots$, $k = k'_n, k'_n+1, \ldots, k''_n - 1$ by (3.46) and the definition of an upcrossing interval. This implies that

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

for all $n = N, N+1, \ldots$, $k = k'_n, k'_n+1, \ldots, k''_n-1$. Using this chain of inequalities, since (E.2) holds for time period t by Lemma 18, we obtain

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

This expression contradicts (3.47). Therefore, we must have a finite number of upcrossing intervals from $\epsilon/2$ to ϵ, which implies that $\limsup_{k\to\infty}|\dot{f}_t(r_t^k)| \le \epsilon$. Since ϵ is arbitrary, we obtain the desired result. □

3.9.2 Multi-Period Newsvendor Problem with Lost Sales and Stationary Cost Parameters

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.

We consider the setting described in Section 3.4 and show that the error function $e_t^k(x_t) = \dot{v}_t(x_t)\,\mathbf{1}(x_t > 0) - \mathbb{E}_k\big\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\,\mathbf{1}(x_t > 0)\big\}$ satisfies the same bound given in Lemma 16. Once we have this bound, we can follow the same induction argument in Proposition 17, Lemmas 18-20 and Proposition 21 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.

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 (3.22) and (3.27), we obtain

$$\mathbb{E}_k\big\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\big\} = \begin{cases} h\,\mathbb{P}_k\big\{d_t^k < x_t\big\} - b\,\mathbb{P}_k\big\{d_t^k \geq x_t\big\} & \\ +\mathbb{E}_k\big\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < x_t)\big\} & \text{if } x_t \geq r_t^k \\ h\,\mathbb{P}_k\big\{d_t^k < r_t^k\big\} - b\;\mathbb{P}_k\big\{d_t^k \geq r_t^k\big\} & \\ +\mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < r_t^k)\big\} & \text{if } x_t < r_t^k \end{cases}$$

$$= \begin{cases} \dot{f}_t(x_t) - c - \mathbb{E}_k\big\{\dot{v}_{t+1}(x_t - d_t^k)\,\mathbf{1}(d_t^k < x_t)\big\} \\ +\mathbb{E}_k\big\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < x_t)\big\} & \text{if } x_t \geq r_t^k \\ \dot{f}_t(r_t^k) - c - \mathbb{E}_k\big\{\dot{v}_{t+1}(r_t^k - d_t^k)\,\mathbf{1}(d_t^k < r_t^k)\big\} \\ +\mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < r_t^k)\big\} & \text{if } x_t < r_t^k. \end{cases} \tag{3.48}$$

Since r_t^* is the minimizer of the convex function $f_t(\cdot)$, we have $\dot{f}_t(r_t^*) = 0$. Using (3.22) and (3.25), we obtain

$$\dot{v}_t(x_t) = \begin{cases} \dot{f}_t(x_t) - c & \text{if } x_t \geq r_t^* \\ -c & \text{if } x_t < r_t^*. \end{cases} \tag{3.49}$$

We consider four cases. First, we assume that $x_t \geq r_t^k$ and $x_t \geq r_t^*$. Using (3.48) and (3.49), we have $e_t^k(x_t) = \mathbb{E}_k\big\{\dot{v}_{t+1}(x_t - d_t^k)\,\mathbf{1}(d_t^k < x_t)\big\} - \mathbb{E}_k\big\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < x_t)\big\} = \mathbb{E}_k\big\{e_{t+1}^k(x_t - d_t^k)\big\}$. Therefore, we obtain $|e_t^k(x_t)| \leq \mathbb{E}_k\big\{|e_{t+1}^k(x_t - d_t^k)|\big\}$ by Jensen's inequality.

Second, we assume that $x_t \geq r_t^k$ and $x_t < r_t^*$. We have $\mathbb{E}_k\big\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\big\} \geq \mathbb{E}_k\big\{\xi_t^k(r_t^k, d_t^k, \ldots, d_\tau^k)\big\}$ by Lemma 22. Using this inequality, (3.48) and (3.49), w e obtain

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

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 (3.48), we also obtain

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

The last two chains of inequalities imply that

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

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 (3.48) and (3.49), we obtain

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

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

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

Fourth, we assume that $x_t < r_t^k$ and $x_t < r_t^*$. In this case, (3.48) and (3.49) imply that

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

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.

3.9.3 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 $t = 1, \ldots, \tau$, 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 $\{v_t(\cdot) : t = 1, \ldots, \tau\}$ and $\{f_t(\cdot) : t = 1, \ldots, \tau\}$ defined in Section 3.4 are convex.

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

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

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

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

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 $\{r_t^* : t = 1, \ldots, \tau\}$ are the minimizers of the functions $\{f_t(\cdot) : t = 1, \ldots, \tau\}$.

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

$$\dot{f}_t(r_t) = c_t + h_t \, \mathbb{P}\big\{d_t < r_t\big\} - b_t \, \mathbb{P}\big\{d_t \geq r_t\big\} + \mathbb{E}\big\{\dot{v}_{t+1}(r_t - d_t) \; \mathbf{1}(d_t < r_t)\big\}, \tag{3.50}$$

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

$$\Delta_t(x_t, d_t) = c_t + h_t \, \mathbf{1}(d_t < x_t) - b_t \, \mathbf{1}(d_t \geq x_t) + \dot{v}_{t+1}(x_t - d_t)\mathbf{1}(d_t < x_t). \tag{3.51}$$

On the other hand, (3.25) implies that

$$\dot{v}_t(x_t) = \begin{cases} \dot{f}_t(x_t) - c_t & \text{if } x_t \geq r_t^* \\ -c_t & \text{if } x_t < r_t^*, \end{cases} \tag{3.52}$$

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

in which case

$$\dot{v}_t(x_t, d_t) = \begin{cases} h_t \, \mathbf{1}(d_t < x_t) - b_t \, \mathbf{1}(d_t \geq x_t) \\ \qquad + \dot{v}_{t+1}(x_t - d_t) \, \mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^* \\ h_t \, \mathbf{1}(d_t < r_t^*) - b_t \, \mathbf{1}(d_t \geq r_t^*) \\ \qquad + \dot{v}_{t+1}(r_t^* - d_t) \, \mathbf{1}(d_t < r_t^*) & \text{if } x_t < r_t^*, \end{cases} \tag{3.53}$$

gives a stochastic gradient of $v_t(\cdot)$ at x_t. To construct tractable approximations to the stochastic gradients of $\{f_t(\cdot) : t = 1, \ldots, \tau\}$, we "mimic" the computation in (3.53) by using the estimates of the optimal base-stock levels. In particular, letting

$\{r_t^k : t = 1, \ldots, \tau\}$ be the estimates of the optimal base-stock levels at iteration k, we recursively define

$$\xi_t^k(x_t, d_t, \ldots, d_\tau) = \begin{cases} h_t\,\mathbf{1}(d_t < x_t) - b_t\,\mathbf{1}(d_t \geq x_t) \\ \qquad + \xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)\,\mathbf{1}(d_t < x_t) & \text{if } x_t \geq r_t^k \\ h_t\,\mathbf{1}(d_t < r_t^k) - b_t\,\mathbf{1}(d_t \geq r_t^k) \\ \qquad + \xi_{t+1}^k(r_t^k - d_t, d_{t+1}, \ldots, d_\tau)\,\mathbf{1}(d_t < r_t^k) & \text{if } x_t < r_t^k, \end{cases} \tag{3.54}$$

with $\xi_{\tau+1}^k(\cdot, \cdot, \ldots, \cdot) = 0$. At iteration k, replacing $\dot{v}_{t+1}(x_t - d_t)$ in (3.51) with $\xi_{t+1}^k(x_t - d_t, d_{t+1}, \ldots, d_\tau)$, we use

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

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

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

with $e_{\tau+1}^k(\cdot) = 0$. We can easily follow the argument in the proof of Lemma 14 to derive bounds on $\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k)$ and $s_t^k(\cdot, d_t^k, \ldots, d_\tau^k)$. The following lemma is analogous to Lemma 15.

Lemma 24 *If $\hat{x}_t$, $\tilde{x}_t$ satisfy $\hat{x}_t \le \tilde{x}_t$, then we have*

$$\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, \ldots, d_\tau^k)\big\} \le \mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, \ldots, d_\tau^k)\big\} + \sum_{s=t+1}^{\tau} \mathbb{E}_k\big\{|e_s^k(\tilde{x}_t - \textstyle\sum_{s'=t}^{s-1} d_{s'}^k)|\big\} \tag{3.55}$$

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

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 15. 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 $r_t^k \le \hat{x}_t \le \tilde{x}_t$. We investigate the conditional expectation $\mathbb{E}_k\big\{\xi_t^k(\cdot, d_t^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\}$, 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 $\phi_t < \hat{x}_t \le \tilde{x}_t$, (3.54) implies that $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h_t + \xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h_t + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$. 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 $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = h_t + \mathbb{E}_k\big\{\xi_{t+1}^k(\hat{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\}$ and $\mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = h_t + \mathbb{E}_k\big\{\xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\}$. Thus, the induction hypothesis implies that

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

Case 1.b. Assume that $\hat{x}_t \le \phi_t < \tilde{x}_t$. We have $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = -b_t$ and $\xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = h_t + \xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$ by (3.54). Taking expectations conditional on $d_t^k = \phi_t$ and noting that $\tilde{x}_t - \phi_t > 0$, we have $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = -b_t$ and $\mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = h_t + \mathbb{E}_k\big\{\xi_{t+1}^k(\tilde{x}_t - \phi_t, d_{t+1}^k, \ldots, d_\tau^k)\big\} = h_t + \dot{v}_{t+1}(\tilde{x}_t - \phi_t) - e_{t+1}^k(\tilde{x}_t - \phi_t)$. Since r_t^* is the minimizer of the convex function $f_t(\cdot)$, (3.52) implies that we have $\dot{v}_t(\cdot) \geq -c_t$ for all $t = 1, \ldots, \tau$. Therefore, we obtain

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

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 $\xi_t^k(\hat{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k) = -b_t = \xi_t^k(\tilde{x}_t, \phi_t, d_{t+1}^k, \ldots, d_\tau^k)$. Taking expectations conditional on $d_t^k = \phi_t$, we obtain $\mathbb{E}_k\big\{\xi_t^k(\hat{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\} = -b_t = \mathbb{E}_k\big\{\xi_t^k(\tilde{x}_t, d_t^k, d_{t+1}^k, \ldots, d_\tau^k) \,|\, d_t^k = \phi_t\big\}$.

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

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. □

The following lemma is analogous to Lemma 16.

Lemma 25 *We have*

$$|e_t^k(x_t)| \le 2\max\Big\{\big|\dot f_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}\big|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}, \sum_{s=t+1}^{\tau} \mathbb{E}_k\{|e_s^k(x_t - \textstyle\sum_{s'=t}^{s-1} d_{s'}^k)|\}\Big\}$$

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

Proof Using (3.50) and (3.54), we have

$$\mathbb{E}_k\{\xi_t^k(x_t, d_t^k, \ldots, d_\tau^k)\} = \begin{cases} h_t\,\mathbb{P}_k\{d_t^k < x_t\} - b_t\,\mathbb{P}_k\{d_t^k \ge x_t\} \\ +\mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < x_t)\} & \text{if } x_t \ge r_t^k \\ h_t\,\mathbb{P}_k\{d_t^k < r_t^k\} - b_t\,\mathbb{P}_k\{d_t^k \ge r_t^k\} \\ +\mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < r_t^k)\} & \text{if } x_t < r_t^k \end{cases}$$

$$= \begin{cases} \dot f_t(x_t) - c_t - \mathbb{E}_k\{\dot v_{t+1}(x_t - d_t^k)\,\mathbf{1}(d_t^k < x_t)\} \\ +\mathbb{E}_k\{\xi_{t+1}^k(x_t - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < x_t)\} & \text{if } x_t \ge r_t^k \\ \dot f_t(r_t^k) - c_t - \mathbb{E}_k\{\dot v_{t+1}(r_t^k - d_t^k)\,\mathbf{1}(d_t^k < r_t^k)\} \\ +\mathbb{E}_k\{\xi_{t+1}^k(r_t^k - d_t^k, d_{t+1}^k, \ldots, d_\tau^k)\,\mathbf{1}(d_t^k < r_t^k)\} & \text{if } x_t < r_t^k. \end{cases} \tag{3.56}$$

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

$$|e_t^k(x_t)| \le \max\Big\{\big|\dot f_t(r_t^k) - \mathbb{E}_k\{e_{t+1}^k(r_t^k - d_t^k)\}\big|, \mathbb{E}_k\{|e_{t+1}^k(r_t^k - d_t^k)|\}, \mathbb{E}_k\{|e_{t+1}^k(x_t - d_t^k)|\}\Big\}$$

for the cases where $x_t \ge r_t^k$ and $x_t \ge r_t^*$, or $x_t < r_t^k$ and $x_t \ge r_t^*$, or $x_t < r_t^k$ and $x_t < r_t^*$ by following the argument in the proof of Lemma 16. 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

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

by Lemma 24. Using this inequality, 3.52) and (3.56), we obtain

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

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 (3.56), we also obtain

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

The last two chains of inequalities imply that

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

The result follows by combining the four cases. □

Finally, we have the following lemma, which is analogous to (E.4) and (E.5) in Proposition 17.

Lemma 26 *There exist constants $\mathcal{A}_t$ and $\mathcal{B}_t$ such that we have*

$$|e_t^k(x_t)| \le \mathcal{A}_t \sum_{s=t}^{\tau} |\dot{f}_s(r_s^k)| \tag{3.57}$$

$$|e_t^k(x_t)|^2 \le \mathcal{B}_t \sum_{s=t}^{\tau} \big[\, |\dot{f}_s(r_s^k)|^2 - \dot{f}_s(r_s^k)\,\mathbb{E}_k\big\{e_{s+1}^k(r_s^k - d_s^k)\big\}\big]^+ \tag{3.58}$$

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

Proof We show the result by induction over the time periods. Since $e_{\tau+1}^k(\cdot) = 0$, Lemma 25 shows that (3.57) and (3.58) hold for time period τ with $\mathcal{A}_\tau = 2$ and $\mathcal{B}_\tau = 4$. Assuming that the result holds for time periods $t+1, \ldots, \tau$, Lemma 25 and the induction hypothesis imply that

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

If we let $\mathcal{A}_t = 2\,(1 + \mathcal{A}_{t+1} + \sum_{s=t+1}^{\tau} \mathcal{A}_s)$, then (3.57) holds for time period t. For all $x_t \in \mathbb{R}$, squaring the bound in Lemma 25 also implies that

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

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

where the second inequality uses the fact that $\left[\sum_{i=1}^n a_i \right]^2 \leq n \sum_{i=1}^n a_i^2$ and Jensen's inequality, and the third inequality uses the induction hypothesis. If we let $\mathcal{B}_t = 4\,[2 + 2\,\mathcal{B}_{t+1} + (\tau - t) \sum_{s=t+1}^{\tau} \mathcal{B}_s]$, then (3.58) holds for time period t. □

Once we have these preliminary results, we can follow the same induction argument in Proposition 17, Lemmas 18-20 and Proposition 21 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 (E.4) and (E.5) in Proposition 17 are satisfied.

3.9.4 Inventory Purchasing Problem under Price Uncertainty

We consider the setting described in Section 3.6 and show that the error function $e_t^k(x_t, \hat{p}_t) = \dot{v}_t(x_t, \hat{p}_t) - \mathbb{E}_k\big\{ \xi_t^k(x_t, \hat{p}_t, p_{t+1}^k, \ldots, p_\tau^k, d^k) \big\}$ satisfies a bound similar to the one given in Lemma 16. Once we have this bound, we can follow the same induction argument in Proposition 17, Lemmas 18-20 and Proposition 21 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 (3.35), we obtain

$$\mathbb{E}_k\big\{\xi_t^k(x_t,\hat{p}_t,p_{t+1}^k\ldots,p_\tau^k,d^k)\big\} = \begin{cases} \mathbb{E}_k\big\{\xi_{t+1}^k(x_t,p_{t+1}^k,\ldots,p_\tau^k,d^k)\big\} & \text{if } x_t \geq r_t^k(\hat{p}_t) \\ \mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k(\hat{p}_t),p_{t+1}^k,\ldots,p_\tau^k,d^k)\big\} & \text{if } x_t < r_t^k(\hat{p}_t). \end{cases} \tag{3.59}$$

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 (3.34) and (3.59), we have $e_t^k(x_t,\hat{p}_t) = \dot{v}_{t+1}(x_t) - \mathbb{E}_k\big\{\xi_{t+1}^k(x_t,p_{t+1}^k\ldots,p_\tau,d)\big\} = \mathbb{E}_k\big\{e_{t+1}^k(x_t,p_{t+1}^k)\big\}$. Therefore, we obtain $|e_t^k(x_t,\hat{p}_t)| \leq \mathbb{E}_k\big\{|e_{t+1}^k(x_t,p_{t+1}^k)|\big\}$ 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 15, it is easy to show that $\mathbb{E}_k\big\{\xi_t^k(\cdot,\hat{p}_t,p_{t+1}^k,\ldots,p_\tau^k,d^k)\big\}$ is increasing. In this case, using (3.33) and (3.59), we obtain $e_t^k(x_t,\hat{p}_t) = -\hat{p}_t - \mathbb{E}_k\big\{\xi_{t+1}^k(x_t,p_{t+1}^k,\ldots,p_\tau^k,d^k)\big\} \leq -\hat{p}_t - \mathbb{E}_k\big\{\xi_{t+1}^k(r_t^k(\hat{p}_t),p_{t+1}^k,\ldots,p_\tau^k,d^k)\big\}$. Therefore, (3.30) implies that

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

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 (3.30), we also obtain

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

The last two chains of inequalities imply that

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

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, (3.30) 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 (3.34) and (3.59), we obtain

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

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

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

Fourth, we assume that $x_t < r_t^k(\hat{p}_t)$ and $x_t < r_t^*(\hat{p}_t)$. In this case, (3.30), (3.33) and (3.59) imply that

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

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

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

Chapter 4

A Stochastic Approximation Method for the Revenue Management Problem on a Single Flight Leg with Discrete Demand Distributions

4.1 Introduction

A recurrent problem in the revenue management literature involves optimally allocating the seats on a single flight leg to the demands from multiple fare classes that arrive sequentially. Given the demand from the current fare class and the number of unsold seats, the decision that needs to be made is how many seats to sell to the current fare class.

Starting with Littlewood (1972), this problem has been studied extensively and it is well-known that the optimal policy is characterized by one protection level for each fare class. Specifically, letting n be the number of fare classes, there exists a set of protection levels $\{y_j^* : j = 1, \ldots, n\}$ such that it is optimal to keep the number of unsold seats just after making the decisions for fare class j as close as possible to y_j^*. In other words, letting x_j be the number of unsold seats just before making the decisions for fare class j and $[\cdot]^+ = \max\{0, \cdot\}$, it is optimal to make $[x_j - y_j^*]^+$ seats available for sale to fare class j. If the demand from fare class j

does not exceed $[x_j - y_j^*]^+$, then all of the demand is satisfied. Otherwise, only $[x_j - y_j^*]^+$ seats are sold. This particular structure of the optimal policy arises from the fact that the value functions in the dynamic programming formulation of the problem are concave in the number of unsold seats. In this case, the computation of the optimal protection levels through the Bellman equations requires solving a number of convex optimization problems, which is a relatively simple task as long as the demand distributions are known.

We propose a stochastic approximation method to compute the optimal protection levels when the demand distributions are not known and we only have access to the samples from the demand distributions. We work with a particular version of the problem where the demand distributions are discrete and the fare classes that generate lower revenues arrive earlier than the fare classes that generate higher revenues. We develop a novel method that uses the dynamic programming formulation of the problem in conjunction with the samples from the demand distributions to approximate the stochastic subgradients of the value functions. By showing that our approximate stochastic subgradients are indeed accurate in the limit, we establish that the iterates of our stochastic approximation method converge to a set of optimal protection levels with probability one (w.p.1). To deal with the case where the demand information is censored by the seat availability, we provide alternative versions of our method that remain applicable when we can only observe the number of seats sold to a fare class, but not necessarily the amount of demand from a fare class.

Although there has been work on using stochastic approximation methods to compute the optimal protection levels, we make the following contributions.

Brumelle and McGill (1993) characterize the conditions that should be satisfied by the optimal protection levels and van Ryzin and McGill (2000) exploit these conditions to develop a stochastic approximation method. However, this method is tightly related to the optimality conditions in Brumelle and McGill (1993) and it is not clear whether it can be extended to another problem class. In contrast, we work with the dynamic programming formulation of the problem and it is possible to extend our method to inventory control problems where the value functions are convex and the base stock policies are optimal. Furthermore, the step directions used by our method are related to the stochastic subgradients of the value functions, whereas this is not the case for the method developed by van Ryzin and McGill (2000). Stochastic subgradients of the value functions can be particularly useful when making tactical decisions such as setting the capacity of the flight leg. Huh and Rusmevichientong (2006) also propose a stochastic approximation method to compute the optimal protection levels. There are similarities between their method and ours as both exploit the dynamic programming formulation of the problem, but Huh and Rusmevichientong (2006) use the results from the online convex optimization literature pioneered by Zinkevich (2003), whereas we use the stochastic approximation theory. Finally, both van Ryzin and McGill (2000) and Huh and Rusmevichientong (2006) work with continuous demand distributions. To our knowledge, our method is the only one that works with discrete demand distributions and has a convergence guarantee for the performance of the policy. To deal with discrete demand distributions, van Ryzin and McGill (2000) propose a randomized version of their method. The iterates of this version converge to a set of optimal protection levels, but the randomization results in suboptimality for the performance of the policy.

It is also important to note that the total expected revenue for the seat allocation problem is not concave when viewed as a function of the protection levels. To illustrate, assuming that there are two fare classes, and using c to denote the initial capacity and $\{r_1, r_2\}$, $\{D_1, D_2\}$ and $\{y_1, y_2\}$ to respectively denote the revenues, demand random variables and protection levels, the total expected revenue is

$$\begin{aligned} R(y_1, y_2) = r_1\, \mathbb{E}\big\{ \min\{[c - y_1]^+, D_1\}\big\} \\ + r_2\, \mathbb{E}\Big\{ \min\Big\{ \Big[c - \min\{[c - y_1]^+, D_1\} - y_2\Big]^+, D_2\Big\}\Big\}, \end{aligned} \tag{4.1}$$

where we use the fact that if there are x_j unsold seats just before making the decisions for fare class j, then we make $[x_j - y_j]^+$ seats available for sale to fare class j and sell $\min\{[x_j - y_j]^+, D_j\}$ seats. Figure 4.1 plots a cross section of $R(\cdot, \cdot)$ for a problem instance and shows that this function may not be concave. Consequently, if we naively attempt to compute the optimal protection levels by solving the problem $\max_{(y_1,y_2)} R(y_1, y_2)$ through a stochastic approximation method, then we do not necessarily obtain the optimal protection levels. We, however, show that it is possible to develop a stochastic approximation method to compute the optimal protection levels as long as we use step directions that are based on the dynamic programming formulation of the problem.

Since we work with discrete demand distributions, our stochastic approximation method has some drawbacks when the demand information is censored by the seat availability. Specifically, if such demand censorship occurs and we sell all of the seats that we make available for sale to a fare class, then we only know that the demand is greater than or equal to the seat availability. However, our method requires knowing whether the demand is strictly greater than the seat availability. If a somewhat *relaxed* view of demand censorship is possible and we can indeed

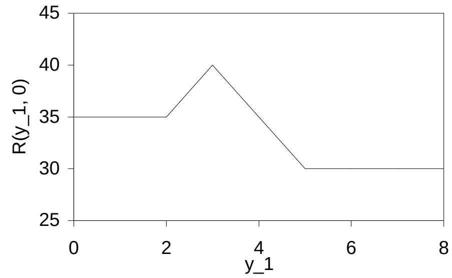

Figure 4.1: Plot of $\{R(y_1, 0) : y_1 = 0, \ldots, 8\}$. The problem parameters are $c = 5$, $r_1 = 5$, $r_2 = 10$. The demands from the two fare classes are deterministic and we have $D_1 = 3$, $D_2 = 3$ w.p.1.

observe whether the demand strictly exceeds the seat availability, then our method remains applicable with no modifications. This relaxed view of demand censorship is equivalent to assuming that we can observe the demand from the first customer that we turn down. We also provide two alternative versions of our method to address the case where the demand information is censored and the relaxed view of demand censorship is not possible. For the first alternative version, we show that the distance between its iterates and the optimal protection levels is bounded by the number of fare classes in the limit w.p.1. The second alternative version is a heuristic modification of the first one with somewhat more desirable practical performance but no convergence guarantee. We emphasize that the drawbacks mentioned in this paragraph arise solely due to the fact that we work with discrete demand distributions. When working with continuous demand distributions, the event that the demand is equal to the seat availability occurs with probability zero. In this case, there is essentially no distinction between knowing that the demand is greater than or equal to or strictly greater than the seat availability.

The rest of this chapter is organized as follows. Section 4.2 briefly reviews the other related literature. Section 4.3 gives a dynamic programming formulation of the seat allocation problem. Section 4.4 describes the stochastic approximation algorithm of van Ryzin and McGill (2000). Section 4.5 describes our stochastic approximation method and Section 4.6 proves its convergence. Section 4.7 considers the case where the demand information is censored by the seat availability. Section 4.8 provides numerical experiments, where we compare the performance of our method with the stochastic approximation algorithm of van Ryzin and McGill (2000).

4.2 Review of Other Related Literature

There is extensive literature on the seat allocation problem on a single flight leg. Although it makes a number of restrictive assumptions and the airlines almost invariably operate hub-and-spoke networks, the problem has important practical implications. For example, it justifies, at least to a certain extent, the use of protection level policies for complex networks. Furthermore, there are a variety of techniques to decompose the revenue management problem over a network into a sequence of single flight legs. Most of the literature on the single flight leg problem assumes that there are multiple fare classes and the demands from different fare classes occur over nonoverlapping time intervals. This ensures that we can formulate the problem as a dynamic program with the number of stages being equal to the number of fare classes. Motivated by the fact that leisure travelers tend to book earlier than the business travelers, it is also a common assumption that the fare classes that generate lower revenues arrive earlier than the fare classes that generate higher revenues. An interesting consequence of this second assumption is that the optimal protection levels are nested. That is, the optimal protection level for a fare class is greater than the optimal protection levels for the fare classes that arrive later.

Littlewood (1972), Curry (1990), Wollmer (1992) and Brumelle and McGill (1993) employ the two assumptions in the previous paragraph and show the optimality of protection level policies. Robinson (1995) characterizes the structure of the optimal policy under the assumption that the demands from different fare classes occur over nonoverlapping time intervals, but the fare classes that generate lower revenues do not necessarily arrive earlier. Lee and Hersh (1993) and Lautenbacher and Stidham (1999) focus on the single leg problem when the demands from different fare classes do not necessarily arrive over nonoverlapping time intervals. In the latter two cases, it is still possible to show that a variation of protection level policies is optimal. We refer the reader to Talluri and van Ryzin (2004) for a coverage of the related revenue management literature.

The use of stochastic approximation methods for solving stochastic optimization problems is well-known. Kushner and Clark (1978), Ermoliev (1988) and Bertsekas and Tsitsiklis (1996) give a coverage of the theory of stochastic approximation methods. There are numerous papers that use these methods for solving the revenue management problem over a network. In particular, van Ryzin and Vulcano (2004), Bertsimas and de Boer (2005) and van Ryzin and Vulcano (2006) describe methods to compute protection levels, Topaloglu (2007) describes a method to compute bid prices and Karaesmen and van Ryzin (2004) describe a method to compute overbooking limits. As far as other application areas are concerned, L'Ecuyer and Glynn (1994), Fu (1994), Glasserman and Tayur (1995), Bashyam and Fu (1998) and Mahajan and van Ryzin (2001) focus on queueing and inventory control. Kunnumkal and Topaloglu (2006) show that a method similar to ours can be used to compute the optimal base stock levels in inventory control problems. However, since they work with continuous demand distributions, their proof technique is considerably different from ours.

4.3 Problem Formulation

We want to use c seats available on a single flight leg to satisfy the demands from n fare classes that arrive sequentially. We index the fare classes such that the demand from fare class 1 arrives first and the demand from fare class n arrives last. If we sell a seat to fare class j, then we generate a revenue of r_j. We assume that the revenues satisfy $0 < r_1 \leq r_2 \leq \ldots \leq r_n$ so that the demands from the cheaper fare classes arrive earlier. The demands from different fare classes are random and we use D_j to denote the demand from fare class j. We assume that D_j is a positive and integer random variable and $\{D_j : j = 1, \ldots, n\}$ are independent of each other. We are interested in maximizing the total expected revenue from n fare classes.

If we use x_j to denote the remaining capacity just before making the decisions for fare class j, u_j to denote the number of seats sold to fare class j and d_j to denote a particular realization of D_j, then the optimal policy can be found by solving the optimality equations

$$v_j(x_j, d_j) = \max_{0 \leq u_j \leq \min\{x_j, d_j\}} r_j \, u_j + \mathbb{E}\big\{v_{j+1}(x_j - u_j, D_{j+1})\big\}, \tag{4.2}$$

with $v_{n+1}(\cdot, \cdot) = 0$. The constraints in the problem above ensure that the number of seats sold do not exceed the remaining capacity and the demand from fare class j. Alternatively, if we let $y_j = x_j - u_j$ be the remaining capacity just after making the decisions for fare class j, then (4.2) can be written as

$$v_j(x_j, d_j) = \left\{ \max_{[x_j - d_j]^+ \le y_j \le x_j} -r_j\, y_j + \mathbb{E}\big\{v_{j+1}(y_j, D_{j+1})\big\} \right\} + r_j\, x_j, \tag{4.3}$$

where the constraints follow from the fact that $x_j - \min\{x_j, d_j\} = \max\{0, x_j - d_j\}$. It is possible to show that $\{v_j(\cdot, D_j) : j = 1, \ldots, n\}$ are piecewise-linear concave functions with points of nondifferentiability being a subset of integers for all realizations of $\{D_j : j = 1, \ldots, n\}$. In this case, it is easy to show that the optimal policy is characterized by a set of protection levels $\{y_j^* : j = 1, \ldots, n\}$, where y_j^* can be computed as a maximizer of the function

$$f_j(y_j) = -r_j\, y_j + \mathbb{E}\big\{v_{j+1}(y_j, D_{j+1})\big\} \tag{4.4}$$

over the interval $[0, c]$. This is to say that if the remaining capacity just before making the decisions for fare class j is x_j and the demand from fare class j is d_j, then it is optimal to sell $\min\{[x_j - y_j^*]^+, d_j\}$ seats to fare class j. The protection level terminology is due to the fact that it is optimal to protect y_j^* seats for the demand from fare classes $\{j+1, \ldots, n\}$ when making the decisions for fare class j.

Since the demands from the cheaper fare classes arrive earlier, it is also possible to show that the optimal protection levels are nested. In other words, the optimal number of seats to protect for the demand from fare classes $\{j, \ldots, n\}$ is at least as large as the optimal number of seats to protect for the demand from fare classes $\{j+1, \ldots, n\}$. To state this mathematically, we let

$$\mathcal{Y}_j^* = \operatorname*{argmax}_{0 \le y_j \le c} f_j(y_j). \tag{4.5}$$

Therefore, we can use any element of $\mathcal{Y}_j^*$ as the optimal protection level when making the decisions for fare class j. The fact that the optimal protection levels are nested implies that $\min_{y_j \in \mathcal{Y}_j^*} y_j \geq \min_{y_{j+1} \in \mathcal{Y}_{j+1}^*} y_{j+1}$ and $\max_{y_j \in \mathcal{Y}_j^*} y_j \geq \max_{y_{j+1} \in \mathcal{Y}_{j+1}^*} y_{j+1}$ for all $j = 1, \ldots, n-1$. In this case, we can choose $y_1^* \in \mathcal{Y}_1^*,\ y_2^* \in \mathcal{Y}_2^*, \ldots, y_n^* \in \mathcal{Y}_n^*$ such that $y_1^* \geq y_2^* \geq \ldots \geq y_n^*$.

Our dynamic programming formulation differs from the existing literature in two aspects. First, we index the fare classes such that fare classes 1 and n respectively correspond to the cheapest and most expensive fare classes, whereas the existing literature usually indexes the fare classes in the reverse order. The motivation for our choice is that it is common to refer to a cheaper fare class as a lower fare class and it is more consistent to index a cheaper fare class with a smaller integer. Second, we use a two-dimensional state variable in (4.2) and (4.3), whereas the existing literature usually uses a one-dimensional state variable. It is possible to use a one-dimensional state variable in (4.2) and (4.3) by simply letting $\hat{v}_j(x_j) = \mathbb{E}\{v_j(x_j, D_j)\}$ and taking the expectations of both sides. However, the way our dynamic programming formulation is presented will be more useful for the subsequent development in the paper. Lastly, we note that all of the results that we mention in this section are quite standard and the details can be found in Brumelle and McGill (1993) and Talluri and van Ryzin (2004).

4.4 Stochastic Approximation Algorithm of van Ryzin and McGill

We briefly describe the stochastic approximation algorithm of van Ryzin and McGill (2000) to compute the optimal protection levels. The starting point of their algorithm is the optimality conditions given in Brumelle and McGill (1993) for the single-leg revenue management problem. Brumelle and McGill (1993) show that if the demands for the fare classes are continuous random variables, then the optimal protection levels satisfy the following fill event conditions. Given a set of protection levels $\{y_j : j = 1, \ldots, n\}$, we define the jth fill event as $E_j(y, D) = \{D_n \geq y_{n-1}, D_n + D_{n-1} \geq y_{n-2}, \ldots, D_n + \ldots + D_{j+1} \geq y_j\}$. Brumelle and McGill (1993) show that if the demands for the fare classes are continuous random variables, then the optimal protection levels $\{y_j^* : j = 1, \ldots, n\}$ satisfy

$$-r_j + r_n \mathbb{P}\{E_j(y^*, D)\} = 0$$

for $j = 1, \ldots, n-1$. The stochastic approximation algorithm of van Ryzin and McGill (2000) iteratively finds a set of protection levels that satisfy the above first order conditions. Letting $H_j(\cdot, D) = -r_j + r_n \mathbf{1}(E_j(\cdot, D))$, van Ryzin and McGill (2000) essentially update their iterates by using $H_j(\cdot, D)$ as the step direction. A difficulty with the above procedure is that the optimality condition in Brumelle and McGill (1993) does not hold if the demands are discrete random variables. van Ryzin and McGill (2000) address this issue by using a randomization scheme that smooths the underlying problem. In particular, letting θ^k be an estimate of the optimal protection levels at iteration k, the protection level for the jth fare class is taken to be $\lfloor \theta_j^k \rfloor$ with probability $\theta_j^k - \lfloor \theta_j^k \rfloor$ and $\lceil \theta_j^k \rceil$ with probability $\lceil \theta_j^k \rceil - \theta_j^k$. The stochastic approximation algorithm of van Ryzin and McGill (2000) is given below.

van Ryzin and McGill's Algorithm

Step 1. Initialize the estimates of the optimal protection levels $\{\theta_j^1 : j = 1, \ldots, n\}$ arbitrarily. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{D_j^k : j = 1, \ldots, n\}$ be the demand random variables at iteration

k and $\{u_j^k : j = 1, \ldots, n\}$ be n independent and identically distributed uniform $[0, 1]$ random variables, set

$$p_j^k = \begin{cases} \lfloor \theta_j^k \rfloor & \text{if } u_j^k \leq \theta_j^k - \lfloor \theta_j^k \rfloor \\ \lceil \theta_j^k \rceil & \text{if } u_j^k > \theta_j^k - \lfloor \theta_j^k \rfloor, \end{cases}$$

and

$$\theta_j^{k+1} = \Pi_{[0,c]}\big(\theta_j^k + \gamma^k H_j(p^k, D^k)\big),$$

for all $j = 1, \ldots, n$, where $\Pi_{[0,c]}(\cdot)$ denotes the projection operator onto the interval $[0, c]$ and γ^k is a step size parameter.

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

Under some technical conditions, van Ryzin and McGill (2000) show that the iterates $\{\theta_j^k : j = 1, \ldots, n\}$ converge to the optimal protection levels w.p.1. However, because of the randomization scheme, a policy that implements the integer protection levels $\{p_j^k : j = 1, \ldots, n\}$ in iteration k is not necessarily optimal. We also note that the iterates $\{\theta_j^k : j = 1, \ldots, n\}$ and $\{p_j^k : j = 1, \ldots, n\}$ do not necessarily have components that are non-increasing in the fare classes. In order to ensure that the fill events $E_j(p^k, D^k)$ are observable when the demands are censored by the seat inventory, van Ryzin and McGill (2000) maintain an additional set of protection levels, which they refer to as the interim protection levels. The interim protection levels are the ones that are actually implemented in practice. Letting $\{\hat{p}_j^k : j = 1, \ldots, n\}$ denote the interim protection levels, we have $\hat{p}_j^k = \max\{p_i^k : j \leq i \leq n\}$. Clearly, the interim protection levels satisfy $\hat{p}_1^k \geq \hat{p}_2^k \geq \ldots \geq \hat{p}_n^k$.

4.5 Stochastic Approximation Method

In this section, we consider computing the optimal protection levels by using our stochastic approximation method. By (4.4), we can compute a stochastic subgradient of $f_j(\cdot)$ at y_j through

$$\Delta_j(y_j, d_{j+1}) = -r_j + \dot{v}_{j+1}(y_j, d_{j+1}), \tag{4.6}$$

where we use $\dot{v}_{j+1}(y_j, d_{j+1})$ to denote a stochastic subgradient of $\mathbb{E}\{v_{j+1}(\cdot, D_{j+1})\}$ at y_j. In other words, if we use $\partial v_{j+1}(y_j, d_{j+1})$ to denote the subdifferential of $v_{j+1}(\cdot, d_{j+1})$ at y_j, then we have $\dot{v}_{j+1}(y_j, d_{j+1}) \in \partial v_{j+1}(y_j, d_{j+1})$. Interchanging the orders of all expectations and subgradients throughout the paper trivially follows from the fact that the demand distributions are discrete and the capacity is finite. In this case, letting $\{y_j^k : j = 1, \ldots, n\}$ be the estimates of the optimal protection levels at iteration k, $\{D_j^k : k = 1, \ldots, n\}$ be the demand random variables at iteration k and $\{\alpha_j^k : j = 1, \ldots, n\}_k$ be a sequence of step size parameters, we can update our estimates of the optimal protection levels by

$$y_j^{k+1} = \min\left\{\left[\, y_j^k + \alpha_j^k\, \Delta_j(y_j^k, D_{j+1}^k)\,\right]^+, c\right\}, \tag{4.7}$$

where the operator $\min\{[\cdot]^+, c\}$ ensures that the estimates of the optimal protection levels always lie in the interval $[0, c]$. If the sequence of protection levels $\{y_j^k : j = 1, \ldots, n\}_k$ is generated by (4.7), then we can use the standard results on stochastic approximation methods to show that $\{y_j^k : j = 1, \ldots, n\}_k$ converges to a set of optimal protection levels w.p.1. However, this approach is clearly not realistic because the computation in (4.6) requires the knowledge of $\{v_j(\cdot, \cdot) : j = 1, \ldots, n\}$. The stochastic approximation method that we propose in this section is based on constructing tractable approximations to the stochastic subgradients of $\{f_j(\cdot) :$

$j = 1, \ldots, n\}$.

Since $f_j(\cdot)$ is concave and the optimal protection level y_j^* is a maximizer of this function over the interval $[0, c]$, we can write (4.3) as

$$v_j(x_j, d_j) = \begin{cases} -r_j\,[x_j - d_j]^+ + r_j\,x_j & \\ \quad + \mathbb{E}\big\{v_{j+1}([x_j - d_j]^+, D_{j+1})\big\} & \text{if } y_j^* < [x_j - d_j]^+ \\ -r_j\,y_j^* + r_j\,x_j + \mathbb{E}\big\{v_{j+1}(y_j^*, D_{j+1})\big\} & \text{if } [x_j - d_j]^+ \leq y_j^* \leq x_j \\ \mathbb{E}\big\{v_{j+1}(x_j, D_{j+1})\big\} & \text{if } x_j < y_j^* \end{cases} \tag{4.8}$$

for $x_j \in [0, c]$. Since $y_j^* \geq 0$, we have $0 \leq y_j^* < [x_j - d_j]^+$ whenever the condition in the first case above holds. Therefore, we can replace $[x_j - d_j]^+$ in the first case by $x_j - d_j$. On the other hand, the condition in the second case is equivalent to $x_j - d_j \leq y_j^* \leq x_j$ and $0 \leq y_j^* \leq x_j$. Since $y_j^* \geq 0$, we can replace the condition in the second case by $x_j - d_j \leq y_j^* \leq x_j$. These imply that we can write (4.8) as

$$v_j(x_j, d_j) = \begin{cases} r_j\,d_j + \mathbb{E}\big\{v_{j+1}(x_j - d_j, D_{j+1})\big\} & \text{if } y_j^* < x_j - d_j \\ r_j\,[x_j - y_j^*] + \mathbb{E}\big\{v_{j+1}(y_j^*, D_{j+1})\big\} & \text{if } x_j - d_j \leq y_j^* \leq x_j \\ \mathbb{E}\big\{v_{j+1}(x_j, D_{j+1})\big\} & \text{if } x_j < y_j^*. \end{cases} \tag{4.9}$$

Therefore, it is easy to see that we can compute a stochastic subgradient of $\mathbb{E}\big\{v_j(\cdot, D_j)\big\}$ at x_j through the recursion

$$\dot{v}_j(x_j, d_j) = \begin{cases} \mathbb{E}\big\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\big\} & \text{if } y_j^* < x_j - d_j \\ r_j & \text{if } x_j - d_j \leq y_j^* \leq x_j \\ \mathbb{E}\big\{\dot{v}_{j+1}(x_j, D_{j+1})\big\} & \text{if } x_j < y_j^*. \end{cases} \tag{4.10}$$

In the appendix, we formally show that (4.10) indeed gives a stochastic subgradient of $\mathbb{E}\big\{v_j(\cdot, D_j)\big\}$.

To construct tractable approximations to the stochastic subgradients of $\{f_j(\cdot) : j = 1, \ldots, n\}$, we *mimic* the computation in (4.10) by using the estimates of the optimal protection levels. In particular, letting $\{y_j^k : j = 1, \ldots, n\}$ be the estimates of the optimal protection levels at iteration k and using $\mathcal{O}(\cdot)$ to denote the operator that rounds a scalar to a nearest integer, we recursively define

$$\rho_j^k(x_j, d_j, d_{j+1}, \ldots, d_n) = \begin{cases} \rho_{j+1}^k(x_j - d_j, d_{j+1}, \ldots, d_n) & \text{if } \mathcal{O}(y_j^k) < x_j - d_j \\ r_j & \text{if } x_j - d_j \leq \mathcal{O}(y_j^k) \leq x_j \\ \rho_{j+1}^k(x_j, d_{j+1}, \ldots, d_n) & \text{if } x_j < \mathcal{O}(y_j^k), \end{cases} \tag{4.11}$$

with $\rho_{n+1}^k(\cdot, \cdot, \ldots, \cdot) = 0$. We propose using $\rho_j^k(x_j, d_j, d_{j+1}, \ldots, d_n)$ to approximate $\dot{v}_j(x_j, d_j)$. More specifically, at iteration k, we replace $\dot{v}_{j+1}(y_j, d_{j+1})$ in (4.6) with $\rho_{j+1}^k(y_j, d_{j+1}, \ldots, d_n)$ and use

$$s_j^k(y_j, d_{j+1}, \ldots, d_n) = -r_j + \rho_{j+1}^k(y_j, d_{j+1}, \ldots, d_n) \tag{4.12}$$

to approximate a stochastic subgradient of $f_j(\cdot)$ at y_j. Therefore, we propose the following algorithm to compute the optimal protection levels.

Algorithm 1

Step 1. Initialize the estimates of the optimal protection levels $\{y_j^1 : j = 1, \ldots, n\}$ such that $c \geq y_1^1 \geq y_2^1 \geq \ldots \geq y_n^1 = 0$. Initialize the iteration counter by setting $k = 1$.

Step 2. Letting $\{D_j^k : j = 1, \ldots, n\}$ be the demand random variables at iteration k, set

$$y_j^{k+1} = \max\big\{ \min\big\{ \big[y_j^k + \alpha_j^k \, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \big]^+, c \big\}, \mathcal{O}(y_{j+1}^{k+1}) \big\} \tag{4.13}$$

for all $j = 1, \ldots, n$.

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

We let $\mathcal{F}^k$ be the filtration generated by $\{\{y_1^1, \ldots, y_n^1\}, \{D_1^1, \ldots, D_n^1\}, \ldots, \{D_1^{k-1}, \ldots, D_n^{k-1}\}\}$. Given $\mathcal{F}^k$, we assume that the conditional distribution of $\{D_j^k : j = 1, \ldots, n\}$ is the same as the distribution of $\{D_j : j = 1, \ldots, n\}$. We assume that the step size parameters $\{\alpha_j^k : j = 1, \ldots, n\}$ are positive and $\mathcal{F}^k$-measurable, in which case the estimates of the optimal protection levels $\{y_j^k : j = 1, \ldots, n\}$ are also $\mathcal{F}^k$-measurable. In the next section, we show that if the sequence $\{y_j^k : j = 1, \ldots, n\}_k$ is generated by Algorithm 1, then it converges to a set of optimal protection levels w.p.1.

Several remarks are in order for our approximation to $\dot{v}_j(x_j, d_j)$ and Algorithm 1. First, we need the realizations of the demand random variables $\{D_j, D_{j+1}, \ldots, D_n\}$ to compute $\rho_j^k(x_j, d_j, d_{j+1}, \ldots, d_n)$, whereas we only need the realization of the demand random variable D_j to compute $\dot{v}_j(x_j, d_j)$. Also, the computation of $\rho_j^k(x_j, d_j, d_{j+1}, \ldots, d_n)$ does not require computing expectations. Second, comparing (4.6) and (4.12) indicates that if $\dot{v}_j(\cdot, \cdot)$ and $\mathbb{E}\{\rho_j^k(\cdot, \cdot, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k\}$ are *close* to each other for all $j = 1, \ldots, n$, then the expected step directions $\mathbb{E}\{\Delta_j(\cdot, D_{j+1}^k) \,|\, \mathcal{F}^k\}$ and $\mathbb{E}\{s_j^k(\cdot, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k\}$ are *close* to each other. In this case, using the step direction $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ instead of $\Delta_j(y_j^k, D_{j+1}^k)$ does not bring too much error in expectation. Our convergence proof is heavily based on this observation, which we make mathematically precise in Lemma 28 below. Third, we round the estimates of the optimal protection levels to the nearest integers when computing $\rho_j^k(x_j, d_j, d_{j+1}, \ldots, d_n)$. This becomes useful when we use Algorithm 1 in a real time setting where we use the protection levels

$\{\mathcal{O}(y_j^k) : j = 1, \ldots, n\}$ to satisfy the demands from the fare classes and we update our estimates of the optimal protection levels after observing the demands. In such situations, it is important to use integer protection levels since we cannot sell a fraction of a seat. Fourth, the way that we update our estimates of the optimal protection levels in (4.13) ensures that we have $c \geq \mathcal{O}(y_1^k) \geq \mathcal{O}(y_2^k) \geq \ldots \geq \mathcal{O}(y_n^k) \geq 0$ for all $k = 1, 2, \ldots$. Therefore, the estimates of the optimal protection levels at each iteration are nested. It is also important to note that the update in (4.13) is a Gauss-Seidel variant. More specifically, we need the value of y_{j+1}^{k+1} to compute the value of y_j^{k+1}. Therefore, Step 2 in Algorithm 1 has to be carried out starting from fare class n and moving backwards through the fare classes.

The step directions in Algorithm 1 are motivated by the dynamic programming formulation of the problem. On the other hand, the step directions in the method proposed by van Ryzin and McGill (2000) are motivated by the fill event optimality conditions in Brumelle and McGill (1993). As we show in Lemma 28 below, our step directions are closely related to the stochastic subgradients of the value functions, but this is not the case for the step directions used by van Ryzin and McGill (2000). We also note that we work with discrete demand distributions. To deal with discrete demand distributions, van Ryzin and McGill (2000) propose the randomized version of their method described in Section 4.4, which randomly chooses between the protection levels $\mathcal{O}(y_j^k)$ and $\mathcal{O}(y_j^k) \mp 1$ at each iteration. The iterates of the randomized version converge to a set of optimal protection levels, but the randomization between the protection levels $\mathcal{O}(y_j^k)$ and $\mathcal{O}(y_j^k) \mp 1$ results in suboptimality for the performance of the policy. Finally, we emphasize that the estimates of the optimal protection levels that we obtain at each iteration are nested, whereas such a condition is not imposed by van Ryzin and McGill (2000).

4.6 Convergence Proof

In this section, we show that the iterates of Algorithm 1 converge to a set of optimal protection levels w.p.1. We begin by some preliminary results in Section 4.6.1 and complete the proof in Section 4.6.2

4.6.1 Preliminaries

The next lemma establishes a uniform bound on our step directions.

Lemma 27 *There exists a finite scalar M such that we have*

$$|\rho_j^k(x_j, D_j^k, D_{j+1}^k, \ldots, D_n^k)| \leq M \text{ and } |s_j^k(x_j, D_{j+1}^k, \ldots, D_n^k)| \leq M$$

w.p.1 for all $x_j \in [0, c]$, $j = 1, \ldots, n$, $k = 1, 2, \ldots$.

Proof If we let $R = \max_{j \in \{1,\ldots,n\}} r_j$, then by using (4.11) and moving backwards through the fare classes, it is easy to see that $|\rho_j^k(x_j, D_j^k, D_{j+1}^k, \ldots, D_n^k)| \leq R$. By (4.12), the result follows by letting $M = 2R$. □

The next lemma shows that if $\{y_j^k : j = 1, \ldots, n\}$ get close to the optimal protection levels, then the step directions in Algorithm 1 are related to the stochastic subgradients of the value functions. In Lemma 28 and throughout the rest of the paper, since $\{y_j^k : j = 1, \ldots, n\}$ are $\mathcal{F}^k$-measurable, we treat $\{y_j^k : j = 1, \ldots, n\}$ as known constants when dealing with a conditional expectation of the form $\mathbb{E}\{\cdot \mid \mathcal{F}^k\}$. Also, since $\{v_j(\cdot, D_j) : j = 1, \ldots, n\}$ are piecewise-linear concave functions with points of nondifferentiability being a subset of integers, it is easy to see that $\mathcal{Y}_j^*$ in

(4.5) is a closed interval with integer end points. We let $\mathcal{Y}_j^* = [L_j^*, U_j^*]$ throughout the rest of the paper, where L_j^* and U_j^* are integers.

Lemma 28 *Assume that the sequence $\{y_j^k : j = 1, \ldots, n\}_k$ is generated by Algorithm 1. If it holds that*

$$y_j^k \in (L_j^* - 1/2, U_j^* + 1/2), y_{j+1}^k \in (L_{j+1}^* - 1/2, U_{j+1}^* + 1/2), \ldots, \\ y_n^k \in (L_n^* - 1/2, U_n^* + 1/2),$$

then we have $\mathbb{E}\{\rho_j^k(x_j, D_j^k, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k,\, D_j^k\} \in \partial v_j(x_j, D_j^k)$ *w.p.1 for all* $x_j \in [0, c]$.

Proof We show the result by induction over the fare classes. It is easy to show the result for fare class n. Assuming that the result holds for fare class $j+1$, we now show that the result holds for fare class j. The assumption in the lemma implies that we can find $y_j^* \in \mathcal{Y}_j^*,\ y_{j+1}^* \in \mathcal{Y}_{j+1}^*, \ldots, y_n^* \in \mathcal{Y}_n^*$ such that $y_j^* = \mathcal{O}(y_j^k),\ y_{j+1}^* = \mathcal{O}(y_{j+1}^k), \ldots, y_n^* = \mathcal{O}(y_n^k)$. Taking the conditional expectations in (4.11) and recalling that we use $\dot{v}_{j+1}(x_j, d_{j+1})$ to denote an element of $\partial v_{j+1}(x_j, d_{j+1})$, we obtain

$$\begin{aligned}
&\mathbb{E}\{\rho_j^k(x_j, d_j, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k\} \\
&\quad = \mathbb{E}\{\mathbb{E}\{\rho_j^k(x_j, d_j, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k, D_{j+1}^k\} \,|\, \mathcal{F}^k\} \\
&\quad = \begin{cases} \mathbb{E}\{\mathbb{E}\{\rho_{j+1}^k(x_j - d_j, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k, D_{j+1}^k\} \,|\, \mathcal{F}^k\} & \text{if } y_j^* < x_j - d_j \\ r_j & \text{if } x_j - d_j \le y_j^* \le x_j \\ \mathbb{E}\{\mathbb{E}\{\rho_{j+1}^k(x_j, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k, D_{j+1}^k\} \,|\, \mathcal{F}^k\} & \text{if } x_j < y_j^* \end{cases}
\end{aligned}$$

$$= \begin{cases} \mathbb{E}\big\{\dot{v}_{j+1}(x_j - d_j, D^k_{j+1}) \,|\, \mathcal{F}^k\big\} & \text{if } y^*_j < x_j - d_j \\ r_j & \text{if } x_j - d_j \le y^*_j \le x_j \\ \mathbb{E}\big\{\dot{v}_{j+1}(x_j, D^k_{j+1}) \,|\, \mathcal{F}^k\big\} & \text{if } x_j < y^*_j, \end{cases} \tag{4.14}$$

where we use the induction assumption that $\mathbb{E}\big\{\rho^k_{j+1}(\cdot, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k, D^k_{j+1}\big\} \in \partial v_{j+1}(\cdot, D^k_{j+1})$ in the third inequality. Comparing (4.14) with (4.10) and noting that the distribution of D^k_{j+1} conditional on $\mathcal{F}^k$ is the same as the distribution of D_{j+1}, we obtain $\mathbb{E}\big\{\rho^k_j(x_j, d_j, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k\big\} \in \partial v_j(x_j, d_j)$. □

Roughly speaking, the next lemma shows that if the estimates of the optimal protection levels at iteration k are close to the optimal protection levels, then the estimates of the optimal protection levels at iteration $k+1$ are also close to the optimal protection levels.

Lemma 29 *Assume that the sequence $\{y^k_j : j = 1, \ldots, n\}_k$ is generated by Algorithm* 1*. If it holds that*

$$y^k_j \in (L^*_j - 1/4, U^*_j + 1/4), y^k_{j+1} \in (L^*_{j+1} - 1/4, U^*_{j+1} + 1/4), \ldots, y^k_n \in (L^*_n - 1/4, U^*_n + 1/4) \text{ and } \alpha^k_j \in [0, 1/(4M)], \alpha^k_{j+1} \in [0, 1/(4M)], \ldots, \alpha^k_n \in [0, 1/(4M)],$$

*then we have $\mathcal{O}(y^{k+1}_j) \in \mathcal{Y}^*_j$ w.p.*1*.*

Proof All statements in the proof are in w.p.1 sense. We show the result by induction over the fare classes. Since $r_n > 0$, we have $\mathcal{Y}^*_n = \{0\}$ by (4.5) and $s^k_n(\cdot) < 0$ by (4.12). Therefore, we have $y^k_n = 0$ by (4.13) for all $k = 1, 2, \ldots$ and the result holds for fare class n. Assuming that the result holds for fare class $j+1$, we now show that the result holds for fare class j. By the assumption in the lemma and Lemma 27, we have

$$L_j^* - 1/2 < y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) < U_j^* + 1/2. \tag{4.15}$$

We consider three cases.

Case 1. Assume that $y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \geq c \geq \mathcal{O}(y_{j+1}^{k+1})$. Since we have $\mathcal{O}(y_{j+1}^{k+1}) \in \mathcal{Y}_{j+1}^* \subset [0, c]$ by the induction assumption, we obtain $y_j^{k+1} = c$ by (4.13). On the other hand, we have $U_j^* + 1/2 > c$ by (4.15). Since U_j^* is an integer smaller than c, we obtain $U_j^* = c$. Therefore, we have $\mathcal{O}(y_j^{k+1}) = c \in [c, c] \subset [L_j^*, U_j^*]$.

Case 2. Assume that $c > y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \geq \mathcal{O}(y_{j+1}^{k+1})$. Since Algorithm 1 ensures that $y_j^k \geq 0$ for all $j = 1, \ldots, n$, $k = 1, 2, \ldots$, we have $\mathcal{O}(y_{j+1}^{k+1}) \geq 0$. Therefore, by (4.13), we have $y_j^{k+1} = y_j^k +$
$\alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. We obtain $\mathcal{O}(y_j^{k+1}) \in [L_j^*, U_j^*]$ by (4.15).

Case 3. Assume that $c \geq \mathcal{O}(y_{j+1}^{k+1}) > y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. We have $y_j^{k+1} = \mathcal{O}(y_{j+1}^{k+1})$ by (4.13) and $\mathcal{O}(y_{j+1}^{k+1}) > L_j^* - 1/2$ by (4.15). Since we have $\mathcal{O}(y_{j+1}^{k+1}) \in \mathcal{Y}_{j+1}^*$ by the induction assumption, the fact that the optimal protection levels are nested implies that $\mathcal{O}(y_{j+1}^{k+1}) \leq U_{j+1}^* \leq U_j^*$. Therefore, we obtain $L_j^* - 1/2 < y_j^{k+1} = \mathcal{O}(y_{j+1}^{k+1}) \leq U_j^*$, which implies that $\mathcal{O}(y_j^{k+1}) \in [L_j^*, U_j^*]$. □

In Section 4.6.2, we give a convergence result for Algorithm 1 that shows that the distance between y_j^k and the optimal protection level that is closest to y_j^k converges to zero w.p.1 as the iterations progress. For fare class j, we define the optimal protection level that is closest to y_j^k as

$$\mathcal{C}_j(y_j^k) = \text{argmin}_{y_j^* \in \mathcal{Y}_j^*} |y_j^* - y_j^k|. \tag{4.16}$$

The next lemma shows a contraction type of result for Algorithm 1.

Lemma 30 *Assume that the sequence* $\{y_j^k : j = 1, \ldots, n\}_k$ *is generated by Algorithm* 1. *If it holds that*

$$y_{j+1}^k \in (L_{j+1}^* - 1/4, U_{j+1}^* + 1/4), y_{j+2}^k \in (L_{j+2}^* - 1/4, U_{j+2}^* + 1/4), \ldots,$$
$$y_n^k \in (L_n^* - 1/4, U_n^* + 1/4) \text{ and } \alpha_{j+1}^k \in [0, 1/(4M)], \alpha_{j+2}^k \in [0, 1/(4M)], \ldots,$$
$$\alpha_n^k \in [0, 1/(4M)],$$

then we have $|y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})| \leq |y_j^k + \alpha_j^k\, s_j^k(y_k^k, D_{j+1}^k, \ldots, D_n^k) - \mathcal{C}_j(y_j^k)|$ *w.p.1.*

Proof All statements in the proof are in w.p.1 sense. We consider the same three cases in the proof of Lemma 29.

Case 1. Assume that $y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \geq c \geq \mathcal{O}(y_{j+1}^{k+1})$. By the same argument in Lemma 29, we have $y_j^{k+1} = c$. Since $\mathcal{C}_j(y_j^{k+1})$ is the closest optimal protection level to y_j^{k+1} and $U_j^* \leq c$, we obtain $\mathcal{C}_j(y_j^{k+1}) = U_j^*$. Using the fact that $\mathcal{C}_j(y_j^k) \leq U_j^* \leq c$, we have $y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \geq c = y_j^{k+1} \geq \mathcal{C}_j(y_j^{k+1}) = U_j^* \geq \mathcal{C}_j(y_j^k)$ and the result follows.

Case 2. Assume that $c > y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \geq \mathcal{O}(y_{j+1}^{k+1})$. By the same argument in Lemma 29, we have $y_j^{k+1} = y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. Therefore, we have $|y_j^k + \alpha_j^k\, s_j^k(y_k^k, D_{j+1}^k, \ldots, D_n^k) - \mathcal{C}_j(y_j^k)| = |y_j^{k+1} - \mathcal{C}_j(y_j^k)| \geq |y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})|$, where the last inequality follows by (4.16).

Case 3. Assume that $c \geq \mathcal{O}(y_{j+1}^{k+1}) > y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. By the same argument in Lemma 29, we have $y_j^{k+1} = \mathcal{O}(y_{j+1}^{k+1})$. If $\mathcal{O}(y_{j+1}^{k+1}) \in \mathcal{Y}_j^*$, then $|y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})| = 0$ and the result follows. We now assume that either $\mathcal{O}(y_{j+1}^{k+1}) > U_j^*$ or $\mathcal{O}(y_{j+1}^{k+1}) < L_j^*$. We immediately eliminate the former possibility, since we have $\mathcal{O}(y_{j+1}^{k+1}) \in \mathcal{Y}_{j+1}^*$ by the assumption in the lemma and Lemma 29, which, together with the fact that the optimal protection levels are nested, implies that $\mathcal{O}(y_{j+1}^{k+1}) \leq U_{j+1}^* \leq U_j^*$. Therefore, we have $y_j^k + \alpha_j^k \, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) < \mathcal{O}(y_{j+1}^{k+1}) = y_j^{k+1} < L_j^*$ and the result follows. $\square$

4.6.2 Convergence of Algorithm 1

We have the following convergence result for Algorithm 1.

Proposition 31 *Assume that the sequence $\{y_j^k : j = 1, \ldots, n\}_k$ is generated by Algorithm 1. If the sequence of step size parameters $\{\alpha_j^k : j = 1, \ldots, n\}_k$ is positive and satisfies $\sum_{k=1}^{\infty} \alpha_j^k = \infty$ and $\sum_{k=1}^{\infty} [\alpha_j^k]^2 < \infty$ w.p.1 for all $j = 1, \ldots, n$, then we have $\lim_{k \to \infty} |y_j^k - \mathcal{C}_j(y_j^k)| = 0$ w.p.1 for all $j = 1, \ldots, n$.*

Proof All statements in the proof are in w.p.1 sense. We show the result by induction over the fare classes. Since we have $\mathcal{Y}_n^* = \{0\}$ and $y_n^k = 0$ for all $k = 1, 2, \ldots$ by the argument in the proof of Lemma 29, the result holds for fare class n. Assuming that the result holds for fare classes $j+1, j+2, \ldots, n$, we now show that the result holds for fare class j. The proof is in three parts. The first part shows that an inequality of the form $\mathbb{E}\{Y^{k+1} \,|\, \mathcal{F}^k\} \leq Y^k - X^k + Z^k$ holds for appropriately defined sequences $\{X^k\}_k$, $\{Y^k\}_k$ and $\{Z^k\}_k$. The second part shows that $\{X^k\}_k$, $\{Y^k\}_k$ and $\{Z^k\}_k$ are positive and $\mathcal{F}^k$-measurable, and $\{Z^k\}_k$ satisfies $\sum_{k=1}^{\infty} Z^k < \infty$. In this case, we can conclude by the supermartingale convergence theorem that the sequence $\{Y^k\}_k$ converges and we have $\sum_{k=1}^{\infty} X^k < \infty$; see Neveu (1975). The third part uses these results to complete the proof.

Part 1. To capture the cases where the assumption of Lemma 30 holds, we define the event A_j^k as

$$A_j^k = \big\{y_{j+1}^k \in (L_{j+1}^* - 1/4, U_{j+1}^* + 1/4), y_{j+2}^k \in (L_{j+2}^* - 1/4, U_{j+2}^* + 1/4), \ldots, \\ y_n^k \in (L_n^* - 1/4, U_n^* + 1/4) \text{ and } \alpha_{j+1}^k \in [0, 1/(4M)], \alpha_{j+2}^k \in [0, 1/(4M)], \\ \ldots, \alpha_n^k \in [0, 1/(4M)]\big\}.$$

Using $\mathbf{1}(\cdot)$ to denote the indicator function, Lemma 30 and the fact that $|y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})| \leq c$ imply that

$$\begin{aligned} &|y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})|^2 \\ &\quad \leq \mathbf{1}(A_j^k)\, |y_j^k + \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) - \mathcal{C}_j(y_j^k)|^2 + [1 - \mathbf{1}(A_j^k)]\, c^2 \\ &\quad \leq |y_j^k - \mathcal{C}_j(y_j^k)|^2 - \mathbf{1}(A_j^k)\, 2\, \alpha_j^k\, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)\, [\mathcal{C}_j(y_j^k) - y_j^k] \, + [\alpha_j^k]^2\, M^2 \\ &\qquad + [1 - \mathbf{1}(A_j^k)]\, c^2, \end{aligned}$$

where the last inequality is by Lemma 27. Taking conditional expectations and noting that $\mathbf{1}(A_j^k)$ is $\mathcal{F}^k$-measurable, we obtain

$$\begin{aligned} \mathbb{E}\big\{|y_j^{k+1} - \mathcal{C}_j(y_j^{k+1})|^2 \,|\, \mathcal{F}^k\big\} \leq |y_j^k - \mathcal{C}_j(y_j^k)|^2 \\ - \mathbf{1}(A_j^k)\, 2\, \alpha_j^k\, [\mathcal{C}_j(y_j^k) - y_j^k]\, \mathbb{E}\big\{s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k\big\} \\ + [\alpha_j^k]^2\, M^2 + [1 - \mathbf{1}(A_j^k)]\, c^2. \end{aligned}$$

If we let $Y^k = |y_j^k - \mathcal{C}_j(y_j^k)|^2$, $X^k = \mathbf{1}(A_j^k)\, 2\, \alpha_j^k\, [\mathcal{C}_j(y_j^k) - y_j^k]\, \mathbb{E}\big\{s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \,|\, \mathcal{F}^k\big\}$ and $Z^k = [\alpha_j^k]^2\, M^2 + [1 - \mathbf{1}(A_j^k)]\, c^2$, then the inequality above is of the form $\mathbb{E}\big\{Y^{k+1} \,|\, \mathcal{F}^k\big\} \leq Y^k - X^k + Z^k$.

Part 2. Clearly, $\{Y^k\}_k$ and $\{Z^k\}_k$ are positive and $\{X^k\}_k$, $\{Y^k\}_k$ and $\{Z^k\}_k$ are $\mathcal{F}^k$-measurable. We now show that $\{X^k\}_k$ is positive. If $\mathbf{1}(A_j^k) = 0$, then we have $X^k = 0$. If, on the other hand, we have $\mathbf{1}(A_j^k) = 1$, then we obtain

$\mathbb{E}\{\rho^k_{j+1}(y^k_{j+1}, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k, D^k_{j+1}\} \in \partial v_{j+1}(y^k_j, D^k_{j+1})$ by Lemma 28 and the definition of the event A^k_j. Therefore, by (4.4) and (4.12), if $\mathbf{1}(A^k_j) = 1$, then we have $\mathbb{E}\{s^k_j(y^k_j, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k\} = -r_j + \mathbb{E}\{\mathbb{E}\{\rho^k_{j+1}(y^k_{j+1}, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k, D^k_{j+1}\} \,|\, \mathcal{F}^k\} = -r_j + \mathbb{E}\{\dot{v}_{j+1}(y^k_j, D^k_{j+1}) \,\,|\, \mathcal{F}^k\}$ and $\mathbb{E}\{s^k_j(y^k_j, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k\}$ is a subgradient of $f_j(\cdot)$ at y^k_j. In this case, we have $[\mathcal{C}_j(y^k_j) - y^k_j]\, \mathbb{E}\{s^k_j(y^k_j, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k\} \geq f_j(\mathcal{C}_j(y^k_j)) - f_j(y^k_j) \geq 0$, where the last inequality follows from the fact that $\mathcal{C}_j(y^k_j) \in \mathcal{Y}^*_j$ and (4.5). Therefore, $\{X^k\}_k$ is positive.

We now show that $\sum_{k=1}^{\infty} Z^k < \infty$. Noting the induction assumption that $\lim_{k\to\infty} |y^k_{j+1} - \mathcal{C}_{j+1}(y^k_{j+1})| = 0,\ \lim_{k\to\infty} |y^k_{j+2} - \mathcal{C}_{j+2}(y^k_{j+2})| = 0, \ldots, \lim_{k\to\infty} |y^k_n - \mathcal{C}_n(y^k_n)| = 0$ and the fact that $\lim_{k\to\infty} \alpha^k_j = 0$ for all $j = 1, \ldots, n$, there exists a finite iteration counter K such that $\mathbf{1}(A^k_j) = 1$ for all $k = K, K+1, \ldots$. Therefore, we have $\sum_{k=1}^{\infty} Z^k \leq \sum_{k=1}^{\infty} [\alpha^k_j]^2 M^2 + K c^2 < \infty$.

Part 3. By the supermartingale convergence theorem and Parts 1 and 2, we conclude that the sequence $\{|y^k_j - \mathcal{C}_j(y^k_j)|\}_k$ converges and we have $\sum_{k=1}^{\infty} X^k < \infty$. Noting the discussion in Part 2, we have $\mathbf{1}(A^k_j) = 1$ and $X^k = 2\,\alpha^k_j\,[\mathcal{C}_j(y^k_j) - y^k_j]\, \mathbb{E}\{s^k_j(y^k_j, D^k_{j+1}, \ldots, D^k_n) \,|\, \mathcal{F}^k\} \geq 2\,\alpha^k_j |f_j(\mathcal{C}_j(y^k_j)) - f_j(y^k_j)|$ for all $k = K, K+1, \ldots$. Therefore, we have $\sum_{k=K}^{\infty} \alpha^k_j\, |f_j(\mathcal{C}_j(y^k_j)) - f_j(y^k_j)| \leq \sum_{k=1}^{\infty} X^k < \infty$, which, together with the fact that $\sum_{k=1}^{\infty} \alpha^k_j = \infty$, implies that $\liminf_{k\to\infty} |f_j(\mathcal{C}_j(y^k_j)) - f_j(y^k_j)| = 0$. Consequently, there exists a subsequence $\{\hat{y}^k_j\}_k$ of $\{y^k_j\}_k$ such that $\lim_{k\to\infty} |f_j(\mathcal{C}_j(\hat{y}^k_j)) - f_j(\hat{y}^k_j)| = 0$. Since the sequence $\{\hat{y}^k_j\}_k$ takes values in the bounded interval $[0, c]$, we can take a further subsequence $\{\tilde{y}^k_j\}_k$ of $\{\hat{y}^k_j\}_k$ such that $\lim_{k\to\infty} \tilde{y}^k_j = \tilde{y}_j$ for some $\tilde{y}_j \in [0, c]$.

Noting the definition of $\mathcal{C}_j(\cdot)$ and letting $F^*_j = \max_{0 \leq y_j \leq c} f_j(y_j)$, we clearly have $f_j(\mathcal{C}_j(\tilde{y}^k_j)) = F^*_j$ for all $k = 1, 2, \ldots$. Therefore, by the fact that $\lim_{k\to\infty} |f_j(\mathcal{C}_j(\tilde{y}^k_j)) -$

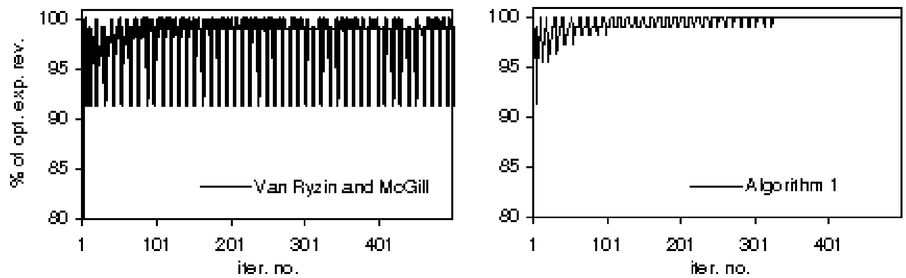

Figure 4.2: Comparison of the randomized version of the method proposed by van Ryzin and McGill (2000) and Algorithm 1. The problem parameters are $c = 15$, $r_1 = 1050$, $r_2 = 100$. The demands from the two fare classes are deterministic and we have $D_1 = 10$ and $D_2 = 10$ w.p.1.

$f_j(\tilde{y}_j^k)| = 0$, we have $\lim_{k\to\infty} f_j(\tilde{y}_j^k) = F_j^*$. On the other hand, by the continuity of $f_j(\cdot)$ and the fact that $\lim_{k\to\infty} \tilde{y}_j^k = \tilde{y}_j$, we have $\lim_{k\to\infty} f_j(\tilde{y}_j^k) = f_j(\tilde{y}_j)$. From the last two statements, we obtain $f_j(\tilde{y}_j) = F_j^*$ and $\tilde{y}_j \in \mathcal{Y}_j^*$, which imply that $|\tilde{y}_j^k - \mathcal{C}_j^k(\tilde{y}_j^k)| \leq |\tilde{y}_j^k - \tilde{y}_j|$ for all $k = 1, 2, \ldots$. Therefore, since $\{|\tilde{y}_j^k - \tilde{y}_j|\}_k$ converges to zero, $\{|\tilde{y}_j^k - \mathcal{C}_j^k(\tilde{y}_j^k)|\}_k$ also converges to zero. Recalling that the whole sequence $\{|y_j^k - \mathcal{C}_j(y_j^k)|\}_k$ converges, we obtain $\lim_{k\to\infty} |y_j^k - \mathcal{C}_j(y_j^k)| = 0$. □

A simple corollary to Proposition 31 is that there exists a finite iteration number K w.p.1 such that we have $L_j^* - 1/2 < y_j^k < U_j^* + 1/2$ for all $j = 1, \ldots, n$, $k = K, K+1, \ldots$. Therefore, we have $\mathcal{O}(y_j^k) \in \mathcal{Y}_j^*$ for all $j = 1, \ldots, n$, $k = K, K+1, \ldots$ and the policy that uses $\{\mathcal{O}(y_j^k) : j = 1, \ldots, n\}$ as the protection levels is optimal w.p.1 after a finite number of iterations. As mentioned before, the randomized version of the method proposed by van Ryzin and McGill (2000) does not guarantee that the performance of the policy is optimal. To illustrate this on a simple example, Figure 4.2 plots the total expected revenues corresponding to the protection levels obtained by the method proposed by van Ryzin and McGill (2000) and Algorithm 1 as a function of the iteration counter. The performance of Algorithm 1 is eventually optimal, whereas the performance of the method proposed by van Ryzin and McGill (2000) fluctuates.

4.7 Censored Demands

Demand censorship refers to the situation where we can observe the number of seats sold to a fare class, but not the actual amount of demand from a fare class. In this case, our demand observations are *truncated* when the amount of demand from a fare class exceeds the number of seats that we make available for sale to a fare class. We begin this section with a negative result that shows that the step direction in (4.12) cannot be computed when the demand information is censored. This implies that Algorithm 1 becomes inapplicable under censored demands. We then propose two alternative versions of Algorithm 1 that remain applicable under censored demands. The first alternative version has a somewhat weak convergence guarantee. The second alternative version is a heuristic modification of the first one, but it has somewhat more desirable practical performance.

If the demand information is censored, then we do not observe the demand random variables $\{D_j^k : j = 1, \ldots, n\}$ in Step 2 of Algorithm 1. Instead, we simulate the behavior of the policy characterized by the protection levels $\{\mathcal{O}(y_j^k) : j = 1, \ldots, n\}$ and observe the number of seats sold to different fare classes. In this case, Step 2 of Algorithm 1 has to be replaced by the following steps.

Step 2.a. Set the initial capacity x_1^k to c and set $j = 1$.

Step 2.b. Make $[x_j^k - \mathcal{O}(y_j^k)]^+$ seats available for sale to fare class j.

Step 2.c. Observe the number of seats sold to fare class j as $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$

and compute the capacity just before making the decisions for fare class $j+1$ as $x_{j+1}^k = x_j^k - \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$.

Step 2.d. If $j < n$, then increase j by 1 and go to Step 2.b.

Step 2.e. For all $j = 1, \ldots, n$, set

$$y_j^{k+1} = \max\Big\{ \min\Big\{ \big[\, y_j^k + \alpha_j^k \, s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k) \,\big]^+, c\Big\}, \mathcal{O}(y_{j+1}^{k+1})\Big\}.$$

Therefore, we only have access to $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$, but not the demand random variables themselves. Unfortunately, this information is not adequate to compute $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ for all $j = 1, \ldots, n$ and Algorithm 1 becomes inapplicable when the demand information is censored.

To illustrate, we consider a numerical example with $n = 3$, $c = 4$, $y_1^k = 3.2$, $y_2^k = 2.1$, $y_3^k = 0$, $D_1^k = 1$, $D_2^k = 1$ and $D_3^k = 2$. By Steps 2.a-2.d above, we have $x_1^k = 4$, $\min\{[x_1^k - \mathcal{O}(y_1^k)]^+, D_1^k\} = 1$, $x_2^k = 4 - 1 = 3$, $\min\{[x_2^k - \mathcal{O}(y_2^k)]^+, D_2^k\} = 1$, $x_3^k = 3 - 1 = 2$ and $\min\{[x_3^k - \mathcal{O}(y_3^k)]^+, D_3^k\} = 2$. By (4.12), computing $s_1^k(y_1^k, D_2^k, D_3^k)$ requires computing $\rho_2^k(y_1^k, D_2^k, D_3^k)$ and we have

$$\rho_2^k(y_1^k, D_2^k, D_3^k) = \begin{cases} \rho_3^k(y_1^k - D_2^k, D_3^k) & \text{if } \mathcal{O}(2.1) < 3.2 - D_2^k \\ r_2 & \text{if } 3.2 - D_2^k \le \mathcal{O}(2.1) \le 3.2 \end{cases}$$

by (4.11). Therefore, to compute $s_1^k(y_1^k, D_2^k, D_3^k)$, we need to know whether $D_2^k < 1.2$ or $D_2^k \ge 1.2$. However, if we only have access to $\{y_1^k, y_2^k, y_3^k\}$, $\{x_1^k, x_2^k, x_3^k\}$ and $\{\min\{[x_1^k - \mathcal{O}(y_1^k)]^+, D_1^k\}, \min\{[x_2^k - \mathcal{O}(y_2^k)]^+, D_2^k\}, \min\{[x_3^k - \mathcal{O}(y_3^k)]^+, D_3^k\}\}$, then we know that $1 = \min\{[x_2^k - \mathcal{O}(y_2^k)]^+, D_2^k\} = \min\{1, D_2^k\} \le D_2^k$, but not whether $D_2^k < 1.2$ or $D_2^k \ge 1.2$. In the next four sections, we describe different ways to deal with this difficulty.

4.7.1 Using Fractional Estimates of the Protection Levels

One obvious approach to deal with the censored demands is to stop rounding the estimates of the optimal protection levels. This amounts to dropping all $\mathcal{O}(\cdot)$ operators throughout the paper. In this case, it is possible to modify Proposition 31 in an obvious manner to get a convergence guarantee and it is easy to check that having access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - y_j^k]^+, D_j^k\} : j = 1, \ldots, n\}$ is adequate to compute $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ as long as $\{y_j^k : j = 1, \ldots, n\}$ are not integers. However, this approach uses fractional protection levels in Steps 2.a-2.e and it is not useful when we use Algorithm 1 in a real time setting.

4.7.2 Using a Relaxed View of Demand Censorship

It is easy to deal with the censored demands under a somewhat relaxed view of demand censorship. This relaxed view of demand censorship assumes that we have access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, D_j^k\} : j = 1, \ldots, n\}$, which amounts to assuming that we can observe the number of seats sold if we were to make $[x_j^k - \mathcal{O}(y_j^k)]^+ + 1$ seats available for sale to fare class j. In other words, the relaxed view of demand censorship assumes that we can observe whether an extra seat would have been sold to a fare class if it had been made available. The next proposition shows that we can compute $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ under the relaxed view of demand censorship. This implies that Algorithm 1 remains applicable with no modifications under the relaxed view of demand censorship.

Proposition 32 *Having access to* $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ *and* $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, D_j^k\} : j = 1, \ldots, n\}$ *is adequate to compute* $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ *for all* $j = 1, \ldots, n$.

Proof It is possible to show the result by induction over the fare classes, but we use a constructive proof that shows the computations involved more clearly. We first use induction over the fare classes to show that

$$x_j^k \geq \mathcal{O}(y_{j-1}^k) \geq \mathcal{O}(y_j^k) \tag{4.17}$$

for all $j = 2, \ldots, n$. This is easy to show for the second fare class. Assuming that the result holds for fare class j, we have $x_{j+1}^k = x_j^k - \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} = x_j^k - \min\{x_j^k - \mathcal{O}(y_j^k), D_j^k\} = \max\{\mathcal{O}(y_j^k), x_j^k - D_j^k\} \geq \mathcal{O}(y_j^k) \geq \mathcal{O}(y_{j+1}^k)$, where the last inequality uses the fact that Algorithm 1 ensures that the estimates of the optimal protection levels are nested. We now focus on computing $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. By (4.12), this requires computing $\rho_{j+1}^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$. We consider two cases.

Case 1. Assume that $\min\{[x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1, D_{j+1}^k\} = [x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1$. In this case, we deduce that $D_{j+1}^k \geq [x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1 \geq \mathcal{O}(y_j^k) - \mathcal{O}(y_{j+1}^k) + 1 \geq y_j^k - \mathcal{O}(y_{j+1}^k)$, where the second inequality follows from (4.17). This chain of inequalities and (4.13) imply that $y_j^k - D_{j+1}^k \leq \mathcal{O}(y_{j+1}^k) \leq y_j^k$ and we obtain $\rho_{j+1}^k(y_j^k, D_{j+1}^k, D_{j+2}^k, \ldots, D_n^k) = r_{j+1}$ by (4.11).

Case 2. Assume that $\min\{[x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1, D_{j+1}^k\} < [x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1$. In this case, we deduce that $D_{j+1}^k = \min\{[x_{j+1}^k - \mathcal{O}(y_{j+1}^k)]^+ + 1, D_{j+1}^k\}$. Therefore, we have access to the value of D_{j+1}^k. We consider two subcases.

Case 2.a. Assume that $D^k_{j+1} \geq y^k_j - \mathcal{O}(y^k_{j+1})$. The same argument in Case 1 implies that $y^k_j - D^k_{j+1} \leq \mathcal{O}(y^k_{j+1}) \leq y^k_j$ and we obtain $\rho^k_{j+1}(y^k_j, D^k_{j+1}, D^k_{j+2}, \ldots, D^k_n) = r_{j+1}$.

Case 2.b. Assume that $D^k_{j+1} < y^k_j - \mathcal{O}(y^k_{j+1})$. We have $\rho^k_{j+1}(y^k_j, D^k_{j+1}, D^k_{j+2}, \ldots, D^k_n)$ $= \rho_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, \ldots, D^k_n)$ by (4.11).

Therefore, if Cases 1 or 2.a holds, then we are done. Otherwise, it remains to compute $\rho^k_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, \ldots, D^k_n)$ for a known value of D^k_{j+1}. The result follows by continuing in the same fashion for the subsequent fare classes. For example, assume that Case 2.b holds, in which case it remains to compute $\rho^k_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, \ldots, D^k_n)$ for a known value of D^k_{j+1}. Since Case 2.b is a subcase of Case 2, we have $D^k_{j+1} = \min\{[x^k_{j+1} - \mathcal{O}(y^k_{j+1})]^+, D^k_{j+1}\}$ and we obtain $x^k_{j+2} = x^k_{j+1} - \min\{[x^k_{j+1} - \mathcal{O}(y^k_{j+1})]^+, D^k_{j+1}\} = x^k_{j+1} - D^k_{j+1}$. We consider two cases similar to Cases 1 and 2.

Case I. Assume that $\min\{[x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1, D^k_{j+2}\} = [x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1$. Using (4.17), we obtain $D^k_{j+2} \geq [x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1 = x^k_{j+2} - \mathcal{O}(y^k_{j+2}) + 1 = x^k_{j+1} - D^k_{j+1} - \mathcal{O}(y^k_{j+2}) + 1 \geq \mathcal{O}(y^k_j) - D^k_{j+1} - \mathcal{O}(y^k_{j+2}) + 1 \geq y^k_j - D^k_{j+1} - \mathcal{O}(y^k_{j+2})$. Since we assume that Case 2.b holds, (4.17) also implies that $D^k_{j+1} < y^k_j - \mathcal{O}(y^k_{j+1}) \leq y^k_j - \mathcal{O}(y^k_{j+2})$. Therefore, we obtain $y^k_j - D^k_{j+1} - D^k_{j+2} \leq \mathcal{O}(y^k_{j+2}) \leq y^k_j - D^k_{j+1}$ and we have $\rho^k_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, D^k_{j+3}, \ldots, D^k_n) = r_{j+2}$ by (4.11).

Case II. Assume that $\min\{[x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1, D^k_{j+2}\} < [x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1$. In this case, we deduce that $D^k_{j+2} = \min\{[x^k_{j+2} - \mathcal{O}(y^k_{j+2})]^+ + 1, D^k_{j+2}\}$ and we have access to the value of D^k_{j+2}. Since Case 2.b is a subcase of Case 2, we also have access to the value of D^k_{j+1}. We consider two subcases similar to Cases 2.a and 2.b.

Case II.a. Assume that $D^k_{j+1} + D^k_{j+2} \geq y^k_j - \mathcal{O}(y^k_{j+2})$. The same argument in Case I implies that $y^k_j - D^k_{j+1} - D^k_{j+2} \leq \mathcal{O}(y^k_{j+2}) \leq y^k_j - D^k_{j+1}$ and we obtain $\rho^k_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, D^k_{j+3}, \ldots, D^k_n) = r_{j+2}$ by (4.11).

Case II.b. Assume that $D^k_{j+1} + D^k_{j+2} < y^k_j - \mathcal{O}(y^k_{j+2})$. We have $\rho^k_{j+2}(y^k_j - D^k_{j+1}, D^k_{j+2}, D^k_{j+3}, \ldots, D^k_n) = \rho^k_{j+3}(y^k_j - D^k_{j+1} - D^k_{j+2}, D^k_{j+3}, \ldots, D^k_n)$ by (4.11).

Therefore, if Cases I or II.a holds, then we are done. Otherwise, it remains to compute $\rho^k_{j+3}(y^k_j - D^k_{j+1} - D^k_{j+2}, D^k_{j+3}, \ldots, D^k_n)$ for known values of D^k_{j+1} and D^k_{j+2}. As mentioned before, the result follows by continuing in the same fashion for the subsequent fare classes. □

Therefore, Algorithm 1 is applicable under the relaxed view of demand censorship.

4.7.3 Perturbing the Demand Random Variables

In certain practical settings, it may not be possible to adopt the relaxed view of demand censorship described in the previous section and Algorithm 1 becomes inapplicable. In this section, we develop an alternative version of Algorithm 1 that is applicable under the assumption that we only have access to $\{y^k_j : j = 1, \ldots, n\}$, $\{x^k_j : j = 1, \ldots, n\}$ and $\{\min\{[x^k_j - \mathcal{O}(y^k_j)]^+, D^k_j\} : j = 1, \ldots, n\}$. Therefore, the alternative version is applicable without the relaxed view of demand censorship. The alternative version, however, does not converge to the optimal protection levels, but we show that the distance between its iterates and the optimal protection levels is bounded by n in the limit w.p.1.

The alternative version of Algorithm 1 that we propose in this section is obtained by replacing Step 2 of Algorithm 1 by the following steps.

Algorithm 2

Step 2.a. Set the initial capacity x_1^k to c and set $j = 1$.

Step 2.b. Make $[x_j^k - \mathcal{O}(y_j^k)]^+$ seats available for sale to fare class j.

Step 2.c. Observe the number of seats sold to fare class j as $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$ and compute the capacity just before making the decisions for fare class $j+1$ as $x_{j+1}^k = x_j^k - \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$.

Step 2.d. If $j < n$, then increase j by 1 and go to Step 2.b.

Step 2.e. For all $j = 1, \ldots, n$, set

$$y_j^{k+1} = \max\Big\{ \min\Big\{ \Big[y_j^k + \alpha_j^k \, s_j^k(y_j^k, D_{j+1}^k + 1, \ldots, D_n^k + 1) \Big]^+, c \Big\}, \mathcal{O}(y_{j+1}^{k+1}) \Big\}. \tag{4.18}$$

In Steps 2.a-2.d, Algorithm 2 uses the demand random variables $\{D_j^k : j = 1, \ldots, n\}$ to simulate the behavior of the policy characterized by the protection levels $\{\mathcal{O}(y_j^k) : j = 1, \ldots, n\}$. In Step 2.e, however, it uses the demand random variables $\{D_j^k + 1 : j = 1, \ldots, n\}$ to compute the step direction. Therefore, Algorithm 2 uses *incorrect* demand random variables when updating the estimates of the optimal protection levels. For this reason, the iterates of Algorithm 2 do not necessarily converge to the optimal protection levels. Nevertheless, as the next corollary to Proposition 32 shows, having access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$ is adequate to compute $s_j^k(y_j^k, D_{j+1}^k + 1, \ldots, D_n^k + 1)$. Therefore, Algorithm 2 is applicable when the demand information is censored and we cannot adopt the relaxed view of demand censorship.

Corollary 33 *Having access to* $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ *and*

$\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$ *is adequate to compute* $s_j^k(y_j^k, D_{j+1}^k + 1, \ldots, D_n^k + 1)$ *for all* $j = 1, \ldots, n$.

Proof Replacing $\{D_j^k : j = 1, \ldots, n\}$ in Proposition 32 by $\{D_j^k + 1 : j = 1, \ldots, n\}$, we know that having access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, D_j^k + 1\} : j = 1, \ldots, n\}$ is adequate to compute $s_j^k(y_j^k, D_{j+1}^k + 1, \ldots, D_n^k + 1)$. Clearly, if we know the values of $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$, then we know the values of $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, D_j^k + 1\} : j = 1, \ldots, n\}$. Therefore, having access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$ is adequate to compute $s_j^k(y_j^k, D_{j+1}^k + 1, \ldots, D_n^k + 1)$. □

The next corollary to Proposition 31 gives a somewhat weak convergence result for Algorithm 2.

Corollary 34 *Assume that the sequence* $\{y_j^k : j = 1, \ldots, n\}_k$ *is generated by Algorithm* 2. *If the sequence of step size parameters* $\{\alpha_j^k : j = 1, \ldots, n\}_k$ *is positive and satisfies* $\sum_{k=1}^{\infty} \alpha_j^k = \infty$ *and* $\sum_{k=1}^{\infty} [\alpha_j^k]^2 < \infty$ *w.p.*1 *for all* $j = 1, \ldots, n$, *then there exists a finite iteration number* K *w.p.*1 *such that we have* $\mathcal{O}(y_j^k) \in [L_j^*, U_j^* + n - j]$ *for all* $j = 1, \ldots, n$, $k = K, K + 1, \ldots$.

Proof We sketch the main ideas of the proof here and defer the details to the appendix. As far as updating the estimates of the optimal protection levels in Step 2.e is concerned, Algorithm 2 assumes that the demand random variables are $\{D_j^k + 1 : j = 1, \ldots, n\}$. Therefore, the iterates of Algorithm 2 converge to the optimal protection levels for the problem

$$\tilde{v}_j(x_j, d_j) = \max_{0 \le u_j \le \min\{x_j, d_j+1\}} r_j\, u_j + \mathbb{E}\big\{\tilde{v}_{j+1}(x_j - u_j, D_{j+1})\big\} \tag{4.19}$$

in the sense of Proposition 31. If we use $\tilde{\mathcal{Y}}_j^*$ to denote the set of optimal protection levels when making the decisions for fare class j in the problem above, then it is possible to show that $\tilde{\mathcal{Y}}_j^* \subset [L_j^*, U_j^* + n - j]$ and the result follows. □

The result in Corollary 34 can be weak especially when the number of fare classes is large. Furthermore, the bound on the protection levels does not imply a bound on the total expected revenue obtained by the corresponding policy. Nevertheless, Brumelle and McGill (1993) and Robinson (1995) provide some evidence that the total expected revenue is robust to small deviations from the optimal protection levels. In the next section, we combine the ideas in Sections 4.7.2 and 4.7.3 to develop another alternative version of Algorithm 1. This alternative version is also applicable when the demand information is censored and has somewhat more desirable practical performance than does Algorithm 2. However, it does not have a convergence guarantee.

4.7.4 Perturbing the Demand Random Variables When Necessary

The main motivation for using the demand random variables $\{D_j^k + 1 : j = 1, \ldots, n\}$ in Algorithm 2 is that this allows us to compute the step direction in Step 2.e under censored demands. It turns out that we do not need to increase all demand random variables by one to be able to compute the step direction. In this section, we propose an alternative version of Algorithm 1, which is obtained by replacing Step 2 of Algorithm 1 by the following steps.

Algorithm 3

Step 2.a. Set the initial capacity x_1^k to c and set $j = 1$.

Step 2.b. Make $[x_j^k - \mathcal{O}(y_j^k)]^+$ seats available for sale to fare class j.

Step 2.c. Observe the number of seats sold to fare class j as $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$ and compute the capacity just before making the decisions for fare class $j+1$ as $x_{j+1}^k = x_j^k - \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$. Also, set

$$\tilde{d}_j^k = \begin{cases} D_j^k & \text{if } \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} < [x_j^k - \mathcal{O}(y_j^k)]^+ \\ D_j^k + 1 & \text{if } \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} = [x_j^k - \mathcal{O}(y_j^k)]^+. \end{cases} \tag{4.20}$$

Step 2.d. If $j < n$, then increase j by 1 and go to Step 2.b.

Step 2.e. For all $j = 1, \ldots, n$, set

$$y_j^{k+1} = \max\Big\{ \min\Big\{ \Big[y_j^k + \alpha_j^k \, s_j^k(y_j^k, \tilde{d}_{j+1}^k, \ldots, \tilde{d}_n^k) \Big]^+, c\Big\}, \mathcal{O}(y_{j+1}^{k+1})\Big\}. \tag{4.21}$$

We note that $\tilde{d}_j^k$ is equal to D_j^k when the number of seats sold to fare class j is strictly less than the number of seats made available for sale to fare class j and it is equal to $D_j^k + 1$ when the number of seats sold to fare class j is equal to the number of seats made available for sale to fare class j. Comparing (4.18) and (4.21), since $\tilde{d}_j^k$ is not always equal to $D_j^k + 1$, the hope is that the step direction used by Algorithm 3 is *closer* to $s_j^k(y_j^k, D_{j+1}^k, \ldots, D_n^k)$ than the step direction used by Algorithm 2.

Contrary to what our description of Algorithm 3 suggests, it is, in fact, not necessary to compute $\{\tilde{d}_j^k : j = 1, \ldots, n\}$ in Step 2.c explicitly. In particular, the next corollary to Proposition 32 shows that having access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$ is adequate to compute $s_j^k(y_j^k, \tilde{d}_{j+1}^k, \ldots, \tilde{d}_n^k)$. Therefore, since $\{\tilde{d}_j^k : j = 1, \ldots, n\}$ are only used in the computation of $s_j^k(y_j^k, \tilde{d}_{j+1}^k, \ldots, \tilde{d}_n^k)$ in (4.21), it is not necessary to compute $\{\tilde{d}_j^k : j = 1, \ldots, n\}$ explicitly. The proof of the next corollary is similar to that of Corollary 33 and is deferred to the appendix.

Corollary 35 *If we let* $\{\tilde{d}_j^k : j = 1, \ldots, n\}$ *be as in* (4.20)*, then having access to* $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ *and* $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$ *is adequate to compute* $s_j^k(y_j^k, \tilde{d}_{j+1}^k, \ldots, \tilde{d}_n^k)$ *for all* $j = 1, \ldots, n$.

Therefore, Corollary 35 shows that Algorithm 3 is applicable when the demand information is censored.

4.8 Numerical Experiments

In this section, we compare the performances of Algorithms 1-3 with the performance of the stochastic approximation method proposed by van Ryzin and McGill (2000). Our test problems are taken from van Ryzin and McGill (2000) and we introduce some variety by using the approach followed by Huh and Rusmevichientong (2006). Specifically, we work with test problem that involve 4, 8 and 12 fare classes. The demand from each fare class is normally distributed, and to satisfy our assumptions, we discretize the demand random variables by rounding them to the nearest integer. Table 4.1 gives the revenues associated with the fare classes, along with the means and standard deviations of the demand random variables. For the test problem with 4 fare classes, we use $c \in \{124, 164\}$, in which case the total expected demand is 25% more or 5% less than the available capacity. Similarly, for the test problems with 8 and 12 fare classes, we respectively use $c \in \{260, 344\}$ and $c \in \{409, 541\}$.

For Algorithms 1-3 and for the method proposed by van Ryzin and McGill (2000), we test three strategies to choose the initial protection levels. In particular, we use

$$y_j^1 = \frac{r_{j+1} + \ldots + r_n}{r_1 + \ldots + r_n}\, c \tag{4.22}$$

$$y_j^1 = \frac{\mu_{j+1} + \ldots + \mu_n}{\mu_1 + \ldots + \mu_n}\, c \tag{4.23}$$

$$y_j^1 = \frac{\mu_{j+1}\, r_{j+1} + \ldots + \mu_n\, r_n}{\mu_1\, r_1 + \ldots + \mu_n\, r_n}\, c, \tag{4.24}$$

where μ_j is the expected value of the demand from fare class j. These initial protection levels are heuristically motivated by the observation that if the revenue associated with a fare class is large and the demand from a fare class is likely to be large, then we should protect a large number of seats for this fare class. We refer to the sets of initial protection levels computed by (4.22), (4.23) and (4.24) respectively as Y_R^1, Y_M^1 and Y_{MR}^1. Although there is no guarantee, for our test problems, the total expected revenues obtained by Y_M^1 are larger than those obtained by Y_{MR}^1, which are, in turn, larger than those obtained by Y_R^1. However, computing Y_M^1 and Y_{MR}^1 requires some a priori information about the demand distributions.

We use the step size parameter $\alpha_j^k = (n - j + 1) / \frac{200}{r_n(10+k)}$ in Algorithms 1-3. This choice of step size parameters results in a bit more aggressive updates for the protection levels for the fare classes that arrive earlier. As evident from our

Table 4.1: Parameters of our test problems.

fare class	1	2	3	4
revenue	1050	567	527	350
mean	17.3	45.1	73.6	19.8
std. dev	5.8	15	17.4	6.6

fare class	1	2	3	4	5	6	7	8
revenue	1155	1050	623.7	579.7	567.7	527	385	350
mean	19	17.3	49.6	81	45.1	73.6	21.8	19.8
std. dev.	6.4	5.8	16.5	19.1	15	17.4	7.3	6.6

fare class	1	2	3	4	5	6	7	8	9	10	11	12
revenue	1260	1155	1050	680.4	632.4	623.7	579.7	567	527	420	385	350
mean	20.8	19	17.3	54.1	88.3	49.6	81	45.1	73.6	23.8	21.8	19.8
std. dev.	7	6.4	5.8	18	20.9	16.5	19.1	15	17.4	7.9	7.3	6.6

backward induction in the proof of Proposition 31, the protection levels for the fare classes that arrive earlier tend to converge more slowly than those for the fare classes that arrive later. This choice of step size parameters ensures that the step sizes that we use for the fare classes that arrive earlier do not get too small prematurely. For the method proposed by van Ryzin and McGill (2000), we use the step size parameter $\alpha_j^k = \frac{200}{r_n(10+k)}$, which is essentially the same step size parameter that is used by van Ryzin and McGill (2000) in their original paper. To be precise, van Ryzin and McGill (2000) use the step size parameter $\alpha_j^k = \frac{200}{10+k}$, but we scale their step size parameters and step directions by $1/r_n$ and r_n, respectively, in which case their method remains unchanged. Using a step size parameter that depends on the fare class does not affect the performance of this method in a systematic fashion.

We label our test problems by $(n, \kappa) \in \{4, 8, 12\} \times \{0.95, 1.25\}$, where n is the number of fare classes and κ is the ratio of the total expected demand to the initial capacity. Since we test three strategies to choose the initial protection levels, this gives us 18 test cases to consider. We use the randomized version of the method proposed by van Ryzin and McGill (2000) as a benchmark strategy, since this method has a convergence guarantee when the demand random variables are discrete. We refer to this method as RA. We refer to Algorithms 1, 2 and 3 respectively as A1, A2 and A3. We run each method for 100 iterations on 25 sample paths and present the average results over 25 sample paths. We use common random numbers when comparing the performances of different methods.

Figure 4.3 compares the performances of A1 and A3 for a few test cases. The two data series in the charts plot the total expected revenues corresponding to the

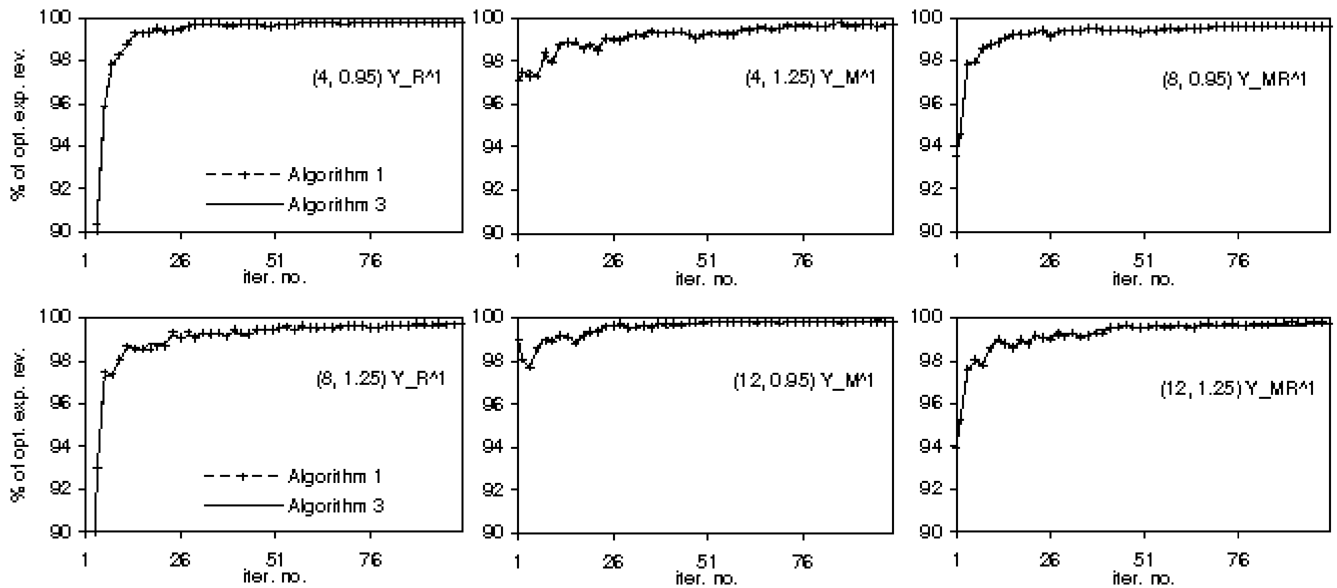

Figure 4.3: Comparison of Algorithms 1 and 3.

protection levels obtained by A1 and A3 as a function of the iteration counter. We normalize the total expected revenues by dividing by the optimal total expected revenue, which we obtain by solving the dynamic programming formulation of the problem. Figure 4.3 indicates that the performance of A3 is indistinguishable from that of A1. Since A3 is applicable when the demand information is censored, whereas A1 is applicable only when the relaxed view of demand censorship described in Section 4.7.2 is possible, we drop A1 from further consideration. We only compare RA, A2 and A3 in the subsequent discussion.

Figures 4.4, 4.5 and 4.6 respectively show the results for the test problems with 4, 8 and 12 fare classes. In these figures, the charts on the left, in the middle and on the right respectively correspond to the cases where the initial protection levels are Y_R^1, Y_M^1 and Y_{MR}^1. The top rows correspond to the cases where $\kappa = 0.95$, whereas the bottom rows correspond to the cases where $\kappa = 1.25$. The thick, dashed and thin data series in the charts respectively plot the total expected revenues corresponding to the protection levels obtained by RA, A2 and A3 as a function of the iteration counter.

In the figures, the performances of A2 and A3 are very close to each other with A2 slightly lagging from behind. The performance of A3 is almost always better than that of RA when the total expected demand exceeds the capacity. If the total expected demand is below the capacity and the initial protection levels are Y_M^1 or Y_{MR}^1, then the performance of RA is slightly better than that of A3.

We note that the protection levels Y_M^1 and Y_{MR}^1 are relatively good since the total expected revenues obtained by these protection levels are about 94-98% of the optimal total expected revenues. Therefore, the figures indicate that if the initial

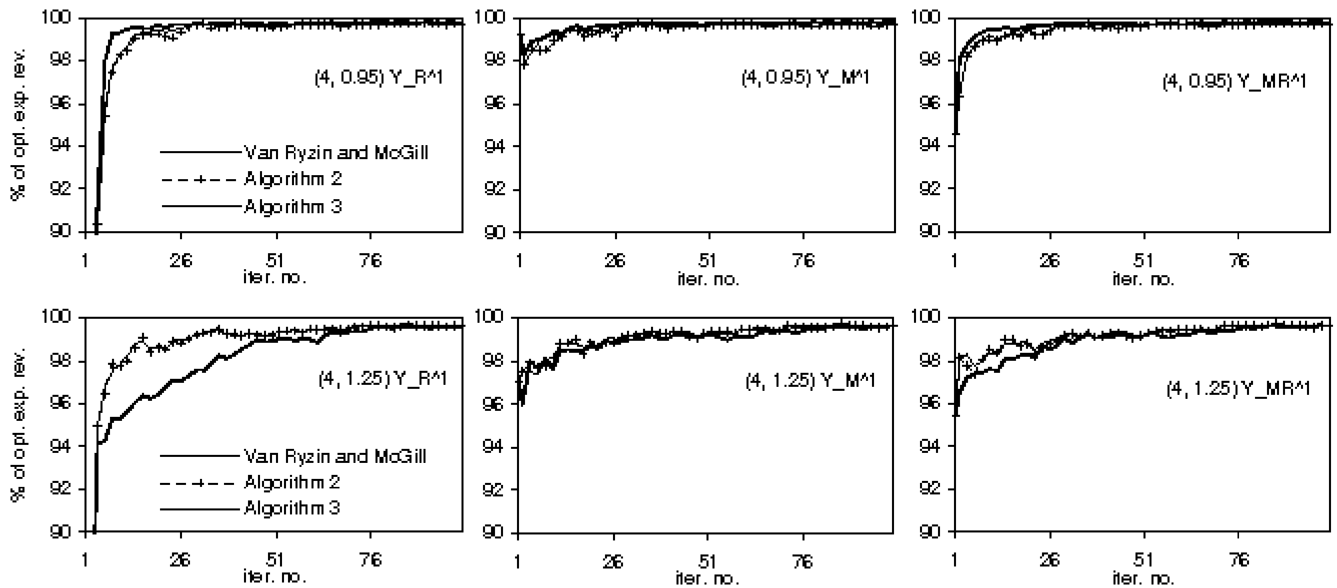

Figure 4.4: Numerical results for the test problems with 4 fare classes.

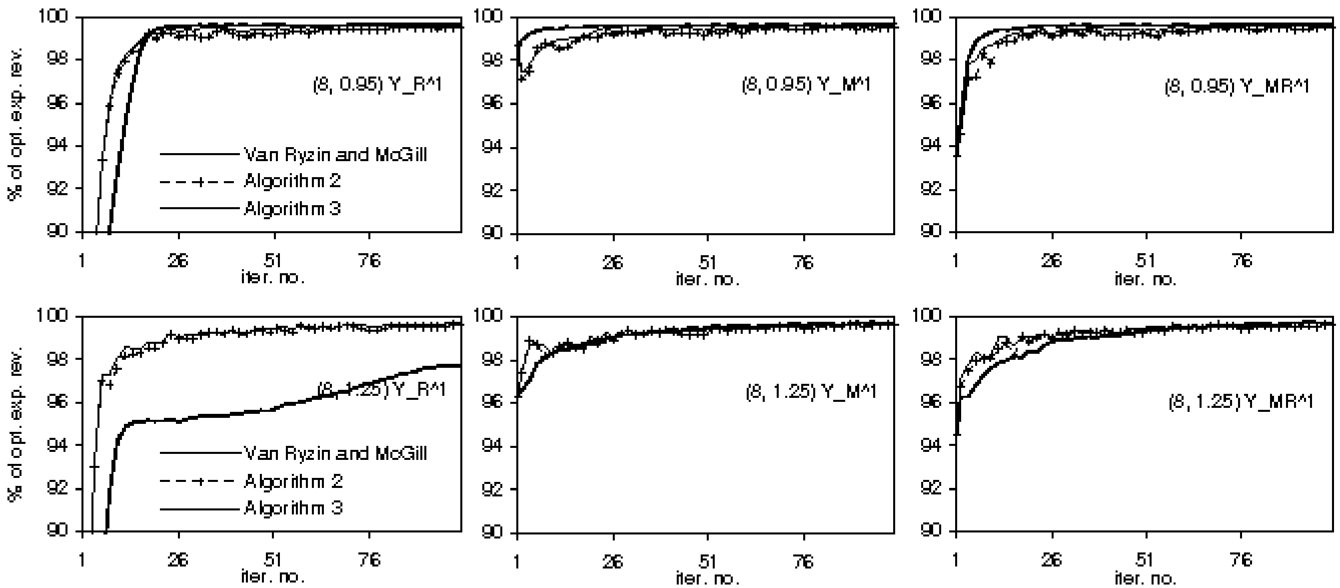

Figure 4.5: Numerical results for the test problems with 8 fare classes.

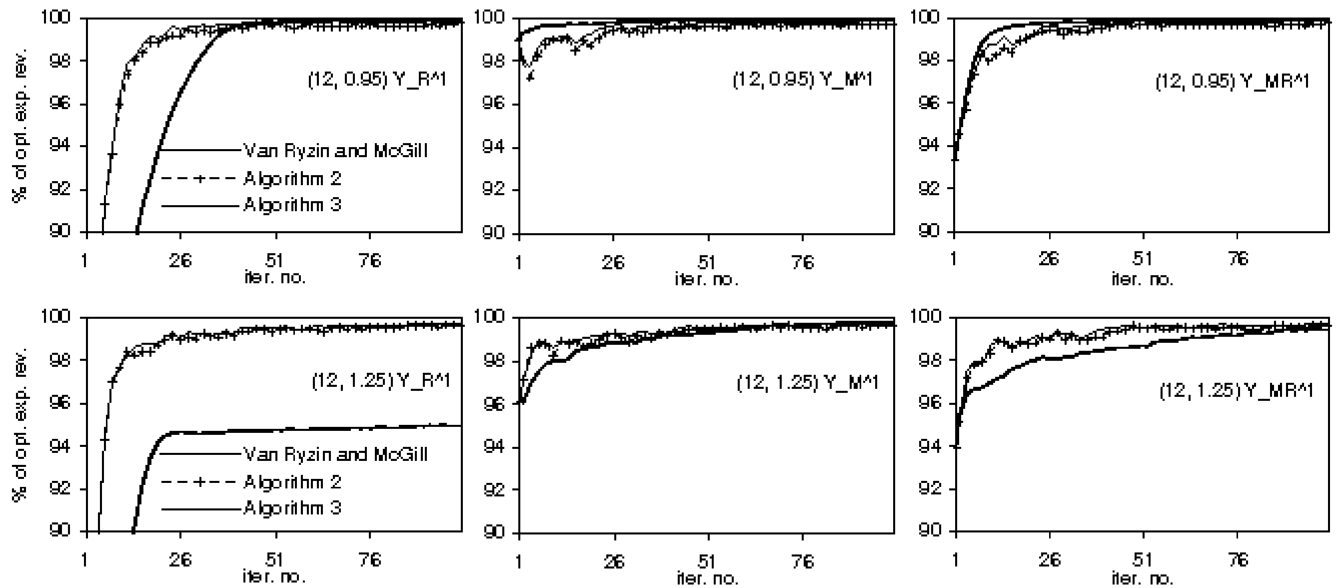

Figure 4.6: Numerical results for the test problems with 12 fare classes.

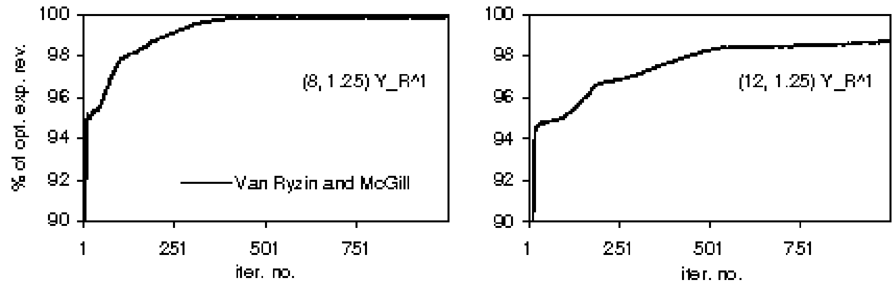

Figure 4.7: Performance of RA on test problems $(8, 1.25)$ and $(12, 1.25)$ when the initial protection levels are Y_R^1.

protection levels are good, then RA and A3 have comparable performances. RA has a slight edge when the total expected demand is below the capacity, whereas A3 has a slight edge when the total expected demand exceeds the capacity. On the other hand, the protection levels Y_R^1 are not very good since the total expected revenues obtained by these protection levels are about 71-75% of the optimal total expected revenues. In this case, the figures indicate that if the initial protection levels are not very good, then RA can lag behind A3 by a significant margin. For example, in the bottom rows of Figures 4.5 and 4.6, it appears that RA cannot get a good set of protection levels within a reasonable number of iterations when the initial protection levels are Y_R^1. To make sure that RA does not prematurely stop making progress, we run RA for 1000 iterations for these two test cases and Figure 4.7 plots the total expected revenues corresponding to the protection levels obtained by RA as a function of the iteration counter. RA eventually obtains good protection levels but this may take a large number of iterations.

Our numerical experiments indicate that the performances of A1, A2 and A3 are at least comparable to that of RA. There are some problem instances with tight capacities where the performance gap between the methods that we present in this paper and RA is significant. Since A1 has a convergence guarantee for the performance of the policy, it appears to be a good substitute for RA when the relaxed view of demand censorship described in Section 4.7.2 is possible. Despite the fact that it does not have a convergence guarantee, A3 also seems to perform quite well and this method is applicable when the demand information is censored. Finally, we note that A1, A2 and A3 provide stochastic subgradients of the value functions with respect to the seat availability, which may be useful when making tactical decisions such as setting the capacity of the flight leg.

4.9 Conclusions

We developed a stochastic approximation method to compute the optimal protection levels for the seat allocation problem under the assumption that the demand distributions are discrete. Although the problem that we consider is nonsmooth and the total expected revenue is not concave when viewed as a function of the protection levels, we were able to show that the iterates of our method converge to a set of optimal protection levels. We provided alternative versions of our method that remain applicable when the demand information is censored. Numerical experiments demonstrated that our methods are especially advantageous when the total expected demand exceeds the capacity by a significant margin and the initial protection levels are not close to the optimal protection levels.

4.10 Appendix

4.10.1 Obtaining a Stochastic Subgradient of $\mathbb{E}\{v_j(\cdot, D_j)\}$

In this section, we use induction over the fare classes to show that the recursion in (4.10) gives a stochastic subgradient of $\mathbb{E}\{v_j(\cdot, D_j)\}$ at x_j. It is easy to show the result for fare class n. Assuming that the result holds for fare class $j+1$ and $\dot{v}_j(x_j, d_j)$ is defined as in (4.10), we show that

$$v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) \leq \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$$

for all $x_j, \tilde{x}_j \in [0, c]$. Since the roles of x_j and $\tilde{x}_j$ are interchangeable, we consider six cases

Case 1. Assume that $y_j^* < x_j - d_j$ and $y_j^* < \tilde{x}_j - d_j$. We have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = \mathbb{E}\{v_{j+1}(\tilde{x}_j - d_j, D_{j+1})\} - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$, where the first equality is by (4.9), the inequality is by the induction assumption and the second equality is by (4.10).

Case 2. Assume that $y_j^* < x_j - d_j$ and $\tilde{x}_j - d_j \leq y_j^* \leq \tilde{x}_j$. By the induction assumption, we have $\mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[y_j^* - x_j + d_j]$. Since y_j^* is a maximizer of $f_j(\cdot)$ over $[0, c]$ and $y_j^* < x_j - d_j$, we also have $\mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\} \leq r_j$. Noting that $\tilde{x}_j - d_j - y_j^* \leq 0$, we obtain $r_j\,[\tilde{x}_j - d_j - y_j^*] \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[\tilde{x}_j - d_j - y_j^*]$. By (4.9), we have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = r_j\,[\tilde{x}_j - y_j^*] + \mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} - r_j\,d_j - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[\tilde{x}_j - d_j - y_j^*] + \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[y_j^* - x_j + d_j] = \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$, where the inequality follows from the two inequalities that we derive at the beginning of this case and the last equality is by (4.10).

Case 3. Assume that $y_j^* < x_j - d_j$ and $\tilde{x}_j < y_j^*$. By the induction assumption, we have $\mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[y_j^* - x_j + d_j]$. Since y_j^* is a maximizer of $f_j(\cdot)$ over $[0, c]$, we have $\mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - \mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} \leq r_j[\tilde{x}_j - y_j^*]$. Adding these two inequalities, we obtain $\mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[y_j^* - x_j + d_j] + r_j\,[\tilde{x}_j - y_j^*]$. Similar to Case 2, since y_j^* is a maximizer of $f_j(\cdot)$ over $[0, c]$ and $y_j^* < x_j - d_j$, we also have $\mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\} \leq r_j$. In this case, by (4.9), we have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = \mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - r_j\,d_j - \mathbb{E}\{v_{j+1}(x_j - d_j, D_{j+1})\} \leq -r_j\,d_j + \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[y_j^* - x_j + d_j] + r_j\,[\tilde{x}_j - y_j^*] \leq \mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\}\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$, where the first inequality follows from the inequality that we derive at the beginning of this case, the second inequality follows from the fact that $\mathbb{E}\{\dot{v}_{j+1}(x_j - d_j, D_{j+1})\} \leq r_j$, $\tilde{x}_j < y_j^*$ and $d_j \geq 0$, and the last equality is by (4.10).

Case 4. Assume that $x_j - d_j \leq y_j^* \leq x_j$ and $\tilde{x}_j - d_j \leq y_j^* \leq \tilde{x}_j$. By (4.9) and (4.10), we have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = r_j\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$.

Case 5. Assume that $x_j - d_j \leq y_j^* \leq x_j$ and $\tilde{x}_j < y_j^*$. Since y_j^* is a maximizer of $f_j(\cdot)$ over $[0, c]$, we have $\mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - \mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} \leq r_j\,[\tilde{x}_j - y_j^*]$. By (4.9), we have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = \mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - r_j\,[x_j - y_j^*] - \mathbb{E}\{v_{j+1}(y_j^*, D_{j+1})\} \leq r_j\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j - x_j]$, where the last equality follows from (4.10).

Case 6. Assume that $x_j < y_j^*$ and $\tilde{x}_j < y_j^*$. We have $v_j(\tilde{x}_j, d_j) - v_j(x_j, d_j) = \mathbb{E}\{v_{j+1}(\tilde{x}_j, D_{j+1})\} - \mathbb{E}\{v_{j+1}(x_j, D_{j+1})\} \leq \mathbb{E}\{\dot{v}_{j+1}(x_j, D_{j+1})\}\,[\tilde{x}_j - x_j] = \dot{v}_j(x_j, d_j)\,[\tilde{x}_j

$-x_j]$, where the first equality is by (4.9), the inequality is by the induction assumption and the second equality is by (4.10).

4.10.2 Proof of Corollary 34

We consider the seat allocation problem in (4.19). Using a computation similar to that in (4.2) and (4.3), we obtain

$$\tilde{v}_j(x_j, d_j) = \left\{ \max_{[x_j - d_j - 1]^+ \leq y_j \leq x_j} -r_j\, y_j + \mathbb{E}\{\tilde{v}_{j+1}(y_j, D_{j+1})\} \right\} + r_j\, x_j. \tag{4.25}$$

Therefore, the optimal policy is characterized by a set of protection levels $\{\tilde{y}_j^* : j = 1, \ldots, n\}$, where $\tilde{y}_j^*$ can be computed as a maximizer of the function

$$\tilde{f}_j(y_j) = -r_j\, y_j + \mathbb{E}\{\tilde{v}_{j+1}(y_j, D_{j+1})\} \tag{4.26}$$

over the interval $[0, c]$. As mentioned before, letting $\tilde{\mathcal{Y}}_j^* = \mathrm{argmax}_{0 \leq y_j \leq c}\, \tilde{f}_j(y_j)$, it is enough to show that $\tilde{\mathcal{Y}}_j^* \subset [L_j^*, U_j^* + n - j]$ for all $j = 1, \ldots, n$. In Lemma 36 below, we give sufficient conditions that ensure that we have $\tilde{\mathcal{Y}}_j^* \subset [L_j^*, U_j^* + n - j]$ for all $j = 1, \ldots, n$. After this, in Lemmas 37 and 38, we show that these sufficient conditions are indeed satisfied under the assumptions of Corollary 34 and this completes the proof. In Lemmas 36-38, we use $\dot{f}_j^+(y_j)$ and $\dot{f}_j^-(y_j)$ to respectively denote the right and left derivatives of the function $f_j(\cdot)$ at y_j. These derivatives exist since $f_j(\cdot)$ is concave.

Lemma 36 *Letting $f_j(\cdot)$ and $\tilde{f}_j(\cdot)$ be respectively as in* (4.4) *and* (4.26)*, if we have*

$$\dot{\tilde{f}}_j^+(y_j) \geq \dot{f}_j^+(y_j) \quad \textit{for all } y_j \in [0, c) \tag{4.27}$$

$$\dot{\tilde{f}}_j^-(y_j) \leq \dot{f}_j^-(y_j - n + j) \quad \textit{for all } y_j \in (n - j, c], \tag{4.28}$$

then we have $\tilde{\mathcal{Y}}_j^* \subset [L_j^*, U_j^* + n - j]$ for all $j = 1, \ldots, n$.

Proof Assume that $\tilde{y}_j^* \in \tilde{\mathcal{Y}}_j^*$. We first consider the case where $\tilde{y}_j^* \in (0, c)$. Since $\tilde{y}_j^*$ is a maximizer of $\tilde{f}_j(\cdot)$ over $[0, c]$, we have $\dot{\tilde{f}}_j^-(\tilde{y}_j^*) \geq 0 \geq \dot{\tilde{f}}_j^+(\tilde{y}_j^*)$. Therefore, we have $\dot{f}_j^+(\tilde{y}_j^*) \leq 0$ by (4.27), which implies that there exists $y_j^* \in \mathcal{Y}_j^*$ such that $y_j^* \leq \tilde{y}_j^*$. Consequently, we obtain $L_j^* \leq \tilde{y}_j^*$. We now show that there exists $y_j^* \in \mathcal{Y}_j^*$ such that $y_j^* \geq \tilde{y}_j^* - n + j$. If we have $\tilde{y}_j^* \leq n - j$, then the result holds trivially. If, on the other hand, we have $\tilde{y}_j^* > n - j$, then we obtain $\dot{f}_j^-(\tilde{y}_j^* - n + j) \geq \dot{\tilde{f}}_j^-(\tilde{y}_j^*) \geq 0$ by (4.28), which implies that there exists $y_j^* \in \mathcal{Y}_j^*$ such that $y_j^* \geq \tilde{y}_j^* - n + j$. Consequently, we obtain $U_j^* \geq \tilde{y}_j^* - n + j$. Therefore, we have $\tilde{y}_j^* \in [L_j, U_j + n - j]$ for any $\tilde{y}_j^* \in \tilde{\mathcal{Y}}_j^*$ and this implies that $\tilde{\mathcal{Y}}_j^* \subset [L_j^*, U_j^* + n - j]$. The cases where $\tilde{y}_j^* = 0$ or $\tilde{y}_j^* = c$ can be handled in a similar manner. □

Lemma 37 *We have $\dot{\tilde{f}}_j^+(y_j) \geq \dot{f}_j^+(y_j)$ for all $y_j \in [0, c)$, $j = 1, \ldots, n$.*

Proof By (4.4) and (4.9), we have

$$v_j(x_j, d_j) = \begin{cases} f_j(x_j - d_j) + r_j\, x_j & \text{if } y_j^* < x_j - d_j \\ f_j(y_j^*) + r_j\, x_j & \text{if } x_j - d_j \leq y_j^* \leq x_j \\ f_j(x_j) + r_j\, x_j & \text{if } x_j < y_j^*, \end{cases}$$

which implies that

$$\dot{v}_j^+(x_j, d_j) = \begin{cases} \dot{f}_j^+(x_j - d_j) + r_j & \text{if } y_j^* \leq x_j - d_j \\ r_j & \text{if } x_j - d_j < y_j^* \leq x_j \\ \dot{f}_j^+(x_j) + r_j & \text{if } x_j < y_j^*. \end{cases} \tag{4.29}$$

Similarly, by (4.25) and (4.26), we have

$$\dot{\tilde{v}}_j^+(x_j, d_j) = \begin{cases} \dot{\tilde{f}}_j^+(x_j - d_j - 1) + r_j & \text{if } \tilde{y}_j^* \leq x_j - d_j - 1 \\ r_j & \text{if } x_j - d_j - 1 < \tilde{y}_j^* \leq x_j \\ \dot{\tilde{f}}_j^+(x_j) + r_j & \text{if } x_j < \tilde{y}_j^*. \end{cases} \tag{4.30}$$

On the other hand, since y_j^* and $\tilde{y}_j^*$ are respectively maximizers of $f_j(\cdot)$ and $\tilde{f}_j(\cdot)$ over $[0, c]$, we have

$$\dot{f}_j^+(x_j) \leq 0 \quad \text{for all } x_j \in [y_j^*, c) \tag{4.31}$$

$$\dot{\tilde{f}}_j^+(x_j) \geq 0 \quad \text{for all } x_j \in [0, \tilde{y}_j^*). \tag{4.32}$$

We show the result by induction over the fare classes. In particular, assuming that the result holds for fare class j, we show that $\dot{\tilde{v}}_j^+(x_j, d_j) \geq \dot{v}_j^+(x_j, d_j)$ for all $x_j \in [0, c)$ and all $d_j \geq 0$. This result, together with (4.4) and (4.26), implies that $\dot{\tilde{f}}_{j-1}^+(y_j) \geq \dot{f}_{j-1}^+(y_j)$ for all $y_j \in [0, c)$ and the result follows.

Since $\dot{\tilde{f}}_j^+(y_j) \geq \dot{f}_j^+(y_j)$ for all $y_j \in [0, c)$ by the induction assumption, we can choose $\tilde{y}_j^*$ and y_j^* such that $\tilde{y}_j^* \geq y_j^*$. To show that $\dot{\tilde{v}}_j^+(x_j, d_j) \geq \dot{v}_j^+(x_j, d_j)$, we consider the cases listed in Table 4.2.

Case 1. We have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j$ and $\dot{\tilde{v}}_j^+(x_j, d_j) = \dot{\tilde{f}}_j^+(x_j - d_j - 1) + r_j$ by (4.29) and (4.30). Since the right derivative of a concave function is decreasing, we have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j \leq \dot{f}_j^+(x_j - d_j - 1) + r_j \leq \dot{\tilde{f}}_j^+(x_j - d_j - 1) + r_j = \dot{\tilde{v}}_j^+(x_j, d_j)$, where the last inequality is by the induction assumption.

Case 2. We have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j$ and $\dot{\tilde{v}}_j^+(x_j, d_j) = r_j$ by (4.29) and (4.30). In this case, we have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j \leq r_j = \dot{\tilde{v}}_j^+(x_j, d_j)$ by (4.31).

Table 4.2: List of cases considered in the proof of Lemma 37.

1	$y_j^* \leq \tilde{y}_j^* \leq x_j - d_j - 1 < x_j - d_j \leq x_j$
2	$y_j^* \leq x_j - d_j - 1 < \tilde{y}_j^* \leq x_j - d_j \leq x_j$
3	$y_j^* \leq x_j - d_j - 1 < x_j - d_j < \tilde{y}_j^* \leq x_j$
4	$y_j^* \leq x_j - d_j - 1 < x_j - d_j \leq x_j < \tilde{y}_j^*$
5	$x_j - d_j - 1 < y_j^* \leq \tilde{y}_j^* \leq x_j - d_j \leq x_j$

6	$x_j - d_j - 1 < y_j^* \leq x_j - d_j < \tilde{y}_j^* \leq x_j$
7	$x_j - d_j - 1 < y_j^* \leq x_j - d_j \leq x_j < \tilde{y}_j^*$
8	$x_j - d_j - 1 < x_j - d_j \leq y_j^* \leq \tilde{y}_j^* \leq x_j$
9	$x_j - d_j - 1 < x_j - d_j \leq y_j^* \leq x_j < \tilde{y}_j^*$
10	$x_j - d_j - 1 < x_j - d_j \leq x_j < y_j^* \leq \tilde{y}_j^*$

Case 3. This case is the same as Case 2.

Case 4. We have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j$ and $\dot{\tilde{v}}_j^+(x_j, d_j) = \dot{\tilde{f}}_j(x_j) + r_j$ by (4.29) and (4.30). In this case, we have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j \le r_j \le \dot{\tilde{f}}_j(x_j) + r_j = \dot{\tilde{v}}_j^+(x_j, d_j)$ by (4.31) and (4.32).

Case 5. We have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j$ and $\dot{\tilde{v}}_j^+(x_j, d_j) = r_j$ by (4.29) and (4.30). In this case, we have $\dot{v}_j^+(x_j, d_j) = \dot{f}_j^+(x_j - d_j) + r_j \le r_j = \dot{\tilde{v}}_j^+(x_j, d_j)$ by (4.31).

The other five cases can be handled in the same manner. □

Lemma 38 *We have* $\dot{\tilde{f}}_j^-(y_j) \le \dot{f}_j^-(y_j - n + j)$ *for all* $x \in (n - j, c]$.

Proof The proof follows from the same induction argument in the proof of Lemma 37. □

4.10.3 Proof of Corollary 35

Assume that we have access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$. If $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} < [x_j^k - \mathcal{O}(y_j^k)]^+$, then we deduce that $D_j^k = \min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\}$. Therefore, we know the value of D_j^k and we can compute $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, \tilde{d}_j^k\}$. If, on the other hand, $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} = [x_j^k - \mathcal{O}(y_j^k)]^+$, then we have $[x_j^k - \mathcal{O}(y_j^k)]^+ + 1 \le D_j^k + 1 = \tilde{d}_j^k$. Therefore, we deduce that $\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, \tilde{d}_j^k\} = [x_j^k - \mathcal{O}(y_j^k)]^+ + 1$. This argument shows that if we have access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+, D_j^k\} : j = 1, \ldots, n\}$, then we have access to $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, \tilde{d}_j^k\} : j = 1, \ldots, n\}$. Proposition 32 implies that if we have access to $\{y_j^k : j = 1, \ldots, n\}$, $\{x_j^k : j = 1, \ldots, n\}$ and $\{\min\{[x_j^k - \mathcal{O}(y_j^k)]^+ + 1, \tilde{d}_j^k\} : j = 1, \ldots, n\}$, then we can compute $s_j^k(y_j^k, \tilde{d}_{j+1}^k, \ldots, \tilde{d}_n^k)$ for all $j = 1, \ldots, n$.

$\square$

Chapter 5

An Alternative to Clark and Scarf's Balance Assumption for Inventory Distribution Systems

5.1 Introduction

We consider an inventory distribution system with a single warehouse that supplies multiple retailers that face random demand. We propose a new method to make the inventory replenishment decisions in such a distribution system. The traditional approach to make the inventory replenishment decisions in such systems is due to Clark and Scarf (1960). In their seminal paper, Clark and Scarf (1960) introduce the well-known balance assumption, which essentially amounts to assuming that the total amount of inventory available at all retailers can be redistributed among the retailers at no cost when needed. Under the balance assumption, Clark and Scarf (1960) show that the optimal inventory replenishment policy can be found by focusing on one installation at a time. Since then, there has been a lot of computational work showing that the inventory replenishment policies obtained under the balance assumption can perform quite well. Nevertheless, there are still a variety of practically important settings where the balance assumption remains inadequate. The method that we propose is a viable alternative when this is the case.

The method that we propose is based on formulating the problem as a dynamic program and relaxing the constraints that ensure that the shipments to the retailers are nonnegative. Although similar relaxation ideas have been used by the existing literature and in particular by Federgruen and Zipkin (1984b,c), the novel aspect of our method is that it explicitly associates Lagrange multipliers with the relaxed constraints, whereas the existing literature uses plain relaxations without penalty terms of any kind. This idea of using Lagrange multipliers in a dynamic program is relatively recent and our method is inspired by the work of Hawkins (2003) and Adelman and Mersereau (2004). It turns out that the presence of the Lagrange multipliers significantly improves the lower bounds on the value functions and the performances of the policies. We show that a good set of values for the Lagrange multipliers can be found by maximizing a concave function and we propose a technique to compute the subgradients of this function. This result provides a sensible approach for choosing a good set of values for the Lagrange multipliers. Computational experiments indicate that although our method does not always provide better performance than the inventory replenishment policies obtained under the balance assumption of Clark and Scarf (1960), it can perform well when the balance assumption remains inadequate. Furthermore, the relaxation ideas of Federgruen and Zipkin (1984b,c) can be visualized as a special case of our method that is obtained by setting all of the Lagrange multipliers to zero. Consequently, our method naturally improves the lower bounds on the value functions and the performances of the policies obtained by the relaxation ideas of Federgruen and Zipkin (1984b,c).

There is extensive literature on distribution systems. Clark and Scarf (1960) were the first to formulate the inventory control problem in a serial system as a dynamic program and to characterize the structure of the optimal inventory replenishment policy. Their work develops fundamental ideas, such as echelon inventory and induced penalty cost, to show that the inventory control problem in a serial system can be solved by focusing on one installation at a time. Clark and Scarf (1960) also introduce the balance assumption to extend their results on serial systems to distribution systems. The ideas introduced by Clark and Scarf (1960), especially the balance assumption, have been widely used since then. For example, Eppen and Schrage (1981) consider a distribution system consisting of identical retailers that face stationary and normally distributed demand and a warehouse that does not carry inventory. They derive closed-form expressions for the inventory control parameters. Federgruen and Zipkin (1984a,b) use a variation of the balance assumption for distribution systems with stockless warehouses, nonidentical retailers and nonstationary demand distributions. Federgruen and Zipkin (1984c) revisit this variation of the balance assumption when the planning horizon is infinite and obtain closed-form expressions for the inventory control parameters. We refer the reader to Axsater (2003) for an extensive and recent review of the related literature. Eppen and Schrage (1981) and Federgruen and Zipkin (1984a,b) demonstrate that the balance assumption provides satisfactory results in a variety of settings. Nevertheless, the computational experiments in Federgruen and Zipkin (1984a) indicate that the balance assumption may not perform too well when the lead time to the warehouse is long and the demands at the retailers are highly variable. Similarly, Axsater et al. (2002) consider a distribution system where the replenishment orders of the warehouse have to be in multiples of a given batch quantity and demonstrate that the balance assumption may not be satisfactory when the lead time to the warehouse is long, the demand variability is high, the batch size is large and there are significant differences in the cost parameters of different installations. Dogru et al. (2005) report on extensive computational experiments on a distribution system and reach similar conclusions to those in Federgruen and Zipkin (1984a) and Axsater et al. (2002).

The idea of relaxing the constraints in a dynamic program by associating Lagrange multipliers with them is used in the literature. Hawkins (2003) and Adelman and Mersereau (2004) recently formalize this idea by introducing the phrase weakly-coupled dynamic program. In such a dynamic program, each component of the state variable is affected by different types of decisions but these decisions interact through a set of linking constraints. Cheung and Powell (1996), Topaloglu and Kunnumkal (2006) and Topaloglu (2006) successfully use the weakly-coupled dynamic programming framework and the Lagrangian relaxation strategy in dynamic fleet management, inventory control and network revenue management settings. Computational experiments in these papers indicate that the Lagrangian relaxation strategy outperforms the existing benchmarks.

Our research contributions are twofold. From the methodology standpoint, we propose a new and tractable method to make the inventory replenishment decisions in a distribution system. Our method is based on formulating the problem as a dynamic program and relaxing certain constraints by associating Lagrange multipliers with them. We show that our method provides lower bounds on the value functions and a good set values for the Lagrange multipliers can easily be obtained by maximizing a concave function. Since the relaxation ideas of Federgruen and Zipkin (1984b,c) can be visualized as a special case of our method that is obtained by setting all of the Lagrange multipliers to zero, our method naturally improves the lower bounds on the value functions and the performances of the policies obtained by the relaxation ideas used by the existing literature. From the computational standpoint, we demonstrate that the lower bounds on the value functions and the performances of the policies obtained by our method can be better than the ones obtained under the balance assumption of Clark and Scarf (1960). In the process, we identify the conditions under which the policies obtained by our method perform better than the policies obtained under the balance assumption.

The rest of this chapter is organized as follows. Section 5.2 formulates the problem as a dynamic program. Section 5.3 describes the Lagrangian relaxation strategy. Section 5.4 briefly reviews the balance assumption of Clark and Scarf (1960) and relates it to our method. This section also compares the relaxation ideas of Federgruen and Zipkin (1984b,c) with the balance assumption of Clark and Scarf (1960). Section 5.5 shows that applying the greedy policies obtained under the Lagrangian relaxation strategy requires solving optimization problems with separable piecewise-linear convex objective functions, which can easily be done by using marginal analysis. Section 5.6 presents our computational experiments.

5.2 Problem Formulation

We consider a distribution system consisting of multiple retailers and a warehouse. The retailers face random demand and they are supplied by the warehouse, which is, in turn, supplied by an external supplier with infinite supply. The problem takes place over the finite planning horizon $\mathcal{T} = \{1, \ldots, \tau\}$. The set of retailers is $\mathcal{I}$ and we denote the warehouse by ϕ. We use the term installation whenever we want to refer to a retailer or the warehouse without making a distinction. We let D_{it} be the demand at retailer i at time period t. We assume that the demands at different retailers or at different time periods are independent. We let $D_{\phi t} = \sum_{i \in \mathcal{I}} D_{it}$ so that we can also refer to the demand at the warehouse.

For notational clarity, we assume that the lead times for all replenishments are zero. Specifically, the product shipped to a certain installation at a certain time period reaches the installation at the same time period. Through standard arguments, one can show that all of our results continue to hold when the lead times are nonzero and when presenting our computational results, we indeed consider test problems with nonzero lead times. Under the assumption that the lead times are zero, the following sequence of events take place at a particular time period. 1) The retailers place their replenishment orders from the warehouse and the warehouse places its replenishment order from the external supplier. 2) Based on the replenishment orders and the product availability, the warehouse supplies the retailers and the external supplier supplies the warehouse. 3) The retailers and the warehouse receive the replenishment orders that were shipped at the same time period. 4) The demands at the retailers are observed. The unsatisfied demands at the retailers are backlogged. The holding and backlogging costs are incurred.

We let x_{it} be the echelon inventory position at installation i at time period t. Since the lead times are zero, $\{x_{it} : i \in \mathcal{I}\}$ are simply the difference between the on-hand inventory and backlogs at the retailers at time period t. For the warehouse, $x_{\phi t}$ includes the on-hand inventory at the warehouse, on-hand inventory at the retailers and backlogs at the retailers at time period t. In particular, $x_{\phi t} - \sum_{i \in \mathcal{I}} x_{it}$

is the on-hand inventory at the warehouse at time period t. We refer the reader to Clark and Scarf (1960) and Zipkin (2000) for a detailed discussion of the echelon inventory concept. The holding and backlogging costs are accounted for by using the echelon inventory position. In particular, given that the echelon inventory position at installation i is x_{it} and q_{it} units of product is shipped to installation i at time period t, the expected holding and backlogging costs incurred at installation i at time period t is given by

$$L_{it}(x_{it} + q_{it}) = h_{it}\,\mathbb{E}\big\{\big[x_{it} + q_{it} - D_{it}\big]^+\big\} + b_{it}\,\mathbb{E}\big\{\big[D_{it} - x_{it} - q_{it}\big]^+\big\},$$

where h_{it} and b_{it} are the per unit holding and backlogging costs at installation i at time period t and $[\cdot]^+ = \max\big\{\cdot, 0\big\}$.

Using $x_t = \big\{x_{it} : i \in \mathcal{I} \cup \{\phi\}\big\}$ as the state variable at time period t, and letting $q_t = \big\{q_{it} : i \in \mathcal{I} \cup \{\phi\}\big\}$ be the vector of shipment quantities to the installations and $D_t = \big\{D_{it} : i \in \mathcal{I} \cup \{\phi\}\big\}$ be the vector of demands at time period t, the optimal policy can be found by computing the value functions $\big\{V_t(\cdot) : t \in \mathcal{T}\big\}$ through the optimality equation

$$\begin{aligned} V_t(x_t) = \min \quad & \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it}\, q_{it} + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(x_{it} + q_{it}) + \mathbb{E}\big\{V_{t+1}(x_t + q_t - D_t)\big\} \\ \text{subject to} \quad & \sum_{i \in \mathcal{I}} q_{it} \le x_{\phi t} - \sum_{i \in \mathcal{I}} x_{it} \\ & q_{it} \ge 0 \qquad\qquad \text{for all } i \in \mathcal{I} \cup \{\phi\}, \end{aligned}$$

where c_{it} is the per unit replenishment cost at installation i at time period t. We emphasize that since $x_{\phi t}$ includes the on-hand inventory and backlogs at the retailers, the echelon inventory position at the warehouse at time period $t+1$ is computed as $x_{\phi t} + q_{\phi t} - D_{\phi t}$. Since $x_{\phi t} - \sum_{i \in \mathcal{I}} x_{it}$ is the on-hand inventory at the warehouse at time period t, the first constraint in the problem above ensures that the shipments to the retailers do not violate the inventory availability at the warehouse. Defining the decision variables $y_t = \{y_{it} : i \in \mathcal{I} \cup \{\phi\}\}$ as $y_{it} = x_{it} + q_{it}$, the optimality equation above becomes

$$V_t(x_t) = \min \quad \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it} \left[y_{it} - x_{it}\right] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(y_{it}) + \mathbb{E}\left\{V_{t+1}(y_t - D_t)\right\} \quad (5.1)$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} y_{it} \leq x_{\phi t} \quad (5.2)$$

$$y_{it} \geq x_{it} \qquad \text{for all } i \in \mathcal{I} \cup \{\phi\}. \quad (5.3)$$

Due to the large number of dimensions of the state vector, solving the optimality equation above through classical dynamic programming techniques is difficult. In the next section, we propose a Lagrangian relaxation strategy that relaxes the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ in problem (5.1)-(5.3) by associating positive Lagrange multipliers with them. In this case, the optimality equation decomposes by the installations. We make this idea precise in the next section.

5.3 Lagrangian Relaxation Strategy

Associating the positive Lagrange multipliers $\lambda = \{\lambda_{it} : i \in \mathcal{I},\ t \in \mathcal{T}\}$ with the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ in problem (5.1)-(5.3), the Lagrangian relaxation strategy solves the optimality equation

$$V_t^L(x_t \,|\, \lambda) = \min \quad c_{\phi t}\big[y_{\phi_t} - x_{\phi t}\big] + L_{\phi t}(y_{\phi t}) + \sum_{i \in \mathcal{I}} \big[c_{it} - \lambda_{it}\big]\,\big[y_{it} - x_{it}\big]$$
$$+ \sum_{i \in \mathcal{I}} L_{it}(y_{it}) + \mathbb{E}\big\{V_{t+1}^L(y_t - D_t \,|\, \lambda)\big\} \tag{5.4}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} y_{it} \leq x_{\phi t} \tag{5.5}$$

$$y_{\phi t} \geq x_{\phi t}, \tag{5.6}$$

where the argument λ in the value functions emphasizes that the solution to the optimality equation above depends on the Lagrange multipliers. We note that since we have $q_{it} = y_{it} - x_{it}$, relaxing the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ is equivalent to relaxing the constraints that ensure the nonnegativity of the shipments to the retailers.

If we assume for the moment that the warehouse has infinite supply, the replenishment quantities of retailer i are not restricted to be positive and the per unit replenishment costs of retailer i are deflated by $\big\{\lambda_{it} : t \in \mathcal{T}\big\}$, then the optimal inventory replenishment policy of retailer i can be found by solving the optimality equation

$$v_{it}^L(x_{it} \,|\, \lambda) = \min_{y_{it}} \Big\{ \big[c_{it} - \lambda_{it}\big]\,\big[y_{it} - x_{it}\big] + L_{it}(y_{it}) + \mathbb{E}\big\{v_{i,t+1}^L(y_{it} - D_{it} \,|\, \lambda)\big\} \Big\}. \tag{5.7}$$

In this case, it is possible to show that the value functions computed under the Lagrangian relaxation strategy through the optimality equation in (5.4)-(5.6) have the form

$$V_t^L(x_t \,|\, \lambda) = \sum_{i \in \mathcal{I} \cup \{\phi\}} v_{it}^L(x_{it} \,|\, \lambda), \tag{5.8}$$

where $\{v^L_{it}(\cdot \,|\, \lambda) : i \in \mathcal{I},\ t \in \mathcal{T}\}$ are as in (5.7) and

$$v^L_{\phi t}(x_{\phi t} \,|\, \lambda) = \min_{y_{\phi t} \geq x_{\phi t}} \Big\{ c_{\phi t} \left[y_{\phi t} - x_{\phi t} \right] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{ v^L_{\phi, t+1}(y_{\phi t} - D_{\phi t} \,|\, \lambda) \big\} \Big\} + \Delta^L_t(x_{\phi t} \,|\, \lambda). \quad (5.9)$$

We shortly characterize the functions $\{\Delta^L_t(\cdot \,|\, \lambda) : t \in \mathcal{T}\}$. Consequently, the value functions computed under the Lagrangian relaxation strategy are separable and the value function components $\{v^L_{it}(\cdot \,|\, \lambda) : i \in \mathcal{I},\ t \in \mathcal{T}\}$ corresponding to the retailers are computed by assuming that there is unlimited supply at the warehouse and the replenishment quantities can take negative values, whereas the value function components $\{v^L_{\phi t}(\cdot \,|\, \lambda) : t \in \mathcal{T}\}$ corresponding to the warehouse are computed by inflating the costs by using the functions $\{\Delta^L_t(\cdot \,|\, \lambda) : t \in \mathcal{T}\}$.

Using induction over the time periods, it is easy to see that (5.8) holds. In particular, if we assume that the result holds for time period $t+1$, then problem (5.4)-(5.6) implies that

$$\begin{aligned} V^L_t(x_t \,|\, \lambda) = \min \quad & c_{\phi t}\left[y_{\phi_t} - x_{\phi t} \right] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{ v^L_{\phi, t+1}(y_{\phi t} - D_{\phi t} \,|\, \lambda) \big\} \\ & + \sum_{i \in \mathcal{I}} \left[c_{it} - \lambda_{it} \right] \left[y_{it} - x_{it} \right] + \sum_{i \in \mathcal{I}} L_{it}(y_{it}) \\ & + \sum_{i \in \mathcal{I}} \mathbb{E}\big\{ v^L_{i, t+1}(y_{it} - D_{it} \,|\, \lambda) \big\} \qquad (5.10) \\ \text{subject to} \quad & (5.5), (5.6). \qquad (5.11) \end{aligned}$$

Noting that the optimal solution to problem (5.7) does not depend on x_{it}, we let $\hat{r}^\lambda_{it}$ be the optimal solution to this problem. Therefore, we have

$$v^L_{it}(x_{it} \,|\, \lambda) = \left[c_{it} - \lambda_{it} \right] \left[\hat{r}^\lambda_{it} - x_{it} \right] + L_{it}(\hat{r}^\lambda_{it}) + \mathbb{E}\big\{ v^L_{i, t+1}(\hat{r}^\lambda_{it} - D_{it} \,|\, \lambda) \big\}.$$

Adding and subtracting $\sum_{i \in \mathcal{I}} v^L_{it}(x_{it} \,|\, \lambda)$ and using the expression above, the objective function of problem (5.10)-(5.11) can be written as

$$c_{\phi t}\big[y_{\phi_t} - x_{\phi t}\big] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{v^L_{\phi,t+1}(y_{\phi t} - D_{\phi t} \,|\, \lambda)\big\} + \sum_{i \in \mathcal{I}} v^L_{it}(x_{it} \,|\, \lambda)$$
$$+ \sum_{i \in \mathcal{I}} \big[c_{it} - \lambda_{it}\big] \big[y_{it} - \hat{r}^{\lambda}_{it}\big] + \sum_{i \in \mathcal{I}} L_{it}(y_{it}) + \sum_{i \in \mathcal{I}} \mathbb{E}\big\{v^L_{i,t+1}(y_{it} - D_{it} \,|\, \lambda)\big\}$$
$$- \sum_{i \in \mathcal{I}} L_{it}(\hat{r}^{\lambda}_{it}) - \sum_{i \in \mathcal{I}} \mathbb{E}\big\{v^L_{i,t+1}(\hat{r}^{\lambda}_{it} - D_{it} \,|\, \lambda)\big\}. \quad (5.12)$$

The decision variables $\big\{y_{it} : i \in \mathcal{I}\big\}$ appear only in constraints (5.5), whereas the decision variable $y_{\phi t}$ appears only in constraint (5.6) in problem (5.10)-(5.11). Therefore, using (5.12) and letting

$$\Delta^L_t(x_{\phi t} \,|\, \lambda) = \min \quad \sum_{i \in \mathcal{I}} \big[c_{it} - \lambda_{it}\big] \big[y_{it} - \hat{r}^{\lambda}_{it}\big] + \sum_{i \in \mathcal{I}} L_{it}(y_{it})$$
$$+ \sum_{i \in \mathcal{I}} \mathbb{E}\big\{v^L_{i,t+1}(y_{it} - D_{it} \,|\, \lambda)\big\} - \sum_{i \in \mathcal{I}} L_{it}(\hat{r}^{\lambda}_{it})$$
$$- \sum_{i \in \mathcal{I}} \mathbb{E}\big\{v^L_{i,t+1}(\hat{r}^{\lambda}_{it} - D_{it} \,|\, \lambda)\big\} \quad (5.13)$$
$$\text{subject to} \quad \sum_{i \in \mathcal{I}} y_{it} \le x_{\phi t}, \quad (5.14)$$

problem (5.10)-(5.11) becomes

$$V^L_t(x_t \,|\, \lambda) = \min_{y_{\phi t} \ge x_{\phi t}} \Big\{c_{\phi t}\big[y_{\phi_t} - x_{\phi t}\big] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{v^L_{\phi,t+1}(y_{\phi t} - D_{\phi t} \,|\, \lambda)\big\}\Big\}$$
$$+ \sum_{i \in \mathcal{I}} v^L_{it}(x_{it} \,|\, \lambda) + \Delta^L_t(x_{\phi t} \,|\, \lambda),$$

in which case (5.9) implies that the result holds for time period t. The induction argument can be completed by showing in a straightforward fashion that the result holds for time period τ as well. This shows that (5.8) holds for all time periods and the functions $\big\{\Delta^L_t(\cdot \,|\, \lambda) : t \in \mathcal{T}\big\}$ are characterized by the optimal objective value of problem (5.13)-(5.14). Consequently, the optimality equation in (5.4)-(5.6) can be solved by focusing on one installation at a time. We first solve the optimality equation in (5.7) to compute $\{v^L_{it}(\cdot \,|\, \lambda) : i \in \mathcal{I},\ t \in \mathcal{T}\}$, using which we can compute $\{\Delta^L_t(\cdot \,|\, \lambda) : t \in \mathcal{T}\}$ by solving problem (5.13)-(5.14). After this, we can compute $\{v^L_{\phi t}(\cdot \,|\, \lambda) : t \in \mathcal{T}\}$ by solving the optimality equation in (5.9). The next lemma shows that the optimality equation in (5.7) can be solved myopically.

Lemma 39 *Using the boundary condition that $c_{i,\tau+1} = \lambda_{i,\tau+1} = 0$, the optimal solution to the problem*

$$\min_{y_{it}} \Big\{ \Big\{ \big[c_{it} - \lambda_{it}\big] - \big[c_{i,t+1} - \lambda_{i,t+1}\big] \Big\} \big[y_{it} - x_{it}\big] + L_{it}(y_{it}) \Big\} \tag{5.15}$$

is also $\hat{r}^\lambda_{it}$. Furthermore, we have

$$\begin{aligned} v^L_{it}(x_{it} \,|\, \lambda) = \big[c_{it} - \lambda_{it}\big] \big[\hat{r}^\lambda_{it} - x_{it}\big] + L_{it}(\hat{r}^\lambda_{it}) \\ + \sum_{t'=t+1}^{\tau} \big[c_{it'} - \lambda_{it'}\big] \big[\hat{r}^\lambda_{it'} - \hat{r}^\lambda_{i,t'-1} + \mathbb{E}\{D_{i,t'-1}\}\big] + \sum_{t'=t+1}^{\tau} L_{it'}(\hat{r}^\lambda_{it'}). \end{aligned} \tag{5.16}$$

Proof We show the result by using induction over the time periods. It is easy to show the result for the last time period. Assuming that the result holds for time period $t+1$, we have

$$\begin{aligned} \hat{r}^\lambda_{it} &= \operatorname*{argmin}_{y_{it}} \Big\{ \big[c_{it} - \lambda_{it}\big] \big[y_{it} - x_{it}\big] + L_{it}(y_{it}) + \mathbb{E}\big\{v^L_{i,t+1}(y_{it} - D_{it} \,|\, \lambda)\big\} \Big\} \\ &= \operatorname*{argmin}_{y_{it}} \Big\{ \big[c_{it} - \lambda_{it}\big] \big[y_{it} - x_{it}\big] + L_{it}(y_{it}) \\ &\quad + \big[c_{i,t+1} - \lambda_{i,t+1}\big] \mathbb{E}\big\{\hat{r}^\lambda_{i,t+1} - y_{it} + D_{it}\big\} \Big\} + L_{i,t+1}(\hat{r}^\lambda_{i,t+1}) \\ &\quad + \sum_{t'=t+2}^{\tau} \big[c_{it'} - \lambda_{it'}\big] \big[\hat{r}^\lambda_{it'} - \hat{r}^\lambda_{i,t'-1} + \mathbb{E}\{D_{i,t'-1}\}\big] + \sum_{t'=t+2}^{\tau} L_{it'}(\hat{r}^\lambda_{it'}) \\ &= \operatorname*{argmin}_{y_{it}} \Big\{ \Big\{ \big[c_{it} - \lambda_{it}\big] - \big[c_{i,t+1} - \lambda_{i,t+1}\big] \Big\} \big[y_{it} - x_{it}\big] + L_{it}(y_{it}) \Big\} \Big\}, \end{aligned}$$

which shows that (5.15) holds. In this case, the fact that $\hat{r}^\lambda_{it}$ is the optimal solution to problem (5.7) and the induction hypothesis imply that

$$\begin{aligned} v^L_{it}(x_{it} \,|\, \lambda) &= \big[c_{it} - \lambda_{it}\big] \big[\hat{r}^\lambda_{it} - x_{it}\big] + L_{it}(\hat{r}^\lambda_{it}) + \mathbb{E}\big\{v^L_{i,t+1}(\hat{r}^\lambda_{it} - D_{it} \,|\, \lambda)\big\} \\ &= \big[c_{it} - \lambda_{it}\big] \big[\hat{r}^\lambda_{it} - x_{it}\big] + L_{it}(\hat{r}^\lambda_{it}) + \big[c_{i,t+1} - \lambda_{i,t+1}\big] \mathbb{E}\big\{\hat{r}^\lambda_{i,t+1} - \hat{r}^\lambda_{it} + D_{it}\big\} \\ &\quad + L_{i,t+1}(\hat{r}^\lambda_{i,t+1}) + \sum_{t'=t+2}^{\tau} \big[c_{it'} - \lambda_{it'}\big] \big[\hat{r}^\lambda_{it'} - \hat{r}^\lambda_{i,t'-1} + \mathbb{E}\big\{D_{i,t'-1}\big\}\big] \\ &\quad + \sum_{t'=t+2}^{\tau} L_{it'}(\hat{r}^\lambda_{it'}). \end{aligned}$$

Collecting the terms in the expression above shows that (5.16) holds. □

We note that although the optimality equation in (5.7) can be solved myopically, our Lagrangian relaxation strategy is not entirely myopic since the Lagrange multipliers play the role of linking different time periods. For comparison with the balance assumption that we explain in the next section, the next lemma gives an alternative characterization of $\big\{\Delta^L_t(\cdot \,|\, \lambda) : t \in \mathcal{T}\big\}$. In this lemma and throughout the rest of this section, we let $\big\{\hat{y}^\lambda_{it}(x_{\phi t}) : i \in \mathcal{I}\big\}$ be the optimal solution to problem (5.13)-(5.14). We also let $\hat{y}^\lambda_{\phi t}(x_{\phi t})$ be the optimal solution to problem (5.9). Therefore, it is easy to see that $\hat{y}^\lambda_t(x_{\phi t}) = \big\{\hat{y}^\lambda_{it}(x_{\phi t}) : i \in \mathcal{I} \cup \{\phi\}\big\}$ is the optimal solution to problem (5.10)-(5.11), which is, noting (5.8), equivalent to problem (5.4)-(5.6).

Lemma 40 *We have*

$$\Delta_t^L(x_{\phi t}\,|\,\lambda) = \begin{cases} 0 & \text{if } \; x_{\phi t} \geq \sum_{i\in\mathcal{I}} \hat{r}_{it}^\lambda \\ \sum_{i\in\mathcal{I}} \big[c_{it}-\lambda_{it}\big]\,\big[\hat{y}_{it}^\lambda(x_{\phi t})-\hat{r}_{it}^\lambda\big] & \\ \quad + \sum_{i\in\mathcal{I}} \big[L_{it}(\hat{y}_{it}^\lambda(x_{\phi t}))-L_{it}(\hat{r}_{it}^\lambda)\big] & \\ \quad\quad + \sum_{i\in\mathcal{I}} \Big[\mathbb{E}\big\{v_{i,t+1}^L(\hat{y}_{it}^\lambda(x_{\phi t})-D_{it}\,|\,\lambda)\big\} & \\ \quad\quad\quad - \mathbb{E}\big\{v_{i,t+1}^L(\hat{r}_{it}^\lambda-D_{it}\,|\,\lambda)\big\}\Big] & \text{otherwise.} \end{cases} \tag{5.17}$$

Proof Since the optimal solution to problem (5.13)-(5.14) is $\big\{y_{it}^\lambda(x_{\phi t}) : i \in \mathcal{I}\big\}$, if we have $x_{\phi t} < \sum_{i\in\mathcal{I}} \hat{r}_{it}^\lambda$, then the result follows by plugging this solution into the objective function of problem (5.13)-(5.14). We now consider the case when $x_{\phi t} \geq \sum_{i\in\mathcal{I}} \hat{r}_{it}^\lambda$. The optimal solution to the problem

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{I}} \big[c_{it}-\lambda_{it}\big]\,\big[y_{it}-\hat{r}_{it}^\lambda\big] + \sum_{i\in\mathcal{I}} L_{it}(y_{it}) + \sum_{i\in\mathcal{I}} \mathbb{E}\big\{v_{i,t+1}^L(y_{it}-D_{it}\,|\,\lambda)\big\} \\ \text{subject to} \quad & \sum_{i\in\mathcal{I}} y_{it} \leq x_{\phi t} \end{aligned}$$

is the same as the optimal solution to problem (5.13)-(5.14). On the other hand, the unconstrained minimizer of the objective function of the problem above is $\big\{\hat{r}_{it}^\lambda : i \in \mathcal{I}\big\}$. Therefore, if we have $x_{\phi t} \geq \sum_{i\in\mathcal{I}} \hat{r}_{it}^\lambda$, then $\big\{\hat{r}_{it}^\lambda : i \in \mathcal{I}\big\}$ is the optimal solution to problem (5.13)-(5.14). In this case, the result follows by plugging the solution $\big\{\hat{r}_{it}^\lambda : i \in \mathcal{I}\big\}$ into the objective function of problem (5.13)-(5.14). □

The next proposition shows that we obtain lower bounds on the value functions by solving the optimality equation in (5.4)-(5.6).

Proposition 41 *If the Lagrange multipliers are positive, then we have* $V_t^L(x_t\,|\,\lambda) \leq V_t(x_t)$.

Proof We show the result by induction over the time periods. It is easy to show the result for the last time period. Assuming that the result holds for time period $t+1$, we let $\hat{y}_t = \{\hat{y}_{it} : i \in \mathcal{I} \cup \{\phi\}\}$ be the optimal solution to problem (5.1)-(5.3), in which case we have

$$\begin{aligned}
V_t^L(x_t \,|\, \lambda) &\leq c_{\phi t}\big[\hat{y}_{\phi_t} - x_{\phi t}\big] + L_{\phi t}(\hat{y}_{\phi t}) + \sum_{i \in \mathcal{I}} \big[c_{it} - \lambda_{it}\big] \big[\hat{y}_{it} - x_{it}\big] + \sum_{i \in \mathcal{I}} L_{it}(\hat{y}_{it}) \\
&\qquad + \mathbb{E}\big\{V_{t+1}^L(\hat{y}_t - D_t \,|\, \lambda)\big\} \\
&\leq \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it}\big[\hat{y}_{it} - x_{it}\big] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(\hat{y}_{it}) - \sum_{i \in \mathcal{I}} \lambda_{it}\big[\hat{y}_{it} - x_{it}\big] \\
&\qquad + \mathbb{E}\big\{V_{t+1}(\hat{y}_t - D_t)\big\} \\
&\leq V_t(x_t),
\end{aligned}$$

where the first inequality follows from the fact that the solution $\hat{y}_t = \{\hat{y}_{it} : i \in \mathcal{I} \cup \{\phi\}\}$ satisfies constraints (5.5) and (5.6), the second inequality follows from the induction hypothesis and the third inequality follows from the fact that the Lagrange multipliers are positive and $\hat{y}_{it} \geq x_{it}$ for all $i \in \mathcal{I}$. □

Given that the initial state variable is x_1, the minimum expected cost over the whole planning horizon is $V_1(x_1)$. Proposition 41 implies that $V_1(x_1)$ is bounded from below by $V_1^L(x_1 \,|\, \lambda)$ as long as the Lagrange multipliers are positive. Therefore, to obtain the tightest possible lower bound on $V_1(x_1)$, we can solve the problem

$$\max_{\lambda \geq 0} \big\{V_1^L(x_1 \,|\, \lambda)\big\}. \tag{5.18}$$

The next proposition shows that $V_1^L(x_1 \,|\, \cdot)$ has a subgradient, in which case Theorem 3.2.6 in Bazaraa et al. (1993) implies that $V_1^L(x_1 \,|\, \lambda)$ is a concave function of the Lagrange multipliers. Consequently, we can solve problem (5.18) by subgradient optimization or Benders decomposition; see Wolsey (1998) and Ruszczynski (2003).

We use some new notation to prove that $V_1^L(x_1 \,|\, \cdot)$ has a subgradient. For the moment, we assume that the inventory replenishment decisions for the retailers and the warehouse are made by solving problem (5.4)-(5.6). We note that this problem does not include the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ and making the inventory replenishment decisions by solving problem (5.4)-(5.6) may require shipping a negative amount of product to the retailers. We assume that such decisions are acceptable. In this case, since the optimal solution to problem (5.4)-(5.6) is given by $\hat{y}_t^\lambda(x_{\phi t}) = \big\{\hat{y}_{it}^\lambda(x_{\phi t}) : i \in \mathcal{I} \cup \{\phi\}\big\}$, the stochastic process $X_\phi^\lambda = \big\{X_{\phi t}^\lambda : t \in \mathcal{T}\big\}$ defined by $X_{\phi 1}^\lambda = x_{\phi 1}$ and

$$X_{\phi,t+1}^\lambda = \hat{y}_{\phi t}^\lambda(X_{\phi t}^\lambda) - D_{\phi t} \tag{5.19}$$

characterizes the echelon inventory position of the warehouse over the whole planning horizon. Similarly, for all $i \in \mathcal{I}$, the stochastic process $X_i^\lambda = \big\{X_{it}^\lambda : t \in \mathcal{T}\big\}$ defined by $X_{i1}^\lambda = x_{i1}$ and

$$X_{i,t+1}^\lambda = \hat{y}_{it}^\lambda(X_{\phi t}^\lambda) - D_{it} \tag{5.20}$$

characterizes the echelon inventory position of retailer i over the whole planning horizon. We emphasize that the stochastic processes $\big\{X_i^\lambda : i \in \mathcal{I}\big\}$ depend on the stochastic process X_ϕ^λ. We are now ready to prove that $V_1^L(x_1 \,|\, \cdot)$ has a subgradient.

Proposition 42 *For any two sets of Lagrange multipliers λ and $\hat{\lambda}$, we have*

$$\begin{aligned} V_1^L(x_1 \,|\, \hat{\lambda}) \leq V_1^L(x_1 \,|\, \lambda) &- \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{i1} - \lambda_{i1}\big] \big[\hat{y}_{i1}^\lambda(x_{\phi 1}) - x_{i1}\big] \\ &- \sum_{t=2}^{\tau} \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it} - \lambda_{it}\big] \, \mathbb{E}\big\{\hat{y}_{it}^\lambda(X_{\phi t}^\lambda) - X_{it}^\lambda\big\}. \end{aligned} \tag{5.21}$$

Proof We use induction over the time periods to show that

$$V_t^L(x_t \,|\, \hat{\lambda}) \leq V_t^L(x_t \,|\, \lambda) - \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it} - \lambda_{it}\big] \, \big[\hat{y}_{it}^{\lambda}(x_{\phi t}) - x_{it}\big]$$
$$- \sum_{t'=t+1}^{\tau} \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it'} - \lambda_{it'}\big] \, \mathbb{E}\big\{\hat{y}_{it'}^{\lambda}(X_{\phi t'}^{\lambda}) - X_{it'}^{\lambda} \,|\, X_{\phi t}^{\lambda} = x_{\phi t}\big\}.$$

Using the expression above with $t = 1$ completes the proof.

It is easy to show the result for the last time period. If we assume that the result holds for time period $t+1$, use it with $x_{t+1} = \hat{y}_t^{\lambda}(x_{\phi t}) - D_t$ and take the conditional expectations, then we have

$$\mathbb{E}\big\{V_{t+1}^L(\hat{y}_t^{\lambda}(x_{\phi t}) - D_t \,|\, \hat{\lambda}) \,|\, X_{\phi t}^{\lambda} = x_{\phi t}\big\}$$
$$\leq \mathbb{E}\big\{V_{t+1}^L(\hat{y}_t^{\lambda}(x_{\phi t}) - D_t \,|\, \lambda) \,|\, X_{\phi t}^{\lambda} = x_{\phi t}\big\}$$
$$- \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{i,t+1} - \lambda_{i,t+1}\big] \, \mathbb{E}\Big\{\hat{y}_{i,t+1}^{\lambda}(\hat{y}_{\phi t}^{\lambda}(x_{\phi t}) - D_{\phi t}) - \big[\hat{y}_{it}^{\lambda}(x_{\phi t}) - D_{it}\big] \,|\, X_{\phi t}^{\lambda} = x_{\phi t}\Big\}$$
$$- \sum_{t'=t+2}^{\tau} \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it'} - \lambda_{it'}\big] \mathbb{E}\big\{\mathbb{E}\big\{\hat{y}_{it'}^{\lambda}(X_{\phi t'}^{\lambda}) - X_{it'}^{\lambda} \,|\, X_{\phi,t+1}^{\lambda} = \hat{y}_{\phi t}^{\lambda}(x_{\phi t}) - D_{\phi t}\big\} | X_{\phi t}^{\lambda} = x_{\phi t}\big\}. \tag{5.22}$$

Since $X_{\phi t}^{\lambda}$ and D_t are independent, the condition $X_{\phi t}^{\lambda} = x_{\phi t}$ in the first two conditional expectations above can be dropped. Furthermore, since having $X_{\phi t}^{\lambda} = x_{\phi t}$ implies that $X_{\phi,t+1}^{\lambda} = \hat{y}_{\phi t}^{\lambda}(x_{\phi t}) - D_{\phi t}$, the last double expectation is equal to $\mathbb{E}\big\{\hat{y}_{it'}^{\lambda}(X_{\phi t'}^{\lambda}) - X_{it'}^{\lambda} \,|\, X_{\phi t}^{\lambda} = x_{\phi t}\big\}$. Since $\hat{y}_t^{\lambda}(x_{\phi t})$ is a feasible but not necessarily the optimal solution to problem (5.4)-(5.6) when we use the Lagrange multipliers $\hat{\lambda}$, we have

$$V_t^L(x_t \,|\, \hat{\lambda}) \leq c_{\phi t}\big[\hat{y}_{\phi_t}^{\lambda}(x_{\phi t}) - x_{\phi t}\big] + L_{\phi t}(\hat{y}_{\phi t}^{\lambda}(x_{\phi t})) + \sum_{i \in \mathcal{I}} \big[c_{it} - \hat{\lambda}_{it}\big] \, \big[\hat{y}_{it}^{\lambda}(x_{\phi t}) - x_{it}\big]$$
$$+ \sum_{i \in \mathcal{I}} L_{it}(\hat{y}_{it}^{\lambda}(x_{\phi t})) + \mathbb{E}\big\{V_{t+1}^L(\hat{y}_t^{\lambda}(x_{\phi t}) - D_t \,|\, \hat{\lambda})\big\}$$

$$
\begin{aligned}
&\leq c_{\phi t}\big[\hat{y}^{\lambda}_{\phi_t}(x_{\phi t}) - x_{\phi t}\big] + L_{\phi t}(\hat{y}^{\lambda}_{\phi t}(x_{\phi t})) + \sum_{i \in \mathcal{I}} \big[c_{it} - \lambda_{it}\big]\,\big[\hat{y}^{\lambda}_{it}(x_{\phi t}) - x_{it}\big] \\
&\quad + \sum_{i \in \mathcal{I}} L_{it}(\hat{y}^{\lambda}_{it}(x_{\phi t})) - \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it} - \lambda_{it}\big]\,\big[\hat{y}^{\lambda}_{it}(x_{\phi t}) - x_{it}\big] \\
&\quad + \mathbb{E}\big\{V^{L}_{t+1}(\hat{y}^{\lambda}_{t}(x_{\phi t}) - D_t \,|\, \lambda)\big\} \\
&\quad - \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{i,t+1} - \lambda_{i,t+1}\big]\,\mathbb{E}\Big\{\hat{y}^{\lambda}_{i,t+1}(\hat{y}^{\lambda}_{\phi t}(x_{\phi t}) - D_{\phi t}) \,|\, X^{\lambda}_{\phi t} = x_{\phi t}\Big\} \\
&\quad + \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{i,t+1} - \lambda_{i,t+1}\big]\,\mathbb{E}\Big\{\big[\hat{y}^{\lambda}_{it}(x_{\phi t}) - D_{it}\big] \,|\, X^{\lambda}_{\phi t} = x_{\phi t}\Big\} \\
&\quad - \sum_{t'=t+2}^{\tau} \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it'} - \lambda_{it'}\big]\,\mathbb{E}\big\{\hat{y}^{\lambda}_{it'}(X^{\lambda}_{\phi t'}) - X^{\lambda}_{it'} \,|\, X^{\lambda}_{\phi t} = x_{\phi t}\big\} \\
&= V^{L}_{t}(x_t \,|\, \lambda) - \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it} - \lambda_{it}\big]\,\big[\hat{y}^{\lambda}_{it}(x_{\phi t}) - x_{it}\big] \\
&\quad - \sum_{t'=t+1}^{\tau} \sum_{i \in \mathcal{I}} \big[\hat{\lambda}_{it'} - \lambda_{it'}\big]\,\mathbb{E}\big\{\hat{y}^{\lambda}_{it'}(X^{\lambda}_{\phi t'}) - X^{\lambda}_{it'} \,|\, X^{\lambda}_{\phi t} = x_{\phi t}\big\},
\end{aligned}
$$

where the second inequality follows from (5.22) and the equality follows from the fact that $\hat{y}^{\lambda}_{t}(x_{\phi t})$ is the optimal solution to problem (5.4)-(5.6) and the definitions of $\big\{X^{\lambda}_{i,t+1} : i \in \mathcal{I} \cup \{\phi\}\big\}$ in (5.19) and (5.20). □

In our computational experiments, we work with discrete demand distributions with finite supports. In this case, the probability laws governing the stochastic processes $\big\{X^{\lambda}_{i} : i \in \mathcal{I} \cup \{\phi\}\big\}$ can be characterized by finite-dimensional transition matrices and the expectation in (5.21) can easily be computed.

5.4 Clark and Scarf's Balance Assumption

If we assume for the moment that the warehouse has infinite supply, then the optimal inventory replenishment policy of retailer i can be found by solving the optimality equation

$$v_{it}^B(x_{it}) = \min_{y_{it} \geq x_{it}} \Big\{ c_{it} \left[y_{it} - x_{it} \right] + L_{it}(y_{it}) + \mathbb{E}\big\{ v_{i,t+1}^B(y_{it} - D_{it}) \big\} \Big\}. \tag{5.23}$$

It is well-known that the solution to the optimality equation above is characterized by a sequence of base-stock levels $\{\hat{r}_{it} : t \in \mathcal{T}\}$ so that it is optimal for retailer i to place an order of $[\hat{r}_{it} - x_{it}]^+$ units when its echelon inventory position at time period t is x_{it}; see Zipkin (2000). It is also well-known that the value functions $\{v_{it}^B(\cdot) : t \in \mathcal{T}\}$ are convex and the base-stock level $\hat{r}_{it}$ can be computed as the optimal solution to the problem

$$\min_{y_{it}} \Big\{ c_{it} \left[y_{it} - x_{it} \right] + L_{it}(y_{it}) + \mathbb{E}\big\{ v_{i,t+1}^B(y_{it} - D_{it}) \big\} \Big\}. \tag{5.24}$$

The balance assumption essentially amounts to the following two assumptions.

(A.1) When the echelon inventory position at the warehouse satisfies $x_{\phi t} < \sum_{i \in \mathcal{I}} \hat{r}_{it}$, the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ in problem (5.1)-(5.3) are redundant.

(A.2) When the echelon inventory position at the warehouse satisfies $x_{\phi t} \geq \sum_{i \in \mathcal{I}} \hat{r}_{it}$, we also have $x_{\phi t} \geq \sum_{i \in \mathcal{I}} \max\{x_{it}, \hat{r}_{it}\}$.

If (A.1) and (A.2) hold, then it is possible to show that the value functions $\{V_t^B(\cdot) : t \in \mathcal{T}\}$ computed through the optimality equation in (5.1)-(5.3) have the form

$$V_t^B(x_t) = \sum_{i \in \mathcal{I} \cup \{\phi\}} v_{it}^B(x_{it}), \tag{5.25}$$

where $\{v_{it}^B(\cdot) : i \in \mathcal{I},\ t \in \mathcal{T}\}$ are as in (5.23) and

$$v_{\phi t}^B(x_{\phi t}) = \min_{y_{\phi t} \geq x_{\phi t}} \Big\{ c_{\phi t} \left[y_{\phi t} - x_{\phi t} \right] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{ v_{\phi,t+1}^B(y_{\phi t} - D_{\phi t}) \big\} \Big\} + \Delta_t^B(x_{\phi t}). \tag{5.26}$$

We shortly characterize the functions $\{\Delta_t^B(\cdot) : t \in \mathcal{T}\}$. We use the superscript B in the value functions to emphasize that they are computed under the balance assumption. We also note that the notation that we use in this section is similar to the one in Section 5.3, but this does not create confusion since the analyses in the two sections are completely unrelated.

We now show that (5.25) holds by using induction over the time periods. It is easy to show the result for the last time period. Assuming that the result holds for time period $t+1$, we consider two cases.

Case 1 First, we consider the case when $x_{\phi t} \geq \sum_{i \in \mathcal{I}} \hat{r}_{it}$. The induction assumption implies that

$$V_t^B(x_t) = \min \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it} \left[y_{it} - x_{it} \right] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(y_{it}) + \sum_{i \in \mathcal{I} \cup \{\phi\}} \mathbb{E}\big\{ v_{i,t+1}^B (y_{it} - D_{it}) \big\} \tag{5.27}$$

$$\text{subject to} \qquad (5.2), (5.3). \tag{5.28}$$

We let $\hat{y}_{\phi t}(x_{\phi t})$ be the optimal solution to the problem

$$\min_{y_{\phi t} \geq x_{\phi t}} \Big\{ c_{\phi t} \left[y_{\phi t} - x_{\phi t} \right] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{ v_{\phi,t+1}^B (y_{\phi t} - D_{\phi t}) \big\} \Big\}. \tag{5.29}$$

If (A.2) holds, then the solution $\tilde{y}_t = \big\{ \tilde{y}_{it} : i \in \mathcal{I} \cup \{\phi\} \big\}$ obtained by letting $\tilde{y}_{it} = \max \big\{ x_{it}, \hat{r}_{it} \big\}$ for all $i \in \mathcal{I}$ and $\tilde{y}_{\phi t} = \hat{y}_{\phi t}(x_{\phi t})$ is the optimal solution to problem (5.27)-(5.28). To see this, we note that the solution $\tilde{y}_t$ is feasible to problem (5.27)-(5.28), since we have $x_{\phi t} \geq \sum_{i \in \mathcal{I}} \max \big\{ x_{it}, \hat{r}_{it} \big\}$. Furthermore, since $\hat{r}_{it}$ is the optimal solution to problem (5.24) and the objective function of this problem is convex, we have

$$\begin{aligned} & c_{it} \left[\tilde{y}_{it} - x_{it} \right] + L_{it}(\tilde{y}_{it}) + \mathbb{E}\big\{ v_{i,t+1}^B (\tilde{y}_{it} - D_{it}) \big\} \\ & \qquad = \min_{y_{it} \geq x_{it}} \Big\{ c_{it} \left[y_{it} - x_{it} \right] + L_{it}(y_{it}) + \mathbb{E}\big\{ v_{i,t+1}^B (y_{it} - D_{it}) \big\} \Big\} \end{aligned} \tag{5.30}$$

for all $i \in \mathcal{I}$. Finally, we note that $\tilde{y}_{\phi t}$ is the optimal solution to problem (5.29). Therefore, the objective value obtained by the solution $\tilde{y}_t$ for problem (5.27)-(5.28) is less than or equal to the objective value obtained by any other feasible solution to problem (5.27)-(5.28) and $\tilde{y}_t$ is the optimal solution. In this case, we have

$$\begin{aligned} V_t^B(x_t) &= \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it} \left[\tilde{y}_{it} - x_{it}\right] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(\tilde{y}_{it}) + \sum_{i \in \mathcal{I} \cup \{\phi\}} \mathbb{E}\big\{v_{i,t+1}^B(\tilde{y}_{it} - D_{it})\big\} \\ &= v_{\phi t}^B(x_{\phi t}) - \Delta_t^B(x_{\phi t}) + \sum_{i \in \mathcal{I}} \min_{y_{it} \geq x_{it}} \Big\{c_{it}\left[y_{it} - x_{it}\right] + L_{it}(y_{it}) \\ &\qquad\qquad + \mathbb{E}\big\{v_{i,t+1}^B(y_{it} - D_{it})\big\}\Big\} \\ &= \sum_{i \in \mathcal{I} \cup \{\phi\}} v_{it}^B(x_{it}) - \Delta_t^B(x_{\phi t}), \end{aligned}$$

where the first equality follows from the fact that $\tilde{y}_t$ is the optimal solution to problem (5.27)-(5.28) and the second equality follows from the fact that $\tilde{y}_{\phi t}$ is the optimal solution to problem (5.26) and the equality in (5.30). Consequently, if we assume that (A.2) holds and let $\Delta_t^B(x_{\phi t}) = 0$ for the values of $x_{\phi t}$ satisfying $x_{\phi t} \geq \sum_{i \in \mathcal{I}} \hat{r}_{it}$, then (5.25) holds for time period t.

Case 2 Second, we consider the case when $x_{\phi t} < \sum_{i \in \mathcal{I}} \hat{r}_{it}$. (A.1) implies that the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ in problem (5.1)-(5.3) are redundant. In this case, the induction assumption implies that

$$V_t^B(x_t) = \min \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it} \left[y_{it} - x_{it}\right] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(y_{it}) + \sum_{i \in \mathcal{I} \cup \{\phi\}} \mathbb{E}\big\{v_{i,t+1}^B(y_{it} - D_{it})\big\} \tag{5.31}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} y_{it} \leq x_{\phi t} \tag{5.32}$$

$$y_{\phi t} \geq x_{\phi t}. \tag{5.33}$$

We let $\{\hat{y}_{it}(x_{\phi t}) : i \in \mathcal{I}\}$ be the optimal solution to the problem

$$\begin{aligned} \min \quad & \sum_{i\in\mathcal{I}} c_{it}\left[y_{it} - x_{it}\right] + \sum_{i\in\mathcal{I}} L_{it}(y_{it}) + \sum_{i\in\mathcal{I}} \mathbb{E}\{v^B_{i,t+1}(y_{it} - D_{it})\} \\ \text{subject to} \quad & \sum_{i\in\mathcal{I}} y_{it} \le x_{\phi t} \end{aligned}$$

and continue using $\hat{y}_{\phi t}(x_{\phi t})$ to denote the optimal solution to problem (5.29). The decision variables $\{y_{it} : i \in \mathcal{I}\}$ appear only in constraints (5.32), whereas the decision variable $y_{\phi t}$ appears only in constraint (5.33) in problem (5.31)-(5.33). This implies that

$$\begin{aligned} V^B_t(x_t) &= \sum_{i\in\mathcal{I}\cup\{\phi\}} c_{it}\left[\hat{y}_{it}(x_{\phi t}) - x_{it}\right] + \sum_{i\in\mathcal{I}\cup\{\phi\}} L_{it}(\hat{y}_{it}(x_{\phi t})) \\ &\quad + \sum_{i\in\mathcal{I}\cup\{\phi\}} \mathbb{E}\{v^B_{i,t+1}(\hat{y}_{it}(x_{\phi t}) - D_{it})\} \\ &= \min_{y_{\phi t}\ge x_{\phi t}} \left\{c_{\phi t}\left[y_{\phi t} - x_{\phi t}\right] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\{v^B_{\phi,t+1}(y_{\phi t} - D_{\phi t})\}\right\} \\ &\quad + \sum_{i\in\mathcal{I}} c_{it}\left[\hat{y}_{it}(x_{\phi t}) - \hat{r}_{it}\right] + \sum_{i\in\mathcal{I}} \left[L_{it}(\hat{y}_{it}(x_{\phi t})) - L_{it}(\hat{r}_{it})\right] \\ &\quad + \sum_{i\in\mathcal{I}} \left[\mathbb{E}\{v^B_{i,t+1}(\hat{y}_{it}(x_{\phi t}) - D_{it})\} - \mathbb{E}\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\}\right] \\ &\quad + \sum_{i\in\mathcal{I}} c_{it}\left[\hat{r}_{it} - x_{it}\right] + \sum_{i\in\mathcal{I}} L_{it}(\hat{r}_{it}) + \sum_{i\in\mathcal{I}} \mathbb{E}\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\}. \end{aligned}$$

Therefore, if we let

$$\begin{aligned} \Delta^B_t(x_{\phi t}) &= \sum_{i\in\mathcal{I}} c_{it}\left[\hat{y}_{it}(x_{\phi t}) - \hat{r}_{it}\right] + \sum_{i\in\mathcal{I}} \left[L_{it}(\hat{y}_{it}(x_{\phi t})) - L_{it}(\hat{r}_{it})\right] \\ &\quad + \sum_{i\in\mathcal{I}} \left[\mathbb{E}\{v^B_{i,t+1}(\hat{y}_{it}(x_{\phi t}) - D_{it})\} - \mathbb{E}\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\}\right] \end{aligned}$$

for the values of $x_{\phi t}$ satisfying $x_{\phi t} < \sum_{i\in\mathcal{I}} \hat{r}_{it}$ and use (5.26), we obtain

$$V^B_t(x_t) = v^B_{\phi t}(x_{\phi t}) + \sum_{i\in\mathcal{I}} c_{it}\left[\hat{r}_{it} - x_{it}\right] + \sum_{i\in\mathcal{I}} L_{it}(\hat{r}_{it}) + \sum_{i\in\mathcal{I}} \mathbb{E}\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\}.$$

By the next lemma, if (A.1) holds, then we have $v^B_{it}(x_{it}) = c_{it}\,\big[\hat{r}_{it} - x_{it}\big] + L_{it}(\hat{r}_{it}) + \mathbb{E}\big\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\big\}$, in which case the expression above shows that (5.25) holds for time period t.

Lemma 43 *If we assume that* $x_{\phi t} < \sum_{i \in \mathcal{I}} \hat{r}_{it}$ *and the optimal solution to the problem*

$$V^B_t(x_t) = min \sum_{i \in \mathcal{I} \cup \{\phi\}} c_{it}\,\big[y_{it} - x_{it}\big] + \sum_{i \in \mathcal{I} \cup \{\phi\}} L_{it}(y_{it}) + \sum_{i \in \mathcal{I} \cup \{\phi\}} \mathbb{E}\big\{v^B_{i,t+1}(y_{it} - D_{it})\big\} \tag{5.34}$$

$$\text{subject to} \qquad (5.2), (5.3) \tag{5.35}$$

is the same as the optimal solution to problem (5.31)-(5.33)*, then we have* $v^B_{it}(x_{it}) = c_{it}\,\big[\hat{r}_{it} - x_{it}\big] + L_{it}(\hat{r}_{it}) + \mathbb{E}\big\{v^B_{i,t+1}(\hat{r}_{it} - D_{it})\big\}$ *for all* $i \in \mathcal{I}$*.*

Proof For all $i \in \mathcal{I}$, we let $\tilde{r}_{it}$ be the largest optimal solution to problem (5.24). If we can show that $\tilde{r}_{it} \geq x_{it}$ for all $i \in \mathcal{I}$, then $\tilde{r}_{it}$ is a feasible solution to problem (5.23) for all $i \in \mathcal{I}$ and the result follows. To get a contradiction, we assume that $y^1_t = \big\{y^1_{it} : i \in \mathcal{I} \cup \{\phi\}\big\}$ is the common optimal solution to problems (5.31)-(5.33) and (5.34)-(5.35), but we have $\tilde{r}_{i't} < x_{i't}$ for some $i' \in \mathcal{I}$. The solution $y^2_t = \big\{y^2_{it} : i \in \mathcal{I} \cup \{\phi\}\big\}$ obtained by letting $y^2_{it} = y^1_{it}$ for all $i \in \mathcal{I} \setminus \{i'\}$, $y^2_{i't} = \tilde{r}_{i't}$ and $y^2_{\phi t} = y^1_{\phi t}$ is a feasible solution to problem (5.31)-(5.33). Since the solution y^1_t satisfies constraints (5.3), we have $\tilde{r}_{i't} < x_{i't} \leq y^1_{i't}$. Therefore, since $\tilde{r}_{i't}$ is the largest minimizer of the convex function $c_{i't}\,\big[y_{i't} - x_{i't}\big] + L_{i't}(y_{i't}) + \mathbb{E}\big\{v^B_{i',t+1}(y_{i't} - D_{i't})\big\}$, the objective value obtained by the solution y^2_t for problem (5.31)-(5.33) is strictly less than the objective value obtained by the solution y^1_t. This contradicts the fact that y^1_t is the optimal solution to problem (5.31)-(5.33). □

The induction argument is now complete and (5.25) holds for all time periods. Furthermore, we can characterize the functions $\{\Delta_t^B(\cdot) : t \in \mathcal{T}\}$ by noting that

$$\Delta_t^B(x_{\phi t}) = \begin{cases} 0 & \text{if } x_{\phi t} \geq \sum_{i \in \mathcal{I}} \hat{r}_{it} \\ \sum_{i \in \mathcal{I}} c_{it} \left[\hat{y}_{it}(x_{\phi t}) - \hat{r}_{it}\right] & \\ \quad + \sum_{i \in \mathcal{I}} \left[L_{it}(\hat{y}_{it}(x_{\phi t})) - L_{it}(\hat{r}_{it})\right] & \\ \quad + \sum_{i \in \mathcal{I}} \mathbb{E}\left\{v_{i,t+1}^B(\hat{y}_{it}(x_{\phi t}) - D_{it})\right\} & \\ \quad - \sum_{i \in \mathcal{I}} \mathbb{E}\left\{v_{i,t+1}^B(\hat{r}_{it} - D_{it})\right\} & \text{otherwise.} \end{cases} \tag{5.36}$$

Consequently, the balance assumption enables us to compute the value function by focusing on one installation at a time. We first solve the optimality equation in (5.23) to compute $\{v_{it}^B(\cdot) : i \in \mathcal{I},\ t \in \mathcal{T}\}$ and the base-stock levels $\{\hat{r}_{it} : i \in \mathcal{I},\ t \in \mathcal{T}\}$, using which we can compute $\{\Delta_t^B(\cdot) : t \in \mathcal{T}\}$ through (5.36). After this, we can compute $\{v_{\phi t}^B(\cdot) : t \in \mathcal{T}\}$ by solving the optimality equation in (5.26).

We close this section by comparing the results in Sections 5.3 and 5.4. First, the form of $\Delta_t^L(\cdot \,|\, \lambda)$ that appears in (5.17) under the Lagrangian relaxation strategy is very similar to the form of $\Delta_t^B(\cdot)$ that appears in (5.36) under the balance assumption. Second, although we do not prove here, it is possible to show that the balance assumption provides lower bounds on the value functions. That is, if $\{V_t^B(\cdot) : t \in \mathcal{T}\}$ are computed through (5.23), (5.25), (5.26) and (5.36), then we have $V_t^B(x_t) \leq V_t(x_t)$. Proposition 41 shows the same result for the value functions computed under the Lagrangian relaxation strategy. Neither of these lower bounds is provably tighter than the other one. Third, under the balance assumption, we assume that the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$ are redundant only when $x_{\phi t} < \sum_{i \in \mathcal{I}} \hat{r}_{it}$ holds, but no penalty is associated with assuming that these constraints are redundant. On the other hand, under the Lagrangian relaxation strategy, we always relax the constraints $y_{it} \geq x_{it}$ for all $i \in \mathcal{I}$, but we associate Lagrange multipliers with these constraints. As our computational experiments in Section 5.6 demonstrate, neither of these strategies is consistently superior to the other one. Finally, we emphasize that using the Lagrangian relaxation strategy by setting all of the Lagrange multipliers to zero is not equivalent to using the balance assumption. This can easily be seen by noting that the constraint $y_{it} \geq x_{it}$ appears in problem (5.23) but not in problem (5.7). Consequently, the relaxation ideas used by Federgruen and Zipkin (1984b,c), which are equivalent to setting all of the Lagrange multipliers to zero, are different from the balance assumption used by Clark and Scarf (1960). We note this distinction mainly because the ideas used by Clark and Scarf (1960) and Federgruen and Zipkin (1984b,c) are sometimes both referred to as the balance assumption. Although these ideas share similarities, they are not equivalent to each other.

5.5 Applying the Greedy Policy

In this section, we examine the greedy policies obtained under the Lagrangian relaxation strategy and the balance assumption. Letting $\hat{\lambda}$ be the optimal solution to problem (5.18), the value functions $\{V_t^L(\cdot \mid \hat{\lambda}) : t \in \mathcal{T}\}$ computed under the Lagrangian relaxation strategy and the value functions $\{V_t^B(\cdot) : t \in \mathcal{T}\}$ computed under the balance assumption are separable functions of the form $\{\sum_{i \in \mathcal{I} \cup \{\phi\}} \vartheta_{it}(\cdot) : t \in \mathcal{T}\}$. Furthermore, it can be shown that the functions $\{\vartheta_{it}(\cdot) : i \in \mathcal{I} \cup \{\phi\},\ t \in \mathcal{T}\}$ are convex.

The greedy policies are obtained by replacing the value function $V_{t+1}(\cdot)$ in problem (5.1)-(5.3) with $\sum_{i \in \mathcal{I} \cup \{\phi\}} \vartheta_{i,t+1}(\cdot)$. In this case, it is easy to see that problem (5.1)-(5.3) decomposes into two problems, one for the retailers and one for the warehouse. The problem for the retailers has the form

$$\min \quad \sum_{i \in \mathcal{I}} c_{it} \big[y_{it} - x_{it} \big] + \sum_{i \in \mathcal{I}} L_{it}(y_{it}) + \sum_{i \in \mathcal{I}} \mathbb{E}\big\{ \vartheta_{i,t+1}(y_{it} - D_{it}) \big\} \tag{5.37}$$

$$\text{subject to} \quad \sum_{i \in \mathcal{I}} y_{it} \le x_{\phi t} \tag{5.38}$$

$$y_{it} \ge x_{it} \qquad \text{for all } i \in \mathcal{I}, \tag{5.39}$$

whereas the problem for the warehouse has the form

$$\min_{y_{\phi t} \ge x_{\phi t}} \Big\{ c_{\phi t} \big[y_{\phi t} - x_{\phi t} \big] + L_{\phi t}(y_{\phi t}) + \mathbb{E}\big\{ \vartheta_{\phi,t+1}(y_{\phi t} - D_{\phi t}) \big\} \Big\}. \tag{5.40}$$

In our computational experiments, we work with test problems that involve discrete demand distributions. In this case, one can show that the functions $\big\{ \vartheta_{it}(\cdot) : i \in \mathcal{I} \cup \{\phi\},\ t \in \mathcal{T} \big\}$ are piecewise-linear convex and we can easily solve the two problems above by using simple marginal analysis.

5.6 Computational Experiments

In this section, we numerically test the performance of the greedy policies obtained under the Lagrangian relaxation strategy.

5.6.1 Experimental Setup and Benchmark Strategies

Although Sections 5.2-5.5 assume that the lead times for all replenishments are zero, our computational experiments involve test problems with nonzero lead times.

We summarize the parameters of our test problems in Table 5.1. The first portion of this table shows the numbers of retailers and the lead times for the replenishments of different installations that we use in our test problems. For example, a test problem labeled with S_2 involves two retailers and the lead time for the replenishments of the warehouse is three time periods, whereas the lead times for the replenishments of the retailers are two time periods.

The second portion of Table 5.1 shows the demand distributions that we use in our test problems. The demands at the retailers always take integer values between zero and 20. For demand distribution D_1, the demand at retailer i at time period t is truncated Poisson distributed with mean α_{it} and we generate $\{\alpha_{it} : i \in \mathcal{I},\ t \in \mathcal{T}\}$ randomly. For demand distribution D_2, the demand at retailer i at time period t is uniformly distributed between l_{it} and u_{it}, and similar to D_1, we generate $\{(l_{it}, u_{it}) : i \in \mathcal{I},\ t \in \mathcal{T}\}$ randomly. For demand distribution D_3, the demands at all retailers and at all time periods are truncated Poisson distributed with mean 10. We obtain five different instances of demand distributions D_1 and D_2 by using five different random seeds to generate $\{\alpha_{it} : i \in \mathcal{I},\ t \in \mathcal{T}\}$ and $\{(l_{it}, u_{it}) : i \in \mathcal{I},\ t \in \mathcal{T}\}$. In Figure 5.1, the chart on the left side shows $\{\alpha_{it} : t \in \mathcal{T}\}$ for one instance of D_1, whereas the chart on the right side shows $\{(l_{it}, u_{it}) : t \in \mathcal{T}\}$ for one instance of D_2 for a particular retailer i. It is important to note in this figure that we generate demand distributions where there are time intervals without any demand followed by time intervals with high demand.

The third portion of Table 5.1 shows the cost structures that we use in our test problems. For cost structure C_1, all cost parameters are stationary and the cost structure at the retailers is significantly different than the one at the ware-

Table 5.1: Problem parameters used in our computational experiments.

	Number of Retailers	Lead Times
S_1	2	$[3, 1, 1]$
S_2	2	$[3, 2, 2]$
S_3	2	$[1, 3, 3]$
S_4	3	$[3, 1, 1, 1]$

	Distribution Type
D_1	Poisson, Nonstationary
D_2	Uniform, Nonstationary
D_3	Poisson, Stationary

	Warehouse Holding	Retailer Holding	Warehouse Backlogging	Retailer Backlogging	Retailer and Warehouse Replenishment
C_1	0.9	0.1	0	9	0, Stationary
C_2	0.7	0.7	1.4	1.4	0.2, Stationary
C_3	0.9	0.1	0	9	Uniform, Nonstationary

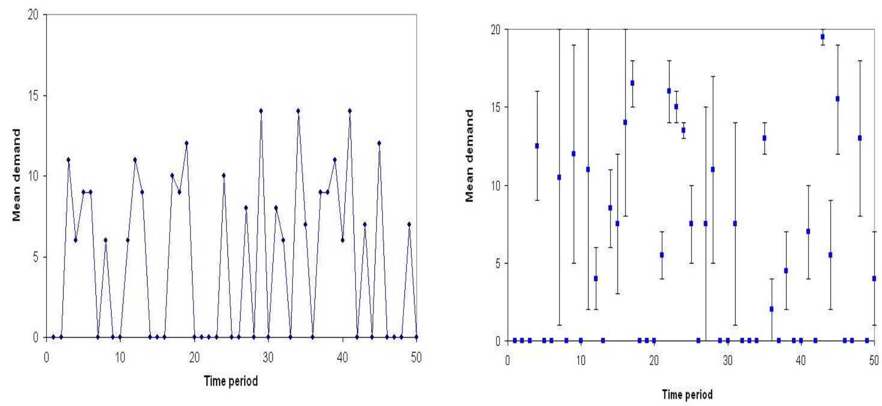

Figure 5.1: Mean demand for one instance of demand distributions D_1 (left) and D_2 (right). The bars in the chart on the right side represent the lower and upper bounds $\{(l_{it}, u_{it}) : t \in \mathcal{T}\}$.

house. This cost structure is similar to the one used by Dogru et al. (2005). For cost structure C_2, all cost parameters are also stationary, but the cost structure at the retailers is similar to the one at the warehouse. For cost structure C_3, the replenishment costs are nonstationary and the cost structure at the retailers is significantly different than the one at the warehouse. We generate the replenishment costs randomly from the uniform distribution over the interval $[0, 1]$, and use five different random seeds to generate five different instances of cost structure C_3. For cost structures C_1 and C_3, we note that the cost parameters at the retailers and at the warehouse differ by an order of magnitude.

All of our test problems involve 50 time periods. We label our test problems by (S_i, D_j, C_k, R_l), where we have $(i, j, k, l) \in \{1, \ldots, 4\} \times \{1, \ldots, 3\} \times \{1, \ldots, 3\} \times \{1, \ldots, 5\}$ and use $\{R_l : l = 1, \ldots, 5\}$ to denote the five random seeds. For example, (S_4, D_1, C_3, R_1) corresponds to a test problem with three retailers. In this test problem, the demand at the retailers are nonstationary and truncated Poisson distributed, the cost structures at the retailers and at the warehouse are significantly different, the replenishment costs are nonstationary, and the parameters of the demand distributions and the replenishment costs are generated by using random seed R_1. Since there are no parameters that are randomly generated in demand distribution D_3 and cost structures C_1 and C_2, we omit the random seed when labeling the test problems that involve this demand distribution and one of these cost structures.

In our computational experiments, we use subgradient optimization to solve problem (5.18); see Wolsey (1998). We use $20/\sqrt{k}$ as the step size at iteration k and terminate the subgradient search after 1000 iterations. These settings provide stable performance and good solutions, although the step size that we use does not guarantee convergence and one can admittedly employ more careful termination rules. After obtaining the optimal solution $\hat{\lambda}$ to problem (5.18), we let $\vartheta_{it}(\cdot) = v^L_{it}(\cdot \,|\, \hat{\lambda})$ for all $i \in \mathcal{I} \cup \{\phi\}$ and $t \in \mathcal{T}$, and solve problems (5.37)-(5.39) and (5.40) to make the inventory replenishment decisions for the retailers and the warehouse. We refer to this solution method as LR, standing for Lagrangian relaxation.

We use two benchmark strategies. The first benchmark strategy simply lets $\vartheta_{it}(\cdot) = v^L_{it}(\cdot \,|\, 0)$ for all $i \in \mathcal{I} \cup \{\phi\}$ and $t \in \mathcal{T}$, and makes the inventory replenishment decisions by using the same method as LR. Therefore, this benchmark strategy uses the trivial value of zero for all of the Lagrange multipliers instead of trying to find a good set of values by solving problem (5.18). We refer to this solution method as LR-0. We note that LR-0 is equivalent to the relaxation ideas used by Federgruen and Zipkin (1984b,c). Performance comparisons between LR and LR-0 show the importance of solving problem (5.18) to find a good set of values for the Lagrange multipliers. Specifically, they give an idea about how much the performance of the relaxation ideas used by Federgruen and Zipkin (1984b,c) can be improved by explicitly associating Lagrange multipliers with the relaxed constraints. The second benchmark strategy we use is the balance assumption described in Section 5.4. Similar to LR, we let $\vartheta_{it}(\cdot) = v_{it}^B(\cdot)$ for all $i \in \mathcal{I} \cup \{\phi\}$ and $t \in \mathcal{T}$, and solve problems (5.37)-(5.39) and (5.40) to make the inventory replenishment decisions. We refer to this solution method as BA, standing for balance assumption.

5.6.2 Computational Results

Our main computational results are summarized in three tables. Specifically, Tables 5.2, 5.3 and 5.4 respectively show the computational results for test problems with cost structures C_1, C_2 and C_3. In all of these tables, the third, fourth and fifth columns show the lower bounds on the value functions obtained by LR, BA and LR-0. The sixth column shows the ratios of the lower bounds obtained by LR and BA. The seventh, eight and ninth columns show the expected costs incurred by the greedy policies obtained by LR, BA and LR-0. We estimate these expected costs by simulating the performance of the greedy policies obtained by LR, BA and LR-0 under different demand realizations. We use enough number of demand realizations so that the performance gaps between LR and BA are always statistically significant. Only in test problems 41-44 in Table 5.2, 85-88 in Table 5.3, and 129 and 146 in Table 5.4, the performance gaps between LR and BA are not statistically significant. Finally, the tenth column shows the ratios of the expected costs incurred by the greedy policies obtained by LR and BA. All of the computational experiments were carried out in MATLAB 7.0 on a Pentium IV Desktop PC with 3.4 GHz CPU and 1 GB RAM running Windows XP. The CPU time required to solve problem (5.18) for a typical test problem with three retailers and 50 time periods is about 15 minutes.

Table 5.2 shows the results for test problems with cost structure C_1. For these test problems, the lower bounds obtained by LR are almost always tighter than the ones obtained by BA. Except for test problems 41-44, which involve stationary demand distributions, there is a significant gap between the expected costs incurred by the greedy policies obtained by BA and the lower bounds on the value functions. Although we do not give these figures in the table, this gap is about 13% on the average. The average performance gap between LR and BA is about 4% and this gap can be as high as 11% as in test problem (S_4, D_2, C_1, R_4). On the average, the performance of LR is about 5% better than that of LR-0 and this indicates that one can obtain significantly better policies by using Lagrange multipliers to penalize violations of the relaxed constraints. For cost structure C_1, the cost parameters at the retailers are significantly different than the cost parameters at the warehouse. The demand distributions for test problems 1-40 are D_1 and D_2, which are highly nonstationary. Consequently, for test problems with nonstationary demand distributions and large differences between the cost structures at different installations, BA may not perform too well and LR may improve on BA. These observations are in alignment with those of Axsater et al. (2002) and Dogru et al. (2005).

Table 5.2: Computational results for test problems with cost structure C_1.

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/V_1^B(x_1)$	LR	BA	LR-0	LR/BA
1	(S_1, D_1, C_1, R_1)	1036	1036	1013	100.05	1148	1177	1213	97.54
2	(S_1, D_1, C_1, R_2)	1133	1132	1113	100.14	1233	1254	1291	98.31
3	(S_1, D_1, C_1, R_3)	1111	1109	1086	100.15	1201	1236	1282	97.16
4	(S_1, D_1, C_1, R_4)	830	827	803	100.37	923	960	1026	96.15
5	(S_1, D_1, C_1, R_5)	880	876	851	100.42	975	1019	1071	95.70
6	(S_2, D_1, C_1, R_1)	1763	1754	1736	100.49	1870	1903	1926	98.23
7	(S_2, D_1, C_1, R_2)	1918	1913	1899	100.29	2009	2055	2078	97.77
8	(S_2, D_1, C_1, R_3)	1881	1871	1857	100.53	1967	2010	2031	97.86
9	(S_2, D_1, C_1, R_4)	1290	1282	1265	100.65	1385	1443	1500	95.98
10	(S_2, D_1, C_1, R_5)	1350	1341	1324	100.69	1445	1500	1518	96.36
11	(S_3, D_1, C_1, R_1)	2028	2002	1992	101.32	2155	2217	2245	97.23
12	(S_3, D_1, C_1, R_2)	2188	2166	2158	101.02	2283	2364	2366	96.58
13	(S_3, D_1, C_1, R_3)	2136	2108	2100	101.33	2236	2313	2340	96.68
14	(S_3, D_1, C_1, R_4)	1571	1545	1535	101.66	1673	1774	1785	94.31
15	(S_3, D_1, C_1, R_5)	1734	1713	1703	101.26	1850	1964	1980	94.19
16	(S_4, D_1, C_1, R_1)	1442	1432	1400	100.76	1591	1666	1717	95.47
17	(S_4, D_1, C_1, R_2)	1519	1508	1479	100.75	1650	1721	1760	95.86
18	(S_4, D_1, C_1, R_3)	1483	1471	1437	100.82	1642	1724	1775	95.23
19	(S_4, D_1, C_1, R_4)	1103	1096	1061	100.71	1268	1353	1450	93.72
20	(S_4, D_1, C_1, R_5)	1333	1320	1284	101.03	1477	1537	1588	96.13
21	(S_1, D_2, C_1, R_1)	973	974	960	99.98	1063	1099	1114	96.72
22	(S_1, D_2, C_1, R_2)	824	826	809	99.75	910	947	974	96.07
23	(S_1, D_2, C_1, R_3)	1119	1120	1106	99.88	1189	1216	1242	97.79
24	(S_1, D_2, C_1, R_4)	675	672	656	100.41	751	816	838	92.05

Table 5.2 (continued).

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/ V_1^B(x_1)$	LR	BA	LR-0	LR/BA
25	(S_1, D_2, C_1, R_5)	791	788	773	100.39	889	900	919	98.71
26	(S_2, D_2, C_1, R_1)	1615	1611	1598	100.26	1694	1721	1727	98.38
27	(S_2, D_2, C_1, R_2)	1331	1331	1318	100.05	1406	1448	1476	97.10
28	(S_2, D_2, C_1, R_3)	1912	1908	1889	100.22	2018	2051	2066	98.41
29	(S_2, D_2, C_1, R_4)	1100	1094	1074	100.56	1190	1266	1286	94.04
30	(S_2, D_2, C_1, R_5)	1330	1323	1306	100.48	1434	1517	1593	94.52
31	(S_3, D_2, C_1, R_1)	1951	1927	1916	101.26	2069	2145	2196	96.46
32	(S_3, D_2, C_1, R_2)	1578	1563	1552	100.94	1681	1761	1764	95.47
33	(S_3, D_2, C_1, R_3)	2148	2133	2119	100.72	2260	2312	2328	97.75
34	(S_3, D_2, C_1, R_4)	1338	1307	1291	102.40	1462	1617	1626	90.38
35	(S_3, D_2, C_1, R_5)	1629	1604	1593	101.56	1772	1940	1942	91.39
36	(S_4, D_2, C_1, R_1)	1313	1305	1283	100.66	1454	1519	1533	95.74
37	(S_4, D_2, C_1, R_2)	1216	1214	1188	100.18	1381	1447	1472	95.45
38	(S_4, D_2, C_1, R_3)	1525	1519	1494	100.40	1729	1796	1843	96.26
39	(S_4, D_2, C_1, R_4)	988	981	957	100.76	1124	1266	1305	88.77
40	(S_4, D_2, C_1, R_5)	1154	1151	1131	100.27	1276	1300	1321	98.10
41	$(S_1, D_3, C_1, .)$	3578	3578	3578	100.00	3581	3579	3581	100.05
42	$(S_2, D_3, C_1, .)$	5447	5447	5447	100.00	5459	5448	5459	100.19
43	$(S_3, D_3, C_1, .)$	5521	5520	5521	100.00	5536	5521	5536	100.27
44	$(S_4, D_3, C_1, .)$	5243	5243	5243	100.00	5250	5250	5250	100.00

Table 5.3 shows the results for test problems with cost structure C_2. For these problems, the lower bounds obtained by BA are always tighter the ones obtained by LR. Furthermore, LR and LR-0 almost always yield the same lower bounds on the value functions, which indicates that the optimal solution to problem (5.18) is very close to zero for a majority of the test problems in Table 5.3. Comparing the expected costs incurred by the greedy policies obtained by BA with the lower bounds on the value functions shows that the performance of BA is very close to optimal. The average gap between the expected costs incurred by the greedy policies obtained by BA and the lower bounds on the value functions is less than 1%. The average performance gap between BA and LR is about 2%. For test problems with similar cost structures at the retailers and at the warehouse, balance assumption provides satisfactory results and LR lags behind BA by a small but consistent margin. LR and LR-0 yield essentially the same results for these test problems, indicating that the constraints that ensure the nonnegativity of the shipments to the retailers are rarely violated even if they are relaxed without explicitly associating penalty terms with them. This also partially explains the success of the balance assumption.

Table 5.3: Computational results for test problems with cost structure C_2.

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/V_1^B(x_1)$	LR	BA	LR-0	LR/BA
45	(S_1, D_1, C_2, R_1)	818	836	818	97.84	859	840	859	102.25
46	(S_1, D_1, C_2, R_2)	860	876	860	98.21	893	877	893	101.80
47	(S_1, D_1, C_2, R_3)	839	858	839	97.76	874	858	874	101.87
48	(S_1, D_1, C_2, R_4)	639	658	639	97.20	675	656	675	102.96
49	(S_1, D_1, C_2, R_5)	686	706	686	97.10	729	710	729	102.64
50	(S_2, D_1, C_2, R_1)	1186	1202	1186	98.73	1225	1211	1225	101.21
51	(S_2, D_1, C_2, R_2)	1264	1278	1264	98.97	1295	1283	1295	100.96
52	(S_2, D_1, C_2, R_3)	1225	1238	1225	98.95	1257	1245	1257	100.97
53	(S_2, D_1, C_2, R_4)	902	917	902	98.35	937	921	937	101.67
54	(S_2, D_1, C_2, R_5)	1002	1017	1002	98.54	1040	1028	1040	101.21
55	(S_3, D_1, C_2, R_1)	1520	1530	1520	99.39	1561	1551	1562	100.68
56	(S_3, D_1, C_2, R_2)	1608	1616	1608	99.53	1637	1632	1638	100.33
57	(S_3, D_1, C_2, R_3)	1563	1572	1563	99.44	1595	1584	1595	100.68
58	(S_3, D_1, C_2, R_4)	1145	1154	1145	99.22	1176	1166	1176	100.89
59	(S_3, D_1, C_2, R_5)	1281	1289	1281	99.33	1319	1308	1319	100.91
60	(S_4, D_1, C_2, R_1)	1148	1173	1148	97.82	1204	1179	1204	102.11
61	(S_4, D_1, C_2, R_2)	1185	1208	1185	98.08	1233	1211	1233	101.80
62	(S_4, D_1, C_2, R_3)	1146	1175	1146	97.55	1202	1177	1202	102.14
63	(S_4, D_1, C_2, R_4)	843	869	843	96.98	904	871	904	103.81
64	(S_4, D_1, C_2, R_5)	984	1013	984	97.11	1042	1016	1042	102.56
65	(S_1, D_2, C_2, R_1)	794	805	794	98.71	825	807	824	102.21
66	(S_1, D_2, C_2, R_2)	655	669	655	97.90	682	671	681	101.68
67	(S_1, D_2, C_2, R_3)	852	865	852	98.52	884	863	885	102.45
68	(S_1, D_2, C_2, R_4)	545	556	545	98.00	582	558	582	104.44

Table 5.3 (continued).

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/$ $V_1^B(x_1)$	LR	BA	LR-0	LR/BA
69	(S_1, D_2, C_2, R_5)	677	689	677	98.33	701	688	702	101.91
70	(S_2, D_2, C_2, R_1)	1160	1170	1160	99.12	1188	1175	1188	101.10
71	(S_2, D_2, C_2, R_2)	946	956	946	99.01	971	964	971	100.74
72	(S_2, D_2, C_2, R_3)	1245	1260	1245	98.78	1279	1262	1280	101.37
73	(S_2, D_2, C_2, R_4)	799	814	799	98.16	838	818	839	102.44
74	(S_2, D_2, C_2, R_5)	1001	1012	1001	98.85	1028	1012	1028	101.60
75	(S_3, D_2, C_2, R_1)	1504	1510	1504	99.61	1531	1522	1531	100.60
76	(S_3, D_2, C_2, R_2)	1217	1223	1217	99.56	1242	1240	1244	100.13
77	(S_3, D_2, C_2, R_3)	1592	1601	1592	99.42	1623	1614	1623	100.56
78	(S_3, D_2, C_2, R_4)	1003	1014	1003	98.93	1042	1027	1042	101.49
79	(S_3, D_2, C_2, R_5)	1274	1281	1273	99.46	1303	1291	1303	100.91
80	(S_4, D_2, C_2, R_1)	1055	1074	1055	98.28	1104	1076	1105	102.64
81	(S_4, D_2, C_2, R_2)	1006	1028	1006	97.89	1050	1030	1051	101.96
82	(S_4, D_2, C_2, R_3)	1170	1190	1170	98.31	1236	1193	1236	103.67
83	(S_4, D_2, C_2, R_4)	830	847	830	98.02	884	848	884	104.24
84	(S_4, D_2, C_2, R_5)	929	947	929	98.11	966	948	966	101.91
85	$(S_1, D_3, C_2, .)$	1953	1953	1953	100.00	1955	1955	1955	100.00
86	$(S_2, D_3, C_2, .)$	2788	2788	2788	100.00	2788	2788	2788	100.00
87	$(S_3, D_3, C_2, .)$	3259	3259	3259	100.00	3262	3262	3262	100.00
88	$(S_4, D_3, C_2, .)$	2871	2871	2871	100.00	2872	2871	2872	100.03

Table 5.4 shows the results for test problems with cost structure C_3. For these test problems, the lower bounds obtained by LR are almost always tighter than the ones obtained by BA. The average gap between the expected costs incurred by the greedy policies obtained by BA and the lower bounds on the value functions is about 11% for test problems 89-128, whereas the same gap is about 1% for test problems 129-148. We emphasize that the demand distributions in test problems 129-148 are stationary and BA seems to provide satisfactory results for these test problems. Similarly, the average gap between the expected costs incurred by the greedy policies obtained by LR and BA is about 3% for test problems 89-128, whereas the same gap is less than 1% for test problems 129-148. For test problems 89-128, for which BA lags behind LR, the performance gap between LR and LR-0 is about 15%, but this gap reduces to 5% for test problems 129-148, for which BA and LR provide comparable results. Consequently, as mentioned in the previous paragraph, the gap between the lower bounds obtained by LR and LR-0 seems to be a good predictor of the success of the balance assumption. In particular, letting EC^{BA} be the expected cost incurred by the greedy policy obtained by BA, Figure 5.2 plots the pairs $\big[V_1^L(x_1 \,|\, \hat{\lambda})/V_1^L(x_1 \,|\, 0),\, EC^{\text{BA}}/\max\big\{V_1^L(x_1 \,|\, \hat{\lambda}), V_1^B(x_1)\big\}\big]$ for the test problems in Table 5.4. The data points in this figure fall roughly along the diagonal. For test problems 129-148, the lower bounds obtained by LR and LR-0 are similar to each other and using a value of zero for all Lagrange multipliers provides a good solution to problem (5.18). Therefore, the constraints that ensure the nonnegativity of the shipments to the retailers are rarely violated even if they are relaxed without associating penalty terms with them. For these problems, the expected costs incurred by the greedy policies obtained by BA are also very close to the tightest lower bounds on the value functions. On the other hand, for test problems 89-128, using a value of zero for all Lagrange multipliers does not provide a good solution to problem (5.18). For these test problems, there are also relatively large gaps between the expected costs incurred by the greedy policies obtained by BA and the tightest lower bounds on the value functions.

It is possible to think of a variant of the Lagrangian relaxation strategy that recomputes the lower bound on the value function at each time period. In particular, if the state variable at time period t is x_t, then we can solve the problem $\max_{\lambda \geq 0} \left\{ V_t^L(x_t \,|\, \lambda) \right\}$ to obtain the optimal solution $\hat{\lambda}_t(x_t)$. In this case, we can let $\vartheta_{i,t+1}(\cdot) = v_{i,t+1}^L(\cdot \,|\, \hat{\lambda}_t(x_t))$ for all $i \in \mathcal{I} \cup \{\phi\}$, and solve problems (5.37)-(5.39) and (5.40) to make the inventory replenishment decisions at time period t. We refer to this solution method as LR-D, standing for the dynamic implementation of Lagrangian relaxation. For a small number of test problems, Table 5.6 shows the ratios of the expected costs incurred by the greedy policies obtained by LR-D and LR. The results indicate that although LR-D always provides better performance than LR, the performance gap is quite small.

5.7 Conclusions

We developed a new method for making the inventory replenishment decisions in a distribution system. Our method is based on formulating the problem as a dynamic program and using Lagrange multipliers to relax the constraints that ensure that the shipments to the retailers are nonnegative. Since the relaxation ideas used by Federgruen and Zipkin (1984b,c) can be visualized as a special case of our method that is obtained by setting all of the Lagrange multipliers to zero, our method naturally improves the lower bounds on the value functions and the performances of the policies obtained by these relaxation ideas. Comparison of the results in Sections 5.3 and 5.4 shows that the relaxation ideas used by Federgruen and Zipkin (1984b,c) are not equivalent to the balance assumption used by Clark and Scarf (1960).

Our computational experiments indicate that one can significantly tighten the lower bounds on the value functions and improve the performances of the greedy policies by finding a good set of values for the Lagrange multipliers to penalize the violations of the relaxed constraints. Although our method does not always perform better than the inventory replenishment policies obtained under the balance assumption of Clark and Scarf (1960), it can be a viable alternative when the balance assumption remains inadequate. Considering the test problems in Tables 5.2-5.4, whenever there is a large gap between the lower bound on the value function and the performance of the greedy policy obtained under the balance assumption, our method seems to perform better than the balance assumption. A large gap between the lower bound on the value function and the performance of the greedy policy obtained under the balance assumption is an indicator of the fact that the balance assumption is not appropriate. In this case, our method provides better performance since it explicitly penalizes the violations of the constraints that the balance assumption assumes to be redundant.

Table 5.4: Computational results for test problems with cost structure C_3 .

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/ V_1^B(x_1)$	LR	BA	LR-0	LR/BA
89	(S_1, D_1, C_3, R_1)	1536	1533	1483	100.18	1638	1659	1891	98.76
90	(S_1, D_1, C_3, R_2)	1657	1652	1613	100.34	1737	1786	2152	97.29
91	(S_1, D_1, C_3, R_3)	1557	1550	1494	100.45	1664	1672	1740	99.50
92	(S_1, D_1, C_3, R_4)	1217	1207	1150	100.77	1319	1400	1772	94.21
93	(S_1, D_1, C_3, R_5)	1261	1258	1205	100.22	1378	1402	1755	98.26
94	(S_2, D_1, C_3, R_1)	2257	2246	2197	100.52	2375	2399	2907	99.02
95	(S_2, D_1, C_3, R_2)	2452	2441	2401	100.44	2573	2579	3049	99.75
96	(S_2, D_1, C_3, R_3)	2303	2285	2250	100.79	2388	2446	2572	97.63
97	(S_2, D_1, C_3, R_4)	1686	1680	1651	100.36	1799	1837	2002	97.90
98	(S_2, D_1, C_3, R_5)	1702	1692	1649	100.63	1872	1840	1977	101.71
99	(S_3, D_1, C_3, R_1)	2522	2496	2468	101.06	2633	2707	2963	97.27
100	(S_3, D_1, C_3, R_2)	2689	2655	2639	101.28	2796	2905	3083	96.24
101	(S_3, D_1, C_3, R_3)	2590	2566	2548	100.92	2684	2753	2833	97.49
102	(S_3, D_1, C_3, R_4)	1942	1912	1882	101.59	2064	2143	2841	96.31
103	(S_3, D_1, C_3, R_5)	2112	2082	2057	101.46	2293	2368	2605	96.84
104	(S_4, D_1, C_3, R_1)	2198	2167	2107	101.47	2345	2425	2923	96.71
105	(S_4, D_1, C_3, R_2)	2227	2205	2130	101.02	2388	2444	2897	97.72
106	(S_4, D_1, C_3, R_3)	2188	2155	2072	101.51	2355	2490	3070	94.58
107	(S_4, D_1, C_3, R_4)	1583	1567	1449	101.07	1764	1796	1986	98.18
108	(S_4, D_1, C_3, R_5)	2007	1981	1888	101.34	2183	2312	2801	94.42
109	(S_1, D_2, C_3, R_1)	1474	1468	1440	100.44	1554	1598	1789	97.21
110	(S_1, D_2, C_3, R_2)	1180	1177	1116	100.24	1271	1364	1658	93.19
111	(S_1, D_2, C_3, R_3)	1631	1635	1585	99.75	1792	1774	2781	101.00
112	(S_1, D_2, C_3, R_4)	1016	1017	973	99.93	1149	1162	2078	98.89
113	(S_1, D_2, C_3, R_5)	1165	1163	1105	100.16	1289	1294	2647	99.58
114	(S_2, D_2, C_3, R_1)	2153	2142	2110	100.55	2230	2305	2667	96.78

Table 5.4 (continued).

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/V_1^B(x_1)$	LR	BA	LR-0	LR/BA
115	(S_2, D_2, C_3, R_2)	1713	1713	1677	99.99	1790	1856	2335	96.45
116	(S_2, D_2, C_3, R_3)	2427	2415	2360	100.52	2546	2567	2769	99.17
117	(S_2, D_2, C_3, R_4)	1437	1425	1383	100.83	1550	1644	1961	94.26
118	(S_2, D_2, C_3, R_5)	1737	1717	1679	101.16	1848	2033	2368	90.89
119	(S_3, D_2, C_3, R_1)	2457	2428	2409	101.16	2575	2678	2747	96.17
120	(S_3, D_2, C_3, R_2)	1965	1953	1930	100.59	2050	2138	2171	95.89
121	(S_3, D_2, C_3, R_3)	2684	2662	2624	100.82	2815	2915	3335	96.57
122	(S_3, D_2, C_3, R_4)	1653	1621	1590	101.96	1789	1936	2002	92.39
123	(S_3, D_2, C_3, R_5)	2014	1993	1964	101.06	2184	2303	2579	94.83
124	(S_4, D_2, C_3, R_1)	1996	1962	1898	101.73	2139	2517	2869	84.96
125	(S_4, D_2, C_3, R_2)	1823	1813	1740	100.53	2078	2031	2342	102.28
126	(S_4, D_2, C_3, R_3)	2213	2193	2112	100.92	2439	2638	3303	92.44
127	(S_4, D_2, C_3, R_4)	1498	1482	1409	101.08	1710	1799	3086	95.02
128	(S_4, D_2, C_3, R_5)	1728	1722	1625	100.34	1880	1909	2419	98.48
129	(S_1, D_3, C_3, R_1)	4602	4571	4550	100.66	4660	4663	5124	99.95
130	(S_1, D_3, C_3, R_2)	4456	4424	4410	100.73	4501	4515	4635	99.69
131	(S_1, D_3, C_3, R_3)	4504	4504	4489	100.01	4551	4569	4754	99.61
132	(S_1, D_3, C_3, R_4)	4575	4560	4548	100.31	4645	4626	4753	100.40
133	(S_1, D_3, C_3, R_5)	4537	4539	4524	99.95	4601	4588	4724	100.27
134	(S_2, D_3, C_3, R_1)	6464	6424	6410	100.63	6531	6540	6745	99.87
135	(S_2, D_3, C_3, R_2)	6335	6292	6281	100.68	6384	6404	6491	99.70
136	(S_2, D_3, C_3, R_3)	6355	6356	6345	99.98	6412	6434	6562	99.67
137	(S_2, D_3, C_3, R_4)	6441	6421	6412	100.31	6512	6508	6597	100.05
138	(S_2, D_3, C_3, R_5)	6390	6394	6381	99.94	6443	6449	6559	99.90
139	(S_3, D_3, C_3, R_1)	6581	6546	6545	100.53	6609	6611	6629	99.96

Table 5.4 (continued).

Prob. No.	Label	$V_1^L(x_1 \mid \hat{\lambda})$	$V_1^B(x_1)$	$V_1^L(x_1 \mid 0)$	$V_1^L(x_1 \mid \hat{\lambda})/ V_1^B(x_1)$	LR	BA	LR-0	LR/BA
140	(S_3, D_3, C_3, R_2)	6386	6345	6342	100.65	6402	6406	6435	99.94
141	(S_3, D_3, C_3, R_3)	6455	6452	6446	100.04	6478	6498	6593	99.69
142	(S_3, D_3, C_3, R_4)	6552	6525	6515	100.41	6592	6627	6752	99.47
143	(S_3, D_3, C_3, R_5)	6571	6570	6569	100.00	6593	6612	6655	99.71
144	(S_4, D_3, C_3, R_1)	6590	6551	6506	100.60	6675	6744	7991	98.98
145	(S_4, D_3, C_3, R_2)	6671	6563	6534	101.65	6745	6772	7130	99.60
146	(S_4, D_3, C_3, R_3)	6700	6695	6666	100.09	6770	6768	8083	100.03
147	(S_4, D_3, C_3, R_4)	6804	6778	6759	100.39	6876	6867	7253	100.13
148	(S_4, D_3, C_3, R_5)	6615	6600	6548	100.22	6719	6823	7937	98.48

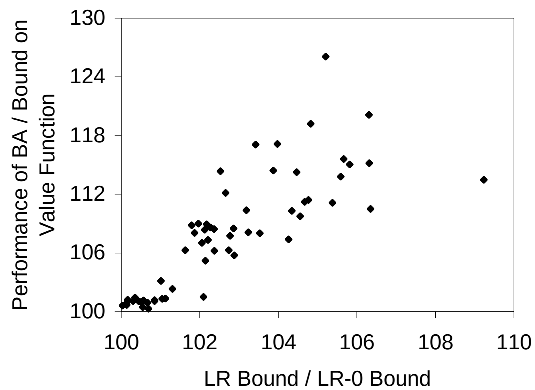

Figure 5.2: The pairs $\left[V_1^L(x_1 \,|\, \hat{\lambda})/V_1^L(x_1 \,|\, 0),\ EC^{\mathrm{BA}}/\max\left\{V_1^L(x_1 \,|\, \hat{\lambda}), V_1^B(x_1)\right\}\right]$ for the test problems in Table 5.4.

Table 5.6: Ratios of the expected costs incurred by the greedy policies obtained by LR-D and LR.

Label	LR-D/LR
(S_4, D_1, C_1, R_1)	98.72
(S_4, D_1, C_1, R_2)	99.19
(S_4, D_2, C_1, R_1)	98.97
(S_4, D_2, C_1, R_2)	98.46
(S_4, D_2, C_1, R_3)	99.16
(S_4, D_2, C_1, R_4)	98.93
(S_4, D_1, C_2, R_1)	99.50
(S_4, D_1, C_2, R_2)	99.57

Chapter 6

Summary

In this thesis, we develop approximate dynamic programming and stochastic approximation methods for problems in inventory control and revenue management. The main contributions of this work lie in establishing convergence guarantees on the solutions obtained and the improved performance when compared with many of the existing methods.

In Chapter 2, we describe a stochastic approximation algorithm for the monotone estimation problem. We use a projection operator with respect to the max norm to project the iterates onto the order simplex. We show that the iterates of the algorithm converge. We present applications to the Q-learning algorithm and the newsvendor problem with censored demands. Numerical studies indicate that exploiting the underlying structure by using the max norm projection operator results in noticeable improvements in the convergence rate.

In Chapter 3, we describe stochastic approximation methods to compute the optimal policies for a number of inventory control problems. The inventory control problems we consider are variants of multiperiod newsvendor problem, where the value functions are convex in the state variable. Consequently, the so-called base-stock policies are optimal for such problems. We show that the iterates of our methods converge to base-stock levels that are globally optimal. Moreover, our methods remain applicable even when we have access to only the sales data and not necessarily the demand data. Numerical experiments indicate that our methods can provide significant advantages over existing stochastic approximation methods.

In Chapter 4, we develop a stochastic approximation algorithm to compute the optimal protection levels for the seat allocation problem on a single flight leg. We work with the version of the problem where the demand random variables are integer valued. We show that the iterates of our algorithm converge to the optimal protection levels. We provide alternative versions of our method that are applicable even when the demand information is censored. Numerical experiments demonstrate that our method is especially advantageous when the load factor is relatively high.

In Chapter 5, we describe an approximate dynamic programming method to make the inventory replenishment decisions in a distribution system consisting of a single warehouse and multiple retailers. Our approach is based on relaxing the constraints that ensure the non-negativity of shipments to the retailers by associating Lagrange multipliers with them. We show that the solution to the relaxed problem is computationally tractable, in that it can be obtained by solving a sequence of newsvendor sub-problems. Numerical studies indicate that our method performs significantly better than traditional methods in settings where the balance assumption is not justified.

For future research, we propose to investigate methods that combine the stochastic approximation and Lagrangian relaxation methods developed in this thesis. Lagrangian relaxation provides a way to decompose a multidimensional dynamic program into a number of single dimensional sub-problems. We also have stochastic approximation methods to compute the optimal solutions to the resulting sub-problems. A natural idea then, is to combine the two methods in an iterative scheme, where we solve the sub-problems by stochastic approximation and use the optimal solutions to the sub-problems to update the Lagrange multipliers. The main benefit of the proposed method would be to make the Lagrangian relaxation idea applicable in settings where we do not have access to the distribution functions of the random variables.

BIBLIOGRAPHY

Adelman, D. and A. J. Mersereau (2004). Relaxations of weakly coupled stochastic dynamic programs. Working paper, Graduate School of Business, The University of Chicago, Chicago, IL.

Arrow, K., S. Karlin, and H. Scarf (1958). *Studies in the Mathematical Theory of Inventory and Production.* Stanford, CA: Stanford University Press.

Axsater, S. (2003). Supply chain operations: Serial and distribution inventory systems. In S. C. Graves and A. G. de Kok (Eds.), Handbook of Operations Research, Supply Chain Management: Design, Coordination and Operation , Amsterdam, North Holland, pp. 525–559. Elsevier Science Publishers B. V.

Axsater, S., J. Marklund, and E. A. Silver (2002). Heuristic methods for centralized control of one-warehouse, n-retailer inventory systems. *Manufacturing and Service Operations Management 4*, 75–97.

Azoury, K. S. (1985). Bayes solutions to dynamic inventory models under unknown demand distribution. *Management Science 31*(9), 1150–1160.

Bashyam, S. and M. C. Fu (1998). Optimizaton of (s, S) inventory systems with random lead times and a service level constraint. *Management Science 44*(12), 243–256.

Bazaraa, M. S., H. D. Sherali, and C. M. Shetty (1993). *Nonlinear Programming: Theory and Algorithms* (second ed.). New York: John Wiley & Sons, Inc.

Bertsekas, D. P. and J. N. Tsitsiklis (1996). *Neuro-Dynamic Programming.* Belmont, MA: Athena Scientific.

Bertsimas, D. and S. de Boer (2005). Simulation-based booking limits for airline revenue management. *Operations Research 53*(1), 90–106.

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

Brumelle, S. L. and J. I. McGill (1993). Airline seat allocation with multiple nested fare classes. *Operations Research 41*, 127–137.

Cheung, R. K. and W. B. Powell (1996). An algorithm for multistage dynamic networks with random arc capacities, with an application to dynamic fleet management. *Operations Research 44*(6), 951–963.

Clark, A. J. and H. Scarf (1960). Optimal policies for a multi-echelon inventory problem. *Management Science 64*, 475–490.

Conrad, S. A. (1976). Sales data and the estimation of demand. *Operations Research Quarterly 27*(1), 123–127.

Curry, R. E. (1990). Optimal airline seat allocation with fare classes nested by origins and destinations. *Transportation Science 24*, 193–204.

Ding, X. (2002). *Estimation and optimization in discrete inventory models.* Ph. D. thesis, The University of British Columbia, Vancouver, British Columbia.

Dogru, M. K., A. G. de Kok, and G. J. van Houtum (2005). A numerical study on the effect of the balance assumption in one-warehouse multi-retailer inventory systems. Beta working paper, WP 135, Department of Technology Management, Technische Universiteit Eindhoven.

Eppen, G. L. and L. Schrage (1981). Centralized ordering policies in a multi-warehouse system with leadtimes and random demand. In L. Schwarz (Ed.), Multi-Level Production/Inventory Control Systems: Theory and Practice , Amsterdam, The Netherlands, pp. 51–69. North Holland.

Ermoliev, Y. (1988). Stochastic quasigradient methods. In Y. Ermoliev and R. Wets (Eds.), Numerical Techniques for Stochastic Optimization. Berlin: Springer-Verlag.

Federgruen, A. and P. Zipkin (1984a). Allocation policies and cost approximations for multilocation inventory systems. *Naval Research Logistics 31*, 97–129.

Federgruen, A. and P. Zipkin (1984b). Approximations of dynamic, multilocation production and inventory problems. *Management Science 30*(1), 69–84.

Federgruen, A. and P. Zipkin (1984c). Computational issues in an infinite horizon multi-echelon inventory model. *Operations Research 32*, 818–836.

Fu, M. (1994). Sample path derivatives for (s, S) inventory systems. *Operations Research 42*(2), 351–363.

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

Glasserman, P. (1994). Perturbation analysis of production networks. In D. D. Yao (Ed.), Stochastic Modeling and Analysis of Manufacturing Systems. New York, NY: Springer.

Glasserman, P. and S. Tayur (1995). Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Science 41*(2), 263–281.

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

Hawkins, J. (2003). *A Lagrangian Decomposition Approach to Weakly Coupled Dynamic Optimization Problems and its Applications*. Ph. D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Huh, W. T. and P. Rusmevichientong (2006). Adaptive capacity allocation with censored demand data: Application of concave umbrella functions. Technical report, Cornell University, School of Operations Research and Industrial Engineering.

Iglehart, D. L. (1964). The dynamic inventory problem with unknown demand distribution. *Managment Science 10*(3), 429–440.

Karaesmen, I. and G. van Ryzin (2004). Overbooking with substitutable inventory classes. *Operations Research 52*(1), 83–104.

Kunnumkal, S. and H. Topaloglu (2005). Exploiting the structural properties of the underlying Markov decision problem in Q-learning algorithm. Technical report, Cornell University, School of Operations Research and Industrial Engineering. Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Kunnumkal, S. and H. Topaloglu (2006). Using stochastic approximation algorithms to compute optimal base-stock levels in inventory control problems. Tech-

nical report, Cornell University, School of Operations Research and Industrial Engineering.
Available at http://legacy.orie.cornell.edu/~huseyin/publications/publications.html.

Kushner, H. J. and D. S. Clark (1978). *Stochastic Approximation Methods for Constrained and Unconstrained Systems*. Berlin: Springer-Verlang.

Kushner, H. J. and G. G. Yin (1997). *Stochastic Approximation Algorithms and Applications*. Springer-Verlag.

Lautenbacher, C. J. and S. Stidham (1999). The underlying markov decision process in the single-leg airline yield management problem. *Transportation Science 33*, 136–146.

L'Ecuyer, P. and P. Glynn (1994). Stochastic optimization by simulation: Convergence proofs for the GI/G/1 queue in steady state. *Management Science 40*, 1245–1261.

Lee, T. and M. Hersh (1993). A model for dynamic airline seat inventory control with multiple seat bookings. *Transportation Science 27*, 252–265.

Levi, R., R. Roundy, and D. M. Shmoys (2005). Computing provably near-optimal sample-based policies for stochastic inventory control models. Technical report, Cornell University, School of Operations Research and Industrial Engineering.

Littlewood, K. (1972). Forecasting and control of passengers. In *Proceedings of the 12th AGIFORS*, New York, pp. 95–117.

Mahajan, S. and G. van Ryzin (2001). Stocking retail assortments under dynamic customer substitution. *Operations Research 49*(3), 334–351.

Nascimento, J. M. and W. B. Powell (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.* Amsterdam: North Holland.

Papadaki, K. and W. B. Powell (2003). An adaptive dynamic programming algorithm for a stochastic multiproduct batch dispatch problem. *Naval Research Logistics 50*(7), 742–769.

Porteus, E. L. (1990). *Handbooks in Operations Research and Management Science: Stochastic Models*, Volume 2, Chapter Stochastic Inventory Theory, pp. 605–652. Amsterdam: North Holland.

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

Robbins, H. and S. Monro (1951). A stochastic approximation method. *Annals of Math. Stat. 22*, 400–407.

Robinson, L. (1995). Optimal and approximate control policies for airline booking with sequential nonmonotonic fare classes. *Operations Research 43*, 252–263.

Ruszczynski, A. (2003). Decomposition methods. In A. Ruszczynski and A. Shapiro (Eds.), Handbook in Operations Research and Management Science*, Volume on* Stochastic Programming, Amsterdam. North Holland.

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.

Talluri, K. T. and G. J. van Ryzin (2004). *The Theory and Practice of Revenue Management*. Kluver Academic Publishers.

Topaloglu, H. (2005). An approximate dynamic programming approach for a product distribution problem. *IIE Transactions 37*, 711–724.

Topaloglu, H. (2006). Using Lagrangian relaxation to compute capacity-dependent bid-prices in network revenue management. Technical report, Cornell University, School of Operations Research and Industrial Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Topaloglu, H. (2007). A stochastic approximation method to compute bid prices in network revenue management problems. Technical report, Cornell University, School of Operations Research and Industrial Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Topaloglu, H. and S. Kunnumkal (2006). Approximate dynamic programming methods for an inventory allocation problem under uncertainty. *Naval Research Logistics Quarterly* (to appear).

Topaloglu, H. and W. B. Powell (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 J. McGill (2000). Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Science 46*(6), 760–775.

van Ryzin, G. and G. Vulcano (2004). Computing virtual nesting controls for network revenue management under customer choice behavior. Technical Report DRO-2004-09, Columbia Business School.

van Ryzin, G. and G. Vulcano (2006). Simulation-based optimization of virtual nesting controls for network revenue management. *Operations Research* (to appear).

Watkins, C. J. C. H. and P. Dayan (1992). *Q*-learning. *Machine Learning 8*, 279–292.

Wollmer, R. D. (1992). An airline seat management model for a single leg route when lower fare classes book first. *Operations Research 40*, 26–37.

Wolsey, L. A. (1998). *Integer Programming*. New York: John Wiley & Sons, Inc.

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the Twentieth International Conference on Machine Learning (ICML-2003)*, Washington, D.C.

Zipkin, P. (2000). *Foundations of Inventory Management*. Boston, MA: McGraw-Hill.