COMPUTING PROVABLY NEAR-OPTIMAL POLICIES FOR STOCHASTIC INVENTORY CONTROL MODELS

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

Retsef Levi

August 2005

COMPUTING PROVABLY NEAR-OPTIMAL POLICIES FOR STOCHASTIC INVENTORY CONTROL MODELS

Retsef Levi, Ph.D.

Cornell University 2005

In this work, we address several fundamental well-studied models in stochastic inventory theory. Most of the models considered here have the following common setting. A single commodity at a single location is required over a planning horizon of discrete T time periods. Specifically, there is a sequence of random demands for the commodity that are going to occur over the planning horizon. The goal is to coordinate a sequence of orders over the planning horizon to satisfy these demands with minimum overall expected ordering, holding and backlogging costs.

A stochastic dynamic programming framework has been the most dominant paradigm in the research on these models. It has been used to analyze and characterize the structure of optimal policies, as well as to provide an algorithmic framework to compute optimal policies. Unfortunately, the inventory models discussed in this work admit tractable dynamic programs only under strong assumptions on the structures of the random demands and on the way their distributions are specified for us. Moreover, these assumptions are unrealistic in most real-life scenarios. Thus, computing optimal policies (and often even good policies) for these models is a fundamental problem in inventory theory and in practice.

In the first part of this thesis, we consider these inventory models under very general assumptions on the structures of demands, specifically, allowing non-stationarity and correlation between demands in different periods. These harder and more realistic models usually admit huge dynamic programs that are not likely to be tractable. This phenomenon is known as the *curse of dimensionality*. We introduce novel and efficient

heuristics to compute provably near-optimal policies for these harder and more realistic models. That is, we analyze the performance of our algorithms and show that the policies they construct are guaranteed to have expected cost near the optimal expected cost. Our policies are simple both computationally and conceptually and do not require solving huge dynamic programs. They are based on a new cost accounting approach for stochastic inventory models and cost-balancing techniques that we believe will have more application in approximating optimal solutions for dynamic programs that arise in the context of supply chain-related models and in other domains.

In the second part of this thesis, we consider models with simple structures of demand but under the assumption that the explicit demand distributions are not known. Instead, we assume that the only information available is a set of independent samples of the demands. We then introduce sample-based policies that are computed efficiently based only on samples without any access to the explicit demand distributions. Moreover, we establish bounds on the number of samples required to guarantee that, with high probability, the expected cost of our sample-based policies will be arbitrarily close to the optimal expected cost under the assumption of full access to the explicit demand distributions. These bounds are general, easy to compute and do not depend on the specific underlying true demand distributions. We believe that this sets the foundations for new classes of sample-based policies for stochastic dynamic programs with analyzed performance.

BIOGRAPHICAL SKETCH

Retsef Levi was born March 7, 1971 in Tel-Aviv, Israel. He served about 12 years as an officer in the Israeli Defence Forces. Levi received a Bachelor's degree in Mathematics with a trend in Operations Research from Tel-Aviv University, Israel in 2001. After several months in a business development position with an Israeli hi-tech company, in the winter of 2002 Levi arrived to Cornell with his family to pursue a PhD degree in Operations Research. In the summer of 2005 he is expected to graduate and begin his Goldstine postdoctoral fellowship at IBM Watson Research Center in Yorktown Heights, NY. In the summer of 2006 Levi will begin an Assistant Professor position at MIT Sloan Business School in Boston, MA.

To Yagev, Nov and Tia, my best achievements in life so far and to my brother Regev who I always miss.

ACKNOWLEDGEMENTS

The past three and a half years at the Cornell ORIE department have been an intellectually exciting period of time for me. I enjoyed every second, expanded my intellectual horizons and was fortunate to get to know and work with some of the most impressive people I have met in my entire life. Being part of top-level organizations in the past, I feel that the department is special in providing its members the best professional environment with friendly and open atmosphere. This has made every day in my years at Cornell exciting and self fulfilling. For all of that, I would like to thank the department's faculty, staff and students as a whole.

Personally, I was fortunate to be advised by two special people, Robin Roundy and David Shmoys. David, being the PhD director at the time, was the first person with whom I communicated at Cornell. He was personally responsible for my decision to come to Cornell with two small children. Later as my advisor, he became a role model for me in terms of how academic research work should be conducted and presented and, what it means to be a professor, a researcher and a teacher. Robin was my first teacher at Cornell, and he is responsible for lighting up my love to inventory theory that has since become my main research domain. Robin has taught me much more than just inventory theory. His integrity and enthusiastic love for research made every minute I spent with him filled with unique intellectual energy. Both Robin and David were as supportive as advisors can be in all professional and personal aspects. For that I deeply thank them.

During my time as a PhD student, I was fortunate to have numerous discussions about research and life in general with many impressive individuals. I have learnt from these discussions about many things. They have contributed to my research work and enriched me as a person. I would like to thank Shane Henderson who served on my committee for our mutual discussions which helped me to significantly improve many of the results included in this thesis; Jim Renegar, who also served on my committee, and has always been extremely supportive and helpful in many personal matters; Jack Muckstadt, with whom I had many hours of interesting and stimulating discussions about research and life; Mike Todd, maybe the smartest and most modest person I happened to know, who always had time to answer questions and discuss crazy research ideas; Adrian Lewis, with whom discussions have lead to some of the results in the third chapter of this thesis; finally, Gennady Samorodnitsky, Bob Bland, David Williamson, Narahari Prabhu, Mark Eisner and Peter Jackson who always were ready to listen. I would also like to specially thank Sid Resnick for many soccer games, good jokes and a lot of fun and interesting discussions.

This is a good place to thank many of my fellow colleague students, Frans Schalekamp who always had time to listen to new research results and to make good comments; Anke van Zuylen, who always had good notes to provide and was very kind and helpful in the last months when I was away and missed many classes; Martin Pál, Gavin Hurely and Van Anh Truong with whom I coauthored papers; Sam Ehrlichman and Yuri Zinchenko who educated me on how to use computers efficiently; and finally Sujin Kim and Jose Prina with whom it was always a pleasure to talk to and have lunch together.

My research was partially supported by the following NSF grants: CCR-9912422, CCR-0430682, DMI-0500263 and DMI-0075627. I was also partially supported by a grant of Motorola.

Last but certainly not least, to Anat, my partner, the love of my life and the mother of our three children. She is definitely the main reason why I arrived to Cornell to start a PhD track. She is definitely the main reason for all of the achievements we have shared together so far. Her love, support and confidence in me were always the fuel in my body, brain and heart.

TABLE OF CONTENTS

Chapter 1

Introduction

In this work, we consider several fundamental models in stochastic inventory theory. Most of the models considered here have the following common setting. A single commodity at a single location is required over a planning horizon of discrete T time periods numbered by $t = 1, \ldots, T$. Specifically, there is a sequence of random demands for the commodity, denoted, respectively, by $D_1, \ldots, D_T$, that are going to occur over the planning horizon. The goal is to coordinate a sequence of orders over the planning horizon to satisfy these demands with minimum expected cost. The cost consists of a per-unit *ordering cost* for ordering supply units of the commodity at the beginning of each period (possibly with capacity constraint on the size of the order), a per-unit *holding cost* for carrying excess inventory from the end of a period to the next period and a per-unit *backlogging cost*, which is a penalty incurred at the end of each period for each unit of unsatisfied demand. The goal is to find an ordering policy that minimizes the overall expected cost over the entire planning horizon.

A stochastic dynamic programming framework has been the most dominant paradigm in the research on these models. It has been used to analyze and characterize the structure of optimal policies, as well as to provide an algorithmic framework to compute optimal policies. The inventory models described above demonstrate the strengths and the limitations of the stochastic dynamic programming framework. That is, these models admit dynamic programming formulations that can be either *easy or hard* to solve, depending on the specific assumptions on the structure of the demands $D_1, \ldots, D_T$ and on the way their distributions are specified for us. By easy and hard to solve, we are distinguishing between formulations and models for which there exist efficient algorithms for computing the optimal policies (i.e., for solving the corresponding dynamic programs), and formulations and models for which it is very unlikely to find such efficient algorithms.

Unfortunately, getting easy models with compact and tractable formulations usually requires strong assumptions both on the structure of the demands (e.g., that demands in different periods are independent random variables) and on the way the demand distributions are specified (e.g., that the distributions are specified explicitly). Moreover, these strong assumptions are unrealistic in most real-life scenarios. In this work, we address the harder and more realistic models that typically give rise to huge dynamic programs that are unlikely to be tractable, both theoretically and in practice. We develop efficient algorithms that approximately solve these dynamic programs and compute policies that are not necessarily optimal but deliver solutions that are provably near-optimal. In particular, we analyze the expected performance of the proposed policies and show that their expected cost is guaranteed to be near the optimal expected cost regardless of the specific instance being solved. Our policies are also easy to implement in practice and we believe they can actually be used in real-life applications. To put this work in context, next we shall discuss how the inventory models described above are formulated as dynamic programs and demonstrate what distinguishes between easy and hard classes of models and formulations.

For the clarity of this discussion, we will demonstrate the above issues in the context of *uncapacitated* models, where at the beginning of each period t we can order any number of units. For ease of notation, assume that the per-unit ordering cost is equal to 0 in all periods and that the per-unit holding and backlogging costs are equal $h \geq 0$ and $p \geq 0$, respectively. Similarly, we will assume that an order placed in period t arrives instantaneously. All of the latter assumptions do not change the models conceptually.

In particular, as discussed in Chapters 2 and 3 below, almost all of the results in this work are valid for models with time-dependent cost parameters (including positive per-unit ordering cost parameters) and with existing *lead times*, i.e., models where an order placed in period t arrives only after several periods.

First consider the *single-period newsvendor problem*, where $T = 1$. A random demand D_1 for the commodity will occur in a single period. At the beginning of the period, before we observe the actual demand, we order y units in our attempt to satisfy the demand. Then the actual demand d_1 (the realization of the random demand D_1) is observed and is satisfied to the maximum extent possible from the available supply units. Next, a cost is incurred. Specifically, we incur, a per-unit holding cost $h \geq 0$ for each unused unit of the commodity, for an overall cost of $h(y-d_1)^+$ (where $x^+ = \max(0, x)$) and a per-unit backlogging cost $p \geq 0$ for each unit of unsatisfied demand, for an overall cost of $p(d_1 - y)^+$. The goal is to find an ordering level y with minimum expected total cost

$$C(y) = E[h(y - D_1)^+ + p(D_1 - y)^+],$$

where the expectation is taken with respect to D_1. It is readily verified that $C(y)$ is convex in y and attains a minimum. Thus, if the demand distribution D_1 is known explicitly (e.g., the CDF function of D_1 is given explicitly), it is straightforward to find a minimizer of $C(y)$. As we will show next, the convexity of the single-period newsvendor cost function is the basic property that enables us to formulate multiperiod models as convex dynamic programs.

Now consider a two-period problem ($T = 2$) and assume that D_1 and D_2 are two independent random variables. Let X_t (for $t = 1, 2$) be the inventory level at the beginning of period t *before ordering*. If $X_t > 0$, this corresponds to excess inventory left from the previous period and if $X_t < 0$, this corresponds to unsatisfied units of demand that are waiting in the system until we will able to satisfy them. We assume that $X_1 = x_1$, where x_1 is the starting inventory level given as part of the input. Without loss of generality, assume $x_1 = 0$. Let Y_t be the inventory level in period t *after ordering* and *before* the demand occurs. Since we can order only non-negative quantities, it is clear that $Y_t \geq X_t$. As in the single-period case above, we can express the expected cost incurred at the end of period t as $C_t(Y_T) = E[h(Y_t - D_t)^+ + p(D_t - Y_t)^+]$. However, even though the demands D_1 and D_2 are independent of each other, this problem does not decompose into two independent single-period newsvendor problems. Specifically, the decision to order up to y_1 (the realization of Y_1) in the first period affects the starting inventory at the beginning of period 2, in that $X_2 = y_1 - D_1$. Since $Y_2 \geq X_2$ this constrains the decision in the second period, and hence, affects the expected cost in that period. Let $V_2(x_2)$ be the minimum expected cost in the second period, given that the starting inventory at the beginning of period 2 is x_2. That is, $V_2(x_2) = \min_{y_2 \geq x_2} C_2(y_2)$. The overall cost as a function of the first period decision y_1 can be expressed as $U_1(y_1) = C_1(y_1) + E[V_2(y_1 - D_1)]$. Note that U_1 consists of two parts, the expected period cost C_1 and the expected future cost $E[V_2(y_1 - D_1)]$ as a function of the decision made in period 1. This captures the fact that given the decision y_1, the inventory level at the beginning of the period 2 is the random variable $X_2 = y_1 - D_1$. Focus now on the function $V_2(x_2)$ above. We have already observed that $C_2(y_2)$ is convex and attains a minimum. Let R_2 denote its minimizer (without loss of generality, assume the minimizer is unique). It is now readily verified that if $x_2 \leq R_2$, then $V_2(x_2) = C_2(R_2)$ (i.e., the minimizer of C_2 is reachable) and if $x_2 > R_2$, then $V_2(x_2) = C_2(x_2)$. This gives rise to what is called *a base-stock policy* in period 2. That is, if the inventory level at the beginning of the period 2 is below R_2, the best policy is to order up-to R_2, otherwise the best policy is to do nothing. Observe that this provides a compact description of the optimal policy in period 2, which is independent of any decision made in period 1. Specifically, R_2 is the minimizer of C_2 which is independent of y_1 and C_1. Thus, if the demand distribution D_2 is known explicitly, R_2 can be computed as a single-period newsvendor minimizer. Suppose that R_2 has been already computed; we can then express the overall cost as a function of the first period decision y_1 as

$$U_1(y_1) = C_1(y_1) + E[\mathbb{1}(y_1 - D_1 \leq R_2)C_2(R_2) + \mathbb{1}(y_1 - D_1 > R_2)C_2(y_1 - D_1)].$$

It is readily verified that $U_1(y_1)$ is the sum of two convex functions, and hence it is a convex function of y_1 that attains a minimum. The function $U_1(y_1)$ accounts for the expected cost of a policy that orders up to y_1 in the first period and then follows the base-stock policy R_2 in the second period. In period 1, we wish to solve $V_1(x_1) = \min_{y_1 \geq x_1} U_1(y_1)$. Since U_1 is convex and attains a minimum, its minimizer, denoted by R_1, gives rise to a base-stock policy R_1, R_2 over the entire horizon. Note that the optimality of R_1 in the first period does depend on the fact that R_2 is followed in the second period. Moreover, if we have already computed R_2 and if the demand distributions are specified explicitly, then the function $U_1(y_1)$ can be evaluated and it is usually straightforward to compute R_1.

This can easily be generalized to T periods. For some $t < T$ assume that we have already computed optimal base-stock levels $R_{t+1}, \ldots, R_T$. Let now $U_t(y_t)$ be the expected cost of a policy that orders up to y_t in period t and then follows the base-stock policy $R_{t+1}, \ldots, R_T$ over $[t+1, T]$. Let $V_t(x_t)$ be the minimum expected cost over $[t, T]$ if the starting inventory at the beginning of period t is x_t and the base-stock policy $R_{t+1}, \ldots, R_T$ is followed over $[t+1, T]$. Specifically,

$$U_t(y_t) = C_t(y_t) + E[V_{t+1}(y_t - D_t)].$$

and

$$V_t(x_t) = \min_{y_t \geq x_t} U_t(y_t).$$

By arguments similar to the ones above, it is readily verified that $U_t(y_t)$ is convex and attains a minimum. Thus, the optimal policy in period t can be described based on its minimizer R_t. That is, if $x_t < R_t$, then we order up-to R_t, otherwise we do nothing. Assuming that $R_{t+1}, \ldots, R_T$, the respective minimizers of $U_{t+1}, \ldots, U_T$ (where $U_T = C_T$), have been already computed, and that the demand distributions are given explicitly, it is again straightforward to compute R_t. We conclude that under the assumptions that the demands in different periods are independent of each other and that the demand distributions are given explicitly, it is relatively straightforward to solve the corresponding dynamic program and find an optimal policy. In particular, we need to solve recursively T subproblems, each of which consists of a single-variable convex minimization problem.

Note that the assumption that the demands in different periods are independent of each other was crucial to get a compact, tractable formulation. This assumption implies that the impact of the evolution of demands and the control policy over the periods $1, \ldots, t-1$ on the control policy over the periods $t, \ldots, T$ is limited only to the inventory level at the beginning of period t, namely X_t. In particular, they make no impact on the way we see the distributions of demands $D_t, \ldots, D_T$ from time t. However, the structure of base-stock policies implies that the base-stock levels in periods over $t, \ldots, T$ can be computed only based on the demand distributions $D_t, \ldots, D_T$ as seen from period t. This gives rise to a very compact dynamic program, with a singletone state-space in each stage. Another essential assumption is that the demand distributions are given explicitly, and that, in each stage, we can compute the minimizer of the function U_t *exactly*. It is not hard to see, that the convexity of the function U_t and the optimality of a base-stock policy in period t depend heavily on the assumption that $R_{t+1}, \ldots, R_T$ are the exact minimizers of the functions $U_{t+1}, \ldots, U_T$, respectively.

Unfortunately, the two assumptions above are not realistic in many if not most real-life scenarios. In many applications the demand in different periods are heavily correlated, that is, the demands we observe in period $1, \ldots, t-1$ do change the conditional distributions of the demands $D_t, \ldots, D_T$. In addition, in most scenarios, the explicit demand distributions are not given. Instead, we usually have access only to *samples* from the true demand distributions through historical data or simulation.

Consider the multi period model discussed above, but under the assumption that the demands in different periods are allowed to have a complex structure, including inter temporal correlation. The dynamic program described above can be extended to capture this more realistic model. However, the dependence between demands in different periods requires a more complex representation of the evolution of the model. Specifically, we let f_t to be the observed *information set* at the beginning of period t. The object f_t consists of the realized demands $d_1, \ldots, d_{t-1}$ (the realizations of $D_1, \ldots, D_{t-1}$, respectively) in past periods, and possibly more information that has become available to us by time t, denoted by $w_1, \ldots, w_{t-1}$. The information set f_t is a specific realization of the random vector $F_t = (D_1, \ldots, D_{t-1}, W_1, \ldots, W_{t-1})$. We let $\mathcal{F}_t$ denote the set of all possible realizations of the random vector F_t. Moreover, since demands in different periods are possibly correlated, each observed information set implies a corresponding joint conditional distribution $I_t = I_t(f_t)$ (as seen from period t) of the future demands $D_t, \ldots, D_T$. Thus, the previous function U_t above is now defined with respect to each pair (t, f_t) as $U_{t,f_t}(y_t)$, where again $f_t \in \mathcal{F}_t$ is a specific observed information set in period t. Similarly, instead of $C_t(y_t)$ and $V_t(x_t)$ we now have C_{t,f_t} and $V_{t,f_t}(x_t)$, respectively, for each pair (t, f_t). By a proof similar to the one outlined above, it is straightforward to show that a *state-dependent base-stock policy* is optimal [56]. That is, the optimal policy can be described through a set of target inventory levels $\bigcup_{t=1}^{T}\{R(f_t)\}_{f_t\in\mathcal{F}_t}$, where again, $R(f_t)$ is the respective minimizer of the convex function $U_{t,f_t}(y_t)$. The function $U_{t,f_t}(y_t)$ now accounts for the expected cost of a policy that in period t with information set f_t orders up-to y_t and then follows the optimal base-stock policy. The recursion of the dynamic program is now

$$U_{t,f_t}(y_t) = C_{t,f_t}(y_1) + E[V_{t+1,F_{t+1}}(y_t - D_t)|f_t].$$

It is readily verified that the size of this dynamic programs grows at least linearly in the cardinality of the sets $\mathcal{F}_t$ (for $t = 1, \ldots, T$). In order to evaluate $E[V_{t+1,F_{t+1}}(y_t - D_t)|f_t]$ above, we need to know $V_{t+1,f_{t+1}}$ for each $f_{t+1} \in F_{t+1}$, i.e., to solve many subproblems. As a result, the corresponding dynamic program quickly becomes intractable. This phenomenon is known as the *curse of dimensionality*.

In Chapter 2 below, we address the more realistic setting of inventory models with more complex structures of demands that allow general correlation between demands in different periods. Rather than solve the huge dynamic programs that arise in this context, we consider the performance of simple heuristics that can be implemented online. Specifically, the control policy of the heuristics in each period does not require any knowledge of the control policy in future periods. Thus, we do not need to maintain huge dynamic programs. Our approach is based on two key ideas. First, we introduce a new cost accounting approach that we call *marginal cost accounting*. This approach significantly departs from the standard dynamic programming cost accounting approach described above. In particular, the standard dynamic programming approach directly associates with the decision made in period t of how many units to order, only the expected cost incurred in period t. Note however, that part of the cost incurred in period t could not be avoided by any decision made in this period, and that the decision made in period t might incur additional cost in future periods. In turn, in the marginal cost accounting approach introduced in Chapter 2, we associate with the decision made in period t *all* of the expected cost that, after this decision is made, become independent of any decision made in future periods. Secondly, we use general *balancing techniques* that, in each period, balance the marginal expected holding cost against the marginal expected backlogging cost associated with the decision made in the current period. The intuition is that we balance between costs incurred by over-ordering and costs incurred by under-ordering. This simple idea appears to be powerful in constructing algorithms that guarantee provably good expected performance guarantees. Specifically, we provide 2-approximation algorithms for the general uncapacitated and capacitated variants of the multiperiod model. That is, for each possible input of the problem, our policies are guaranteed to have expected cost at most twice the optimal expected cost. The new class of balancing algorithms are relatively straightforward to implement both computationally and conceptually. We believe, that our new marginal cost accounting approach and the balancing algorithms will have application in many other supply chain-related models and possibly in other domains that admit huge stochastic dynamic programs.

In Chapter 3 below, we consider models with simple structure of demands, i.e., models where the demands in different periods are independent of each other (not necessarily identically distributed), but under the assumption that the explicit demand distributions are not given. Instead, we assume that the only information about the true underlying demand distributions is available through a black box that, on request, can generate independent samples from the true demand distributions. This corresponds to having available historical data or to a simulation setting. We propose *sample-based policies*, that is, policies that are computed efficiently based only on samples without any access to the explicit true demand distributions. Moreover, we provide a novel analysis of the number of samples required to guarantee that our sample-based policies will have, with high probability, expected cost arbitrarily close to the optimal expected cost in the case we allow *full* access to the demand distributions. The bounds that we establish on the number of required samples are general, easy to compute and surprisingly do not depend on the specific underlying demand distributions. We believe that this approach sets the foundations for new classes of sample-based algorithms for stochastic dynamic programs with analyzed worst-case performance.

Chapter 2

Near-Optimal Balancing Policies for Stochastic Inventory Control Models

2.1 Introduction

In this chapter, we address the fundamental problem of finding computationally efficient and provably good inventory control policies in supply chains in which the demands are stochastic, correlated and non-stationary (time-dependent). This problem arises in many domains and has many practical applications (see, for example, [11, 29]). We consider several classical models, the *periodic-review stochastic inventory control problem* with and without capacity constrains, and the *stochastic lot-sizing problem*. All the models are considered with correlated and non-stationary demands. Here the correlation is inter-temporal, i.e., what we observe in period s changes our forecast for the demand in future periods. We provide what we believe to be the first computationally efficient policies with constant worst-case performance guarantees; that is, there exists a constant C such that, for any instance of the problem, the expected cost of the policy is at most C times the expected cost of an optimal policy.

A major domain of applications in which demand correlation and non-stationarity are commonly observed in supply chains is where dynamic demand forecasts are employed. Demand forecasts often serve as an essential managerial tool, especially when the demand environment is highly dynamic. How these demand forecasts, that evolve over time, can be used to devise an efficient and cost-effective inventory control policy is of great interest to managers, and has attracted the attention of many researchers over the years (see, for example, [22, 31, 37]). However, it is well known that such environments often induce high correlation between demands in different periods that makes it very hard to compute optimal inventory control policies, and in many cases, even computing a 'good' policy is rather a hard task. Another relevant and important domain of applications is for new products and/or new markets. These scenarios are often accompanied by an intensive promotion campaign and involve many uncertainties, which create high levels of correlation and non-stationarity in the demands over time. Correlation and non-stationarity also arise for products with strong cyclic demand patterns, and as products being phased out of the market.

The stochastic inventory control models considered here, capture many if not most of the application domains in which correlation and non-stationarity arise. Specifically, we consider single-item models with one location and a finite planning horizon of T discrete time periods. The demands over the T periods are random variables that can be non-stationary and correlated. In the periodic-review stochastic inventory control problem, the cost consists of per-unit, time-dependent ordering cost, holding cost for carrying excess inventory from period to period and backlogging cost, which is a penalty incurred, in each period, for each unit of unsatisfied demand (where all shortages are fully backlogged). In the *uncapacitated model* we can order at the beginning of each period any number of units of the commodity. In the *capacitated model* there is a capacity constraint on the number of units ordered in each period. There is also a lead time between the time an order is placed and the time that it actually arrives. In the stochastic lot-sizing problem, we consider, in addition, a fixed ordering cost that is incurred in each period in which an order is placed (regardless of its size), but with no lead time. In all the models, the goal is to find a policy of orders with minimum expected overall cost over the given planning horizon. The assumptions that we make on the demand distributions are very mild and generalize all of the currently known approaches in the literature to model correlation and non-stationarity of demands over time. This includes classical approaches like the *martingale model of forecast evolution model* (MMFE), exogenous Markovian demand, time series, order-one auto-regressive demand and random walks. For an overview of the different approaches and models, and for relevant references, we refer the reader to [22, 9]. Moreover, we believe that the models we consider are general enough to capture almost any other reasonable way of modelling correlation and non-stationarity of demands over time.

These models have attracted the attention of many researchers over the years and there exists a huge body of related literature. The dominant paradigm in almost all of the existing literature has been to formulate these models using a dynamic programming framework. The optimization problem is defined recursively over time using subproblems for each possible state of the system. The state usually consists of a given time period, the level of the echelon inventory at the beginning of the period, a resulted conditional distribution on the future demands over the rest of the horizon, and possibly more information that is available by that period. For each subproblem, we compute an optimal solution to minimize the expected overall discounted cost from time t until the end of the horizon.

This framework has turned out to be very effective in characterizing the optimal policy of the overall system. Surprisingly, the optimal policies for these rather complex models follow simple forms. In the models with only per-unit ordering cost and no capacity constraints, the optimal policy is a *state-dependent base-stock policy*. In each period, there exists an optimal target base-stock level that is determined only by the given conditional distribution (at that period) of future demands and possibly by additional information that is available, but it is independent of the starting inventory level at the beginning of the period, that is, independent of the control policy in past periods.

The optimal policy aims to keep the inventory level in each period as close as possible to the target base-stock level. That is, it orders up to the target level whenever the inventory level at the beginning of the period is below that level, and orders nothing otherwise. The optimality of base-stock policies has been proven in many settings, including models with correlated demand and forecast evolution (see, for example, [22, 36]). Similarly, in the presence of capacity constraints on the size of the order in each period, the optimal policies are *state-dependent modified base-stock policies*. As before, in each period, there exists a target inventory level, and the optimal policy aims to keep its inventory level as close as possible to that target level. However, in the case where the inventory level at the beginning of the period is below that target level, it might not be possible to order up to the target level because of the capacity constraint. In this case the order placed would be up to capacity. There are several proofs of the optimality of modified base-stock policies in different settings (with independent demands) that are based on dynamic programming approach (see, for example, [12, 26, 1]).

For the models with fixed ordering cost, the optimal policy follows a slightly more complicated pattern. Now, in each period, there are lower and upper thresholds that are again determined only by the given conditional distribution (at that period) on future demands. Following an optimal policy, an order is placed in a certain period if and only if the inventory level at the beginning of the period has dropped below the lower threshold. Once an order is placed, the inventory level is increased up to the upper threshold. This class of policies is usually called *state-dependent* (s, S) *policies*. The optimality of state-dependent (s, S) policies was proven for the case of non-stationary but independent demand (see [56]). Gallego and Özer [14] have established their optimality for a model with correlated demands. We refer the reader to [9, 22, 56, 14] for the details on some of the results along these lines, as well as a comprehensive discussion of relevant literature.

Unfortunately, the rather simple forms of these policies do not always lead to efficient algorithms for computing the optimal policies. The corresponding dynamic programs are relatively straightforward to solve if the demands in different periods are independent. Dynamic programming approach can still be tractable for uncapcitated models with Markov-modulated demand but under rather strong assumptions on the structure and the size of the state space of the underlying Markov process (see, for example, [52, 7]). However, in many scenarios with more complex demand structure the state space of the corresponding dynamic programs grows exponentially and explodes very fast (see [22, 9] for relevant discussions on the MMFE model). Capacitated models are even harder to solve using a dynamic programming approach. The difficulty again comes from the fact that we need to solve 'too many' subproblems. This phenomenon is known as *the curse of dimensionality*. Moreover, because of this phenomenon, it seems unlikely that there exists an efficient algorithm to solve these huge dynamic programs. This gap between the knowledge on the structure of the optimal policies and the inability to compute them efficiently provides the stimulus for future theoretical interest in these problems.

For the uncapacitated periodic-review stochastic inventory control problem, Muharremoglu and Tsitsiklis [35] have proposed an alternative approach to the dynamic programming framework. They have observed that this problem can be decoupled into a series of *unit supply-demand subproblems*, where each subproblem corresponds to a single unit of supply and a single unit of demand that are matched together. This novel approach enabled them to substantially simplify some of the dynamic programming based proofs on the structure of optimal policies, as well as to prove several important new structural results. In particular, they have established the optimality of state-dependent base-stock policies for the uncapacitated model with general Markov-modulated demand. Using this unit decomposition, they have also suggested new methods to compute the optimal policies. However, their computational methods are essentially dynamic programming approaches applied to the unit subproblems, hence they suffer from similar problems in the presence of correlated and non-stationary demand. Although our approach is very different from theirs, we use some of their ideas as technical tools in some of the proofs in this chapter. Janakriman and Muckstadt [24] have extended this approach to capacitated models and established the optimality of state-dependent modified base-stock policies for models with Markov-modulated demand.

As a result of this apparent computational intractability, many researchers have attempted to construct computationally efficient (but suboptimal) heuristics for these problems. However, we are aware of very few attempts to analyze the worst-case performance of these heuristics (see for example [31]). Moreover, we are aware of no computationally efficient policies for which there exist constant performance guarantees. For details on some of the proposed heuristics, and a discussion of others, see [9, 31, 22, 16, 27].

One specific class of suboptimal policies for the uncapacitated model that has attracted a lot of attention is the class of *myopic policies*. In a myopic policy, in each period, we attempt to minimize the expected cost for that period, ignoring the potential effect on the cost in future periods. The myopic policy is attractive since it yields a base-stock policy that is easy to compute on-line, that is, it does not require information on the control policy in the future periods. In each period, we need to solve a single-variable convex minimization problem. In many cases, the myopic policy seems to perform well. However, in many other cases, especially when the demand can drop significantly from period to period, the myopic policy performs poorly. Veinott [55] and

Ignall and Veinott [21] have shown that myopic policy can be optimal even in models with nonstationary demand as long as the demands are stochastically increasing over time. Iida and Zipkin [22] and Lu, Song and Regan [31] have focused on the martingale model of forecast evolution (MMFE) and shown necessary conditions and rather strong sufficient conditions for myopic policies to be optimal. They have also used myopic policies to compute upper and lower bounds on the optimal base-stock levels, as well as bounds on the relative difference between the optimal cost and the cost of different heuristics. However, the bounds they provide on this relative error are not constants. For the capacitated model, it is known that myopic policies often perform very badly since they do not consider possible capacity limitations in future periods.

Chan and Muckstadt [5] have considered a different way for approximating huge dynamic programs that arise in the context of inventory control problems. More specifically, they have considered uncapacitated and capacitated multi-item models. Instead of solving the one period problem (as in the myopic policy) they have added to the one period problem a penalty function which they call Q-function. This function accounts for the holding cost incurred by the inventory left at the end of the period over the entire horizon. Their look ahead approach with respect to the holding cost is somewhat related to our approach, though significantly different.

We note that our work is also related to a huge body of approximation results for stochastic and on-line combinatorial problems. The work on approximation results for stochastic combinatorial problems goes back to the work of Möhring, Radermacher and Weiss [32, 33] and the more recent work of Möohring, Schulz and Uetz [34]. They have considered stochastic scheduling problems. However, their performance guarantees are dependent on the specific distributions (namely on second moment information). Recently, there is a growing stream of approximation results for several 2-stage stochastic combinatorial problems. For a comprehensive literature review we refer the reader to [53, 10, 49, 6]. We note that the problems we consider in this work are by nature *multi-stage* stochastic problems, which are usually much harder (see [8] for a recent result on the stochastic knapsack problem).

Another approach that was applied to these models is the robust optimization approach (see [2]). Here the assumption is of a distribution-free model, where instead the demands are assumed to be drawn from some specified uncertainty set. Each policy is then evaluated with respect to the worst possible sequence of demands within the given uncertainty set. The goal is to find the policy with the best worst-case (i.e., a min-max approach). This objective is very different from the objective of minimizing expected (average cost) discussed in most of the existing literature, including this work.

Our work is distinct from the existing literature in several significant ways, and is based on three key ideas:

Marginal cost accounting scheme. We introduce a novel approach for cost accounting in stochastic inventory control problems. The standard dynamic programming approach directly assigns to the decision of how many units to order in each period only the expected holding and backlogging costs incurred in that period although this decision might effect the costs in future periods. Instead, our new cost accounting scheme assigns to the decision in each period *all* the expected costs that, once this decision is made, become independent of any decision made in future periods, and are dependent only on the future demands. The marginal holding cost accounting approach is based on the key observation that once we place an order for a certain number of units in some period, then the expected ordering and holding cost that these units are going to incur over the rest of the planning horizon is a function only of the realized demands over the rest of the horizon, not of future orders. Hence, with each period, we can associate the overall expected ordering and holding cost that is incurred by the units ordered in this period, over the entire horizon. We note that similar ideas of holding cost accounting were previously used in the context of models with continuous time, infinite horizon and stationary (Poisson distributed) demand (see, for example, the work of Axsäter and Lundell [43] and Axsäter [42]). In an uncapacitated model the decision of how many units to order in each period effect the expected backlogging cost in only a single future period, namely, a lead time ahead. However, this is not necessarily true in a capacitated model, where this decision might effect the expected backlogging cost in several periods into the future. Thus, for capacitated models we introduce a marginal backlogging cost accounting approach. Suppose that in the current period the order placed was not up to capacity, we wish to account for the potential backlogging cost in future periods incurred directly by the decision not to use the full available capacity. Assume temporarily that we order up to capacity in each one of the periods. Suppose now that in the current period we do not order up to capacity. Then expected marginal backlogging cost associated with the current period is the overall increase in the expected backlogging cost over the entire horizon resulting from this decision. The marginal backlogging cost accounting scheme for the capacitated model is in fact a generalization of the traditional period backlogging cost accounting scheme. As we will show it turns out that both the expected marginal holding and backlogging costs are straightforward to compute in most common scenarios. We believe that this new approach will have more applications in the future in analyzing stochastic inventory control problems.

Cost balancing. The idea of cost balancing was used in the past to construct heuristics with constant performance guarantees for deterministic inventory problems. The most well-known examples are the Silver-Meal Part-Period balancing heuristic for the lot-sizing problem (see [51]) and the Cost-Covering heuristic of Joneja for the joint-replenishment problem [25]. We are not aware of any application of these ideas to stochastic inventory control problems. The key observation is that any policy in any period incurs potential expected costs due to overordering (namely, expected holding costs) and underordering (namely, expected backlogging costs). For the periodic-review stochastic inventory control problem (both uncapacitated and capacitated variants), we use the marginal cost accounting approach to construct policies that, in each period, balance the expected (marginal) ordering and holding cost against the expected (marginal) backlogging cost. For the stochastic lot-sizing problem, we construct a policy that balances the expected fixed ordering cost, holding cost and backlogging cost over each interval between consecutive orders. As we shall show, the simple idea of balancing is powerful and leads to policies that have constant worst-case performance guarantees. We again believe that the balancing idea will have more applications in constructing and analyzing algorithms for other stochastic inventory control models.

Non base-stock policies. Our policies are not state-dependent base-stock policies, in that the order up-to level order of the policy in each period does depend on the inventory control in past periods. However, this enable us to use, in each period, the distributional information about the future demands beyond the current period (unlike the myopic policy), without the burden of solving huge dynamic programs. Moreover, our policies can be easily implemented on-line (like the myopic policy) and are simple, both conceptually and computationally (see [23]).

Using these ideas we provide what is called a 2-approximation algorithm for the uncapacitated and capacitated variants of the periodic-review stochastic inventory control problem; that is, the expected cost of our policies is no more than twice the expected cost of an optimal policy. Note that this is not the same requirement as stipulating that, for each realization of the demands, the cost of our policy is at most twice the optimal cost, which is a much more stringent requirement. We also note that these guarantees refer only to the worst-case performance and it is likely that the typical performance would be significantly better (see [23]). We then use a standard cost transformation to achieve significantly better guarantees if the ordering cost is the dominant part in the overall cost, as it is the case in many real life situations. Our results are valid for all known approaches used to model correlated and non-stationary demands. We note that the analysis of the worst-case performance is tight. In particular, we describe a family of examples for which the ratio between the expected cost of the balancing policy and the expected cost of the optimal policy is asymptotically 2. For the uncapacitated periodic-review stochastic inventory control problem, we also present an extended class of myopic policies that provides easily computed upper bounds and lower bounds on the optimal base-stock levels. As shown in [23], these bounds combined with the balancing techniques lead to improved balancing policies. These policies have a worst-case performance guarantee of 2 and they seem to perform significantly better in practice. We establish similar bounds for the capacitated model, and show again how to use them to get improved policies.

An interesting question that is left open in the current literature is whether the myopic policy for the uncapacitated model has a constant worst-case performance guarantee. We provide a negative answer to this question, by showing a family of examples for which the expected cost of the myopic policy can be arbitrarily more expensive than the expected cost of an optimal policy. Our example provides additional insight into situations in which the myopic policy performs poorly.

For the stochastic lot-sizing problem we provide a 3-approximation algorithm. This is again a worst-case analysis and we would expect the typical performance to be much better.

The rest of this chapter is organized as follows. In Section 2.2 we present a mathematical formulation of the periodic-review stochastic inventory control problem. Then in Section 2.3 we explain the details of the new marginal holding cost accounting approach followed by section 2.4 in which we describe the balancing policy for the uncapacitated periodic-review stochastic inventory control problem and present its worst-case analysis. Then in Section 2.5 we describe the marginal backlogging cost accounting for the capacitated model followed by Section 2.6 in which we describe the balancing policy and its worst-case analysis for the capacitated model. Several important extensions are discussed in Section 2.7. In Section 2.8 we present an extended class of myopic policies for the uncapacitated model, develop upper and lower bounds on the optimal base-stock levels, and discuss the example in which the performance of the myopic policy is arbitrarily bad. Similarly, in Section 2.9, we consider the capacitated case, develop lower and upper bounds on the optimal inventory levels, and show how to use them to get improved policies. The stochastic lot-sizing problem is discussed in Section 2.10, where we present a 3-approximation algorithm for the problem. We then conclude the chapter with some remarks and open research questions.

2.2 Periodic-Review Stochastic Inventory Control Problem

In this section, we provide the mathematical formulation of the periodic-review stochastic inventory problem and introduce some of the notation used in the coming sections of this chapter. As a general convention throughout this chapter, we distinguish between a random variable and its realization using capital letters and lower case letters, respectively. Script font is used to denote sets. We consider a finite planning horizon of T periods numbered $t = 1, \dots, T$ (note that t and T are both deterministic unlike the convention above). The demands over these periods are random variables, denoted by

$D_1, \ldots, D_T$.

As part of the model, we assume that at the beginning of each period s, we are given what we call an *information set* that is denoted by f_s. The information set f_s contains all of the information that is available at the beginning of time period s. More specifically, the information set f_s consists of the realized demands $(d_1, \ldots, d_{s-1})$ over the interval $[1, s)$, and possibly some more (external) information denoted by $(w_1, \ldots, w_s)$. The information set f_s in period s is one specific realization in the set of all possible realizations of the random vector $F_s = (D_1, \ldots, D_{s-1}, W_1, \ldots, W_s)$. This set is denoted by $\mathcal{F}_s$. In addition, we assume that in each period s, there is a known conditional joint distribution of the future demands $(D_s, \ldots, D_T)$, denoted by $I_s := I_s(f_s)$, which is determined by f_s (i.e., knowing f_s, we also know $I_s(f_s)$). For ease of notation, D_t will always denote the random demand in period t according to the conditional joint distribution I_s for some $s \leq t$, where it will be clear from the context to which period s we refer. We will use t as the general index for time, and s will always refer to the current period.

The only assumption on the demands is that for each $s = 1, \ldots, T$, and each $f_s \in \mathcal{F}_s$, the conditional expectation $E[D_t|f_s]$ is well defined and finite for each period $t \geq s$. In particular, we allow non-stationarity and correlation between the demands in different periods. We note again that by allowing correlation we let I_s be dependent on the realization of the demands over the periods $1, \ldots, s-1$ and possibly on some other information that becomes available by time s (i.e., I_s is a function of f_s). However, the information set f_s as well as the conditional joint distribution I_s are assumed to be independent of the specific inventory control policy being considered.

In the periodic-review stochastic inventory control problem, our goal is to supply each unit of demand while attempting to avoid ordering it either too early or too late. In period t, ($t = 1, \dots, T$) three types of costs are incurred, a per-unit ordering cost c_t for ordering up to u_t units, where $u_t \geq 0$ is the available order capacity in period t (in the uncapacitated model we assume $u_t = \infty$ for each $t = 1, \dots, T$), a per-unit holding cost h_t for holding excess inventory from period t to $t + 1$, and a unit backlogging penalty p_t that is incurred for each unsatisfied unit of demand at the end of period t. Unsatisfied units of demand are usually called *backorders*. The assumption is that backorders fully accumulate over time until they are satisfied. That is, each unit of unsatisfied demand will stay in the system and will incur a backlogging penalty in each period until it is satisfied. In addition, we consider a model with a lead time of L periods between the time an order is placed and the time at which it actually arrives. We first assume that the lead time is a known integer L. In Section 2.7, we will show that our policy can be modified to handle stochastic lead times under the assumption of no order crossing (i.e., any order arrives no later than orders placed later in time).

There is also a discount factor $\alpha \leq 1$. The cost incurred in period t is discounted by a factor of α^t. Since the horizon is finite and the cost parameters are time-dependent, we can assume without loss of generality that $\alpha = 1$. We also assume that there are no speculative motivations for holding inventory or having back orders in the system. To enforce this, we assume that, for each $t = 2, \dots, T - L$, the inequalities $c_t \leq c_{t-1} + h_{t+L-1}$ and $c_t \leq c_{t+1} + p_{t+L}$ are maintained (where $C_{T+1} = 0$). (In case there is a discount factor, we require that $\alpha c_t \leq c_{t-1} + \alpha^L h_{t+L-1}$ and $c_t \leq \alpha c_{t+1} + \alpha^L p_{t+L}$.) We also assume that the parameters $h_t,\ p_t$ and c_t are all non-negative. Note that the parameters h_T and p_T can be defined to take care of excess inventory and back orders at the end of the planning horizon. In particular, p_T can be set to be high enough to ensure that there are very few back orders at the end of time period T.

The goal is to find a feasible ordering policy (i.e., one that respects the capacity constraints) that minimizes the overall expected discounted ordering cost, holding cost and backlogging cost. We consider only policies that are *non-anticipatory*, i.e., at time s, the information that a feasible policy can use consists only of f_s and the current inventory level. In particular, given any feasible policy P and conditioning on a specific information set f_s, we know the inventory level x_s^P deterministically.

We will use $D_{[s,t]}$ to denote the accumulated demand over the interval $[s,t]$, i.e., $D_{[s,t]} := \sum_{j=s}^{t} D_j$. We will also use superscripts P and OPT to refer to a given policy P and the optimal policy respectively.

Given a feasible policy P, we describe the dynamics of the system using the following terminology. We let NI_t denote the *net inventory* at the end of period t, which can be either positive (in the presence of physical on-hand inventory) or negative (in the presence of back orders). Since we consider a lead time of L periods, we also consider the orders that are on the way. The sum of the units included in these orders, added to the current net inventory is referred to as the *inventory position* of the system. We let X_t be the inventory position at the beginning of period t *before* the order in period t is placed, i.e., $X_t := NI_{t-1} + \sum_{j=t-L}^{t-1} Q_j$ (for $t = 1, \ldots, T$), where Q_j denotes the number of units ordered in period j (we will sometime denote $\sum_{j=t-L}^{t-1} Q_j$ by $Q_{[t-L,t-1]}$). Similarly, we let Y_t be the inventory position *after* the order in period t is placed, i.e., $Y_t = X_t + Q_t$. Note that once we know the policy P and the information set $f_s \in \mathcal{F}_s$, we can easily compute ni_{s-1}, x_s and y_s, where again these are the realizations of NI_{s-1}, X_s and Y_s, respectively.

Since time is discrete, we next specify the sequence of events in each period s:

1. The order placed in period $s - L$ of q_{s-L} units arrives and the net inventory level increases accordingly to $ni_{s-1} + q_{s-L}$.

2. The decision of how many units to order in period s is made. Following a given

policy P, q_s units are ordered, where $0 \leq q_s \leq u_s$, i.e., the order in period s can not exceed the available capacity u_s. Consequently, the inventory position is raised by q_s units (from x_s to y_s). This incurs a linear cost $c_s q_s$.

3. We observe the demand in period s which is realized according to the conditional joint distribution I_s. We also observe the new information set $f_{s+1} \in \mathcal{F}_{s+1}$, and hence we also know the updated conditional joint distribution I_{s+1}. The net inventory and the inventory position each decrease by d_s units. In particular, we have $x_{s+1} = x_s + q_s - d_s$ and $ni_{s+1} = ni_s + q_{s-L} - d_s$.

4. If $ni_{s+1} > 0$, then we incur a holding cost $h_s ni_{s+1}$ (this means that there is excess inventory that needs to be carried to time period $s + 1$). On the other hand, if $ni_{s+1} < 0$ we incur a backlogging penalty $p_t|ni_{s+1}|$ (this means that there are currently unsatisfied units of demand).

2.3 Marginal Holding Cost Accounting

In this section, we present a new approach to the holding cost accounting of stochastic inventory control problems. Our approach differs from the traditional dynamic programming based approach. In particular, we account for the holding cost incurred by a feasible policy in a different way, which enables us to design and analyze new approximation algorithms. We believe that this approach will be useful in other stochastic inventory models.

2.3.1 Dynamic Programming Framework

Traditionally, stochastic inventory control problems of the kind described in Section 2.2 are formulated using a dynamic programming framework. For simplicity, we discuss the case with $L = 0$, where $x_s = ni_s$ (for a detailed discussion see Zipkin [56]).

In a dynamic programming framework, the problem is defined recursively over time through subproblems that are defined for each possible state. A state usually consists of a time period t, an information set $f_t \in \mathcal{F}_t$ and the inventory position at the beginning of period t, denoted by x_t. For each subproblem let $V_t(x_t, f_t)$ be the optimal expected over the interval $[t, T]$ given that the inventory position at the beginning of period t was x_t and the observed information set was f_t. We seek to compute an optimal policy in period t that minimizes the expected cost over $[t, T]$ (i.e., minimizes $V_t(x_t, f_t)$) under the assumption that we are going to make optimal decisions in future periods. The space of feasible decisions consists of all orders of size $0 \leq q_t \leq u_t$ (in the uncapacitated case $u_t = \infty$), or alternatively the level y_t to which the inventory position is raised, where $x_t \leq y_t \leq x_t + u_t$ (and $q_t = y_t - x_t$). Assuming that the optimal policy for all subproblems of states with periods $t+1, \ldots, T$ has been already computed, the dynamic programming formulation for computing the optimal policy for the subproblem of period t is

$$V_t(x_t, f_t) = \min_{x_t \leq y_t \leq x_t + u_t} \{c_t(y_t - x_t) + E[h_t(y_t - D_t)^+ + p_t(D_t - y_t)^+ | f_t] + E[V_{t+1}(y_t - D_t, \mathcal{F}_{t+1}) | f_t]\}.$$

As can be seen the cost of any feasible decision $x_t \leq y_t \leq x_t + u_t$ is divided into two parts. The first part is the *period cost* associated with period t, namely the ordering cost incurred by the order placed in period t and the resulted expected holding cost and backlogging cost in this period, i.e.,

$$c_t(y_t - x_t) + E[h_t(y_t - D_t)^+ + p_t(D_t - y_t)^+ | f_t].$$

In addition, there are the future costs over $[t+1, T]$ (again, assuming that optimal decisions are made in future periods). The impact of the decision in period t on the future costs is captured through the state in the next period, namely $y_t - D_t$. In particular, in a dynamic programming framework, the cost accounted directly in each period t, is only the expected period cost, although the decision made in this period might imply additional costs in the future periods. We note that if $L > 0$, then the period cost is always computed a lead time ahead. That is, the period cost associated with the decision to order up to y_t in period t is

$$c_t(y_t - x_t) + E[h_{t+L}(y_t - D_{[t,t+L]})^+ + p_{t+L}(D_{[t,t+L]} - y_t)^+ | f_t],$$

where $D_{[t,t+L]}$ is the accumulated demand over the lead time.

Dynamic programming approach has turned out to be very effective in characterizing the structure of optimal policies. As was noted in Section 2.1, this yields an optimal base-stock policy, $\{R(f_t) : f_t \in \mathcal{F}_t\}$. Given that the information set at time s is f_s, then the optimal base-stock level is $R(f_s)$. The optimal policy then follows the following pattern. In case the inventory position level at the beginning of period s is lower than $R(f_s)$ (i.e., $x_s < R(f_s)$), then the inventory position is increased to $y_s = R(f_s)$ by placing an order of the appropriate number of units. Suppose that the target level $R(f_s)$ is not reachable because of the capacity constraint, i.e., $x_s + u_s < R(f_s)$. We then set $y_s = x_s + u_s$, i.e., order up to capacity. In the case $x_s \geq R(f_s)$, the inventory position is kept the same (i.e., nothing is ordered) and $y_s = x_s$.

Unfortunately, in scenarios where the demands in different periods are correlated, obtaining the optimal policy using this dynamic programming formulation is likely to be intractable. To compute the optimal policy we need to consider a subproblem for every possible period and possible state of the system. However, the set $\mathcal{F}_s$ can be exponentially large or infinite. This phenomenon is known as the 'curse of dimensionality'.

2.3.2 Marginal Holding Cost Accounting

We take a different approach for accounting for the holding cost associated with each period. Observe that once we decide to order q_s units at time s (where $q_s = y_s - x_s$), then the holding cost they are going to incur from period s until the end of the planning horizon is independent of any future decision in subsequent time periods. It is dependent only on the demand to be realized over the time interval $[s, T]$.

To make this rigorous, we use a *ground distance-numbering scheme* for the units of demand and supply, respectively. More specifically, we think of two infinite lines, each starting at 0, the *demand line* and the *supply line*. The demand line $\mathcal{L}_D$ represents the units of demands that can be potentially realized over the planning horizon, and similarly, the supply line $\mathcal{L}_S$ represents the units of supply that can be ordered over the planning horizon. Each 'unit' of demand, or supply, now has a *distance-number* according to its respective distance from the origin of the demand line and the supply line, respectively. If we allow continuous demand (rather then discrete) and continuous order quantities the unit and its distance-number are defined infinitesimally. We can assume, without loss of generality, that the units of demands are realized according to increasing distance-number. For example, if the accumulated realized demand up to time t is $d_{[1,t)}$ and the realized demand in period t is d_t, we then say that the demand units numbered $(d_{[1,t)}, d_{[1,t)} + d_t]$ were realized in period t. Similarly, we can describe each policy P in terms of the periods in which it orders each supply unit, where all unordered units are "ordered" in period $T + 1$. It is also clear that we can assume without loss of generality that the supply units are ordered in increasing distance-number. Specifically, the supply units that ordered in period t are numbered $(ni_0 + q_{[1-L,t)}, ni_0 + q_{[1-L,t]}]$, where ni_0 and $q_j,\ 1 - L \leq j \leq 0$ are the net inventory and the sequence of the last L orders, respectively, given as an input at the beginning of the planning horizon (in time 0). We further assume (again without loss of generality) that as demand is realized, the units of supply are consumed on a *first-ordered-first-consumed basis*. Therefore, we can *match* each unit of supply that is ordered to a certain unit of demand that has the same number. We note that Muharremoglu and Tsitsiklis [35] have used the idea of matching units of supply to units of demand in a novel way to characterize and compute the optimal policy in different stochastic inventory models. However, their computational methods are based on applying dynamic programming to the single-unit problems. Therefore, their cost accounting within each single-unit problem is still additive, and differs fundamentally from ours.

Suppose now that at the beginning of period s we have observed an information set f_s. Assume that the inventory position is x_s and q_s additional units are ordered. Then the expected additional (marginal) holding cost that these q_s units are going to incur from time period s until the end of the planning horizon is equal to

$$\sum_{j=s+L}^{T} E[h_j(q_s \;-\; (D_{[s,j]} - x_s)^+)^+ | f_s],$$

(recall that we assume without loss of generality that $\alpha = 1$), where $x^+ = \max(x, 0)$. Recall that at time s we assume to know a given joint distribution I_s of the demands $(D_s, \ldots, D_T)$.

Using this approach, consider any feasible policy P and let $H_t^P := H_t^P(Q_t^P)$ (for $t = 1, \ldots, T$) be the discounted ordering and expected holding cost incurred by the additional Q_t^P units ordered in period t by policy P. Thus,

$$H_t^P = H_t^P(Q_t^P) := c_t Q_t^P + \sum_{j=t+L}^{T} h_j(Q_t^P \;-\; (D_{[t,j]} - X_t)^+)^+$$

(assume again $\alpha = 1$). Now let Π_t^P be the discounted expected backlogging cost incurred in period $t + L$ ($t = 1 - L, \ldots, T - L$). That is, $\Pi_t^P := p_{t+L}(D_{[t,t+L]} - (X_{t+L} + Q_t^P))^+$ (where $D_j := 0$ with probability 1 for each $j \leq 0$, and $Q_t^P = q_t$ for each $t \leq 0$).

Let $\mathcal{C}(P)$ be the cost of the policy P. Clearly,

$$\mathcal{C}(P) := \sum_{t=1-L}^{0} \Pi_t^P + H_{(-\infty,0]} + \sum_{t=1}^{T-L} (H_t^P + \Pi_t^P), \tag{1}$$

where $H_{(-\infty,0]}$ denotes the total holding cost incurred by units ordered before period 1. We note that the first two expressions $\sum_{t=1-L}^{0} \Pi_t^P$ and $H_{(-\infty,0]}$ are not affected by our decisions (i.e., they are the same for any feasible policy and each realization of the demand), and therefore we will omit them. Since they are non-negative, this will not effect our approximation results. Also observe that without loss of generality, we can assume that $Q_t^P = H_t^P = 0$ for any policy P and each period $t = T-L+1, \dots, T$, since nothing that is ordered in these periods can be used within the given planning horizon. We now can write

$$\mathcal{C}(P) = \sum_{t=1}^{T-L} (H_t^P + \Pi_t^P). \tag{2}$$

In some sense, we change the accounting of the holding cost from periodical to marginal. In the next section, we shall demonstrate that this new cost accounting approach serves as a powerful tool for designing simple approximation algorithms that can be analyzed with respect to their worst-case expected performance.

2.4 Dual-Balancing Policy - Uncapacitated Model

In this section, we consider a new policy for the uncapacitated periodic-review stochastic inventory control problem, which we call a *dual-balancing policy*. In this policy we aim to balance the expected marginal holding cost against the expected marginal backlogging penalty cost. In each period $s = 1, \dots, T - L$, we focus on the units that we order in period s only, and balance the expected and holding cost they are going to incur over $[s, T]$ against the expected backlogging cost in period $s + L$. We do that using the marginal accounting of the holding cost as introduced in Section 2.3 above.

We next describe the details of the policy, which is very simple to implement, and then analyze its expected performance. In particular, we will show that for any input of demand distributions and cost parameters, the expected cost of the dual-balancing policy is at most twice the expected cost of an optimal policy. At the end of this section we will show that the worst-case guarantee of 2 is tight. Specifically, we will show that there exists a set of instances for which the ratio between the expected cost of the dual-balancing policy and the expected cost of the optimal policy converges to 2 asymptotically. A superscript B will refer to the dual-balancing policy described below.

Recall the assumption discussed in Section 2.2 that the cost parameters imply no motivation for holding inventory or backorders. This implies that, without loss of generality, for each $t = 1, \ldots, T$, that $c_t = 0$ and $h_t,\ p_t \geq 0$ (using a standard cost transformation from inventory theory). Moreover, we first describe the algorithm and its analysis under the latter assumption. Then in Section 2.7.3 we discuss in detail the generality of this assumption. In that section, we will also show how a simple transformation of the costs can yield a better worst-case performance guarantee and certainly a better typical (average) performance in many cases in practice.

2.4.1 The Algorithm

We first describe the algorithm and its analysis in the case where fractional orders are allowed. In Section 2.7, we will show how to extend the algorithm and the analysis to the case in which the demands and the order sizes are integer-valued. In each period $s = 1, \ldots, T - L$, we consider a given information set f_s (where again $f_s \in \mathcal{F}_s$) and the resulting pair (x_s^B, I_s) the dual-balancing policy's inventory position at the beginning of period s and the conditional joint distribution I_s of the demands $(D_s, \ldots, D_T)$. We then consider the following two functions:

(i) The expected holding cost over $[s, T]$ incurred by the *additional* q_s units ordered in period s, conditioned on f_s. We denote this function by $l_s^B(q_s)$, where $l_s^B(q_s) := E[H_s^B(q_s)|f_s]$. As we have seen in Section 2.3,

$$H_t^B(Q_t) := \sum_{j=t+L}^{T} h_j(Q_t \ - \ (D_{[t,j]} - X_t)^+)^+$$

(recall that $c_t = 0$).

(ii) The expected backlogging cost incurred in period $s + L$ as a function of the additional q_s units ordered in period s, conditioned again on f_s. We denote this function by $\pi_s^B(q_s)$, where $\pi_s^B(q_s) := E[\Pi_s^B(q_s)|f_s]$. In Section 2.3 we have defined $\Pi_t^B := p_t(D_{[t,t+L]} - (X_t^B + Q_t))^+ = p_t(D_{[t,t+L]} - Y_t^B)^+$. We note that conditioned on a specific $f_s \in \mathcal{F}_s$ and given any policy P, we already know x_s, the starting inventory position in time period s. Hence, the backlogging cost in period s, $\Pi_s^B|f_s$, is indeed a function only of q_s and future demands.

The dual-balancing policy now orders $q_s^B = q_s'$ units in period s, where q_s' is such that $l_s^B(q_s') = \pi_s^B(q_s')$. In other words, we set q_s' so that the expected holding cost incurred over the time interval $[s, T]$ by the additional q_s' units we order at s is equal to the expected backlogging cost in period $s + L$, i.e., $E[H_s^B(q_s')|f_s] = E[\Pi_s^B(q_s')|f_s]$. Since we assume that fractional orders are allowed, we know that the functions $l_t^P(q_t)$ and $\pi_t^P(q_t)$ are continuous in q_t, for each $t = 1, \ldots, T - L$ and each feasible policy P.

Note again that for any given policy P, once we condition on a specific information set $f_s \in \mathcal{F}_s$, we already know x_s^P deterministically. It is then straightforward to verify that both $l_s^P(q_s)$ and $\pi_s^P(q_s)$ are convex functions of q_s. Moreover, the function $l_s^P(q_s)$ is equal to 0 for $q_s = 0$ and is an increasing function in q_s, which goes to infinity as q_s goes to infinity. In addition, the function $\pi_s^P(q_s)$ is non-negative for $q_s = 0$ and is a decreasing function in q_s, which goes to 0 as q_s goes to infinity. Thus, q'_s is well-defined and we can indeed balance the two functions.

Observe the difference between the marginal holding cost function l_s that accounts for costs over an entire time interval, and the backlogging cost function π_s that accounts for costs incurred in a single period. The intuitive explanation is that in an uncapacitated model, underordering (i.e., ordering 'too little') can always be fixed in the next period to avoid further costs. On the other hand, since we can not order a negative number of units, overordering (i.e., ordering 'too many' units) can not be fixed by any decision made in future periods, and the resulting costs are only a function of future demands, not of future orders. We also point out that q'_s can be computed as the minimizer of the function

$g_s(q_s^B) := \max\{l_s^B(q_s), \pi_s^B(q_s)\}$. Since $g_s(q_s)$ is the maximum of two convex functions of q_s, it is also a convex function of q_s. This implies that in each period s we need to solve a single-variable convex minimization problem and this can be solved efficiently. In particular, if for each $j \geq s$, $D_{[s,j]}$ has any of the distributions that are commonly used in inventory theory, then it is extremely easy to evaluate the functions $l_s^P(q_s)$ and $\pi_s^P(q_s)$ (observe that x_s is known at time s). More generally, the complexity of the algorithm is of order T (number of time periods) times the complexity of solving the single variable convex minimization defined above. The complexity of this minimization problem can vary depending on the level of information we assume on the demand distributions and their characteristics. In all of the common scenarios there exist straightforward methods to solve this problem efficiently (see also [23]). In particular, q'_s lies at the intersection of two monotone convex functions, which suggests that bi-section methods can be effective in computing q'_s. Note that in the presence of positive lead times even computing a simple myopic policy requires to know the distribution of the accumulated demand over the lead time.

Finally, observe that the dual-balancing policy is not a state-dependent base-stock policy. That is, the control of the dual-balancing policy does depend on the inventory control policy in past periods, namely on x_s^B. However, it can be implemented on-line, i.e., it does not require any knowledge of the control policy in future periods. Thus, we avoid the burden of solving large dynamic programming problems. Moreover, unlike the myopic policy, the dual-balancing policy is using in each period, available information about the future demands.

This concludes the description of the algorithm for continuous-demand case. Next we describe the analysis of the worst-case expected performance of this policy.

2.4.2 Analysis

Next we shall show that, for each instance of the problem, the expected cost of the dual-balancing policy described above is at most twice the expected cost of an optimal policy. We will use the marginal cost accounting approach described in Section 2.3 (see (2) above), and amortize the period cost of the dual-balancing policy with the cost of the optimal policy.

Using the marginal holding cost accounting approach discussed in Section 2.3, the expected cost of the dual-balancing policy can be expressed as

$$E[\mathcal{C}(B)] = \sum_{t=1}^{T-L} E[H_t^B + \Pi_t^B].$$

For each $t = 1, \ldots, T - L$, let Z_t be the *random balanced cost* by the dual-balancing policy in period t, i.e., $Z_t = E[H_t^B|\mathcal{F}_t] = E[H_t^B|\mathcal{F}_t]$. Note that Z_t is realized in period t as a function of the observed information set f_t (we will denote its realization by z_t). By the construction of the dual-balancing policy, we know that, with probability 1,

$E[H_t^B|\mathcal{F}_t] = E[\Pi_t^B|\mathcal{F}_t]$, for each period $t = 1, \ldots, T-L$. This implies that, for each period t, $E[H_t^B + \Pi_t^B|\mathcal{F}_t] = 2Z_t$, that proves the following lemma follows.

Lemma 2.4.1 *The expected cost of the dual-balancing policy is equal to twice the expected sum of the Z_t variables, i.e.,* $E[\mathcal{C}(B)] = 2\sum_{t=1}^{T-L} E[Z_t]$.

Proof : Using the marginal cost accounting discussed in Section 2.3 and a standard argument of conditional expectations we express

$$E[\mathcal{C}(B)] = \sum_{t=1}^{T-L} E[H_t^B + \Pi_t^B] = \sum_{t=1}^{T-L} E[E[H_t^B + \Pi_t^B|\mathcal{F}_t]] = 2\sum_{t=1}^{T-L} E[Z_t].$$

■

Next we will state and prove two lemmas which imply that the expected cost of an optimal policy is at least $\sum_{t=1}^{T-L} E[Z_t]$. For each realization of the demands $D_1, \ldots, D_T$, let $\mathcal{T}_H$ be the set of periods in which the optimal policy had more inventory than the dual-balancing policy, i.e., the set of periods t such that $Y_t^B < Y_t^{OPT}$. Let $\mathcal{T}_\Pi$ be the set of periods in which the dual-balancing had at least as much inventory as OPT, i.e., the set of periods t such that $Y_t^B \geq Y_t^{OPT}$. Observe that $\mathcal{T}_H$ and $\mathcal{T}_\Pi$ are random sets that induce a random partition of the planning horizon. The next lemma shows that, with probability 1, the marginal holding cost incurred by the dual-balancing policy in periods $t \in \mathcal{T}_H$ is at most the overall holding cost incurred by OPT, denoted by H^{OPT}, i.e., $\sum_{t\in\mathcal{T}_H} H_t^B \leq H^{OPT}$ with probability 1.

Recall the concepts of $\mathcal{L}_D$, the line of potential units of demand to be realized over the horizon, and $\mathcal{L}_S$, the line of supply units to be ordered over the planning horizon, discussed in Section 2.3 above. Since the demand is independent from the inventory policy, we can compare between any two feasible policies by looking at the respective periods in which each supply unit in $\mathcal{L}_S$ was ordered. The proof technique in the next lemma will be based on such comparison between the dual-balancing policy and an optimal policy

Lemma 2.4.2 *For each realization $f_T \in \mathcal{F}_T$, the marginal holding cost incurred by the dual-balancing policy in all periods $t \in \mathcal{T}_H$ is at most the overall holding cost incurred by OPT, denoted by H^{OPT}, i.e., $\sum_{t \in \mathcal{T}_H} H_t^B \leq H^{OPT}$ with probability 1.*

Proof : Consider an information set $f_T \in \mathcal{F}_T$ which corresponds to a complete evolution over the planning horizon, and some period $s \in \mathcal{T}_H$. We slightly abuse the notation and let $\mathcal{T}_H$ denote the deterministic set of periods that corresponds to the specific information set f_T. Let $\mathcal{Q}_s \subseteq \mathcal{L}_S$ be the set of supply units ordered by the dual-balancing policy in period s, where clearly, $|\mathcal{Q}_s| = q_s'$. By the definition of $\mathcal{T}_H$, we know that in period s we had $y_s^B < y_s^{OPT}$. This implies that the units in $\mathcal{Q}_s$ were ordered by OPT either in period s or even prior to s. Since we assume that $c_s = 0$ and that $h_t \geq 0$ for each period t, we conclude that the holding cost that these units have incurred in OPT is at least as much as the holding cost they have incurred in the dual-balancing policy.

We conclude the proof by observing that the sets $\{\mathcal{Q}_s : s \in \mathcal{T}_H\}$ are of disjoint supply units since they consist of units ordered by the dual-balancing policy in different periods. This implies that indeed $\sum_{t \in \mathcal{T}_H} H_t^B \leq H^{OPT}$, with probability 1. ∎

The next lemma shows that, with probability 1, the marginal backlogging penalty cost of the dual-balancing policy associated with periods $t \in \mathcal{T}_\Pi$ is at most the overall backlogging penalty incurred by OPT, denoted by Π^{OPT}.

Lemma 2.4.3 *For each realization $f_T \in \mathcal{F}_T$, the marginal backlogging penalty cost of the dual-balancing policy associated with all periods $t \in \mathcal{T}_\Pi$ is at most the overall backlogging penalty incurred by OPT, denoted by Π^{OPT}, i.e., $\sum_{t \in \mathcal{T}_\Pi} \Pi_t^B \leq \Pi^{OPT}$ with probability 1.*

Proof : Consider a realization $f_T \in \mathcal{F}_T$ and some period $s \in \mathcal{T}_\Pi$ (where again we abuse the notation and use $\mathcal{T}_\Pi$ to denote a deterministic set). Note that period s is associated with the backlogging cost incurred in period $s + L$. By definition of $\mathcal{T}_\Pi$ we know that $y_s^B \geq y_s^{OPT}$. However, this implies that, with probability 1, the backlogging cost incurred by the dual-balancing policy in period $s + L$ are no greater than the respective backlogging cost incurred by the optimal policy in period $s+L$. The proof then follows.

■

As a corollary of Lemmas 2.4.1, 2.4.2 and 2.4.3 we get the following theorem.

Theorem 2.4.4 *The dual-balancing policy for the uncapacitated model has a worst-case performance guarantee of 2, i.e., for each instance of the capacitated periodic-review stochastic inventory control problem, the expected cost of the dual-balancing policy is at most twice the expected cost of an optimal solution, i.e.,*
$E[\mathcal{C}(B)] \leq 2E[\mathcal{C}(OPT)]$.

Proof : From Lemma 2.4.1, we know that the expected cost of the dual-balancing policy is equal to twice the expected cost of the sum of the Z_t variables, i.e.,
$E[\mathcal{C}(B)] = 2\sum_{t=1}^{T-L} E[Z_t]$. From Lemmas 2.4.2 and 2.4.3 we know that, with probability 1, the cost of OPT is at least as much as the holding cost incurred by units ordered by the dual-balancing policy in periods $t \in \mathcal{T}_H$ plus the backlogging cost of the dual-balancing policy that is associated with periods $t \in \mathcal{T}_\Pi$. In other words, with probability 1, $H^{OPT} + \Pi^{OPT} \geq \sum_{t\in\mathcal{T}_H} H_t^B + \sum_{t\in\mathcal{T}_\Pi} \Pi_t^B$. Using again conditional expectations and the definition of Z_t, this implies that indeed,

$$E[\mathcal{C}(OPT)] \geq E[\sum_{t \in \mathcal{T}_H} H_t^B + \sum_{t \in \mathcal{T}_\Pi} \Pi_t^B] =$$
$$\sum_t E[H_t^B \cdot \mathbb{1}(t \in \mathcal{T}_H) + \Pi_t^B \cdot \mathbb{1}(t \in \mathcal{T}_\Pi)] =$$
$$\sum_t E[E[H_t^B \cdot \mathbb{1}(t \in \mathcal{T}_H) + \Pi_t^B \cdot \mathbb{1}(t \in \mathcal{T}_\Pi)|\mathcal{F}_t]] =$$
$$\sum_t E[(\mathbb{1}(t \in \mathcal{T}_H) + \mathbb{1}(t \in \mathcal{T}_\Pi))Z_t] = \sum_t E[Z_t].$$

We note that if the optimal policy is deterministic (i.e., it makes deterministic decisions in each period t given the observed information set f_t), then if we condition on $\mathcal{F}_t$, y_t^B and y_t^{OPT} are known deterministically, and so are the indicators $\mathbb{1}(t \in \mathcal{T}_H)$ and $\mathbb{1}(t \in \mathcal{T}_\Pi)$. Suppose that the optimal policy is a randomized policy, i.e., it selects an order of size Q_t^{OPT} as a random function of f_t, then the same arguments above still work. We now need to condition not only on $\mathcal{F}_t$ but also on Q_t^{OPT}. Since the inventory control policy does not have any effect on the evolution of the future demands, the arguments above are still valid. This concludes the proof of the theorem. ■

Observe that in a capacitated model, the dual-balancing policy described above might not work out. In particular, the balancing order q_s' might not be reachable in the case where $q_s' > u_s$. In the next section, we describe a novel marginal backlogging cost accounting approach that gives rise to a dual-balancing policy for the capacitated model. Next we will show that the analysis above is in fact tight by demonstrating a set of instances for which the ratio between the expected cost of the dual-balancing policy and the expected cost of an optimal policy converges to 2 asymptotically.

Dual-Balancing - Bad Example

The following example was constructed based on a suggestion in [39]. Consider an instance with $h > 0$, $p = h\sqrt{L}$ where $L > 0$ is again the a positive integer that denotes the lead time between the time an order is placed and the time it arrives. Also assume that $T = 1 + 2L$ and $\alpha = 1$. The random demands have the following structure. there is one unit of demand that is going to occur with equal probability either in period $L + 1$ or in period $2L + 1$. For each $t \neq L + 1,\ 2L + 1$, we have $D_t = 0$ with probability 1. Fractional orders are allowed.

It is readily verified that the optimal policy orders 1 unit in period 1 and incurs expected cost of $\frac{1}{2}hL$. On the other hand, the dual-balancing policy will order in each one of the periods $1, \ldots, \sqrt{L}$ just a small amount of the commodity. In particular, in period 1, the dual-balancing orders $\frac{1}{\sqrt{L}+1}$ of a unit (this can be calculated by equating $\frac{1}{2}(\sqrt{L}h)(1 - q) = \frac{1}{2}Lhq$, where q is the size of the order). It can be easily verified that when L goes to ∞, the ratio between the expected cost of the dual-balancing policy and the expected cost of the optimal policy converges to 2 (the calculations are rather messy to present but can be easily coded).

2.5 Marginal Backlogging Cost Accounting Approach in Capacitated Model

Recall the observation regarding the fundamental difference in uncapacitated models between holding cost and backlogging penalty cost. That is, any mistake of ordering ‘too little’ can be fixed in the next period to avoid further backlogging penalty cost, while the effect of ordering ‘too much’, may last for a number of periods depending on the realized future demands. In particular, no future decision can fix this mistake, since we can not order a negative quantity. Consequently, in the uncapacitated case Π_t^P only accounts for costs incurred in a single period, namely, backlogging cost in period $t + L$, whereas H_t^P accounts for holding costs incurred over multiple periods. By way of contrast, in models with capacity constraints on the size of the order in each period, the above observation is not valid anymore. More specifically, because of the capacity constraints, it is not longer true that a mistake of ordering 'too little' in the current period can be always fixed by decisions made in future periods.

We now consider the capacitated model, and present a new backlogging penalty cost accounting approach that associates with the decision of how many units to order in period s what we shall call the *forced backlogging penalty cost* resulting from this decision in future periods.

Consider some period s. Suppose that x_s is the inventory position at the beginning of period s and that the number of units ordered in the period is $q_s < u_s$. Let $\bar{q}_s$ be the resulting *unused slack capacity* in period s, i.e., $\bar{q}_s = u_s - q_s > 0$. Focus now on some future period $t \geq s + L$ after this order arrives and becomes available. Suppose that for some realization of the demands $d_{[s,t]} - (x_s + q_s + \sum_{j \in (s,t-L]} u_j) > 0$. This implies that there exists a shortage in period t, and moreover, that even if in every period *after* period s and until period $t - L$ the orders had been up to the maximum available capacity, this part of the shortage in period t would still exist and incur the corresponding backlogging penalty cost. The actual shortage may be even bigger and is equal to $d_{[s,t]} - (x_s + q_s + \sum_{j \in (s,t-L]} q_j) > 0$ (recall that $q_j \leq u_j$ for each period j). In other words, given our decision in period s, this part of the shortage could not be avoided by any decision made over the interval $(s, t - L]$ (clearly, any order placed after period $t - L$ will not be available by time t). We conclude that, if more units had been ordered in period s, then at least some of the shortage in period t could have been avoided. More precisely, the maximum number of units of shortage that could have been avoided by ordering more units in period s is equal to $\min\{\bar{q}_s,\ (d_{[s,t]} - (x_s + q_s + \sum_{j \in (s,t-L]} u_j))^+\}$. The intuition is that by ordering more units in period s, we could have averted part of the shortage in period t, but clearly not more than the unused slack capacity $\bar{q}_s$, since we could not have ordered in period s more than additional $\bar{q}_s$ units. In this case, we would say that this part of the backlogging penalty cost in period t was *forced* by the decision in period s, and hence period s is associated with a backlogging penalty of $p_t \min\{\bar{q}_s,\ (d_{[s,t]} - (x_s + q_s + \sum_{j \in (s,t-L]} u_j))^+\}$. This is significantly different from the *traditional* backlogging penalty cost accounting, in which this penalty cost would be associated with period $t - L$.

We let W_{st} be the shortage in period t that is forced by the decision in period s (where again $s \leq t - L$), i.e.,

$$W_{st} := \min\{\bar{Q}_s,\ (D_{[s,t]} - (X_s + Q_s + \sum_{j \in (s,t-L]} u_j))^+\}.$$

An alternative way to express W_{st} is

$$W_{st} = (D_{[s,t]} - (X_s + Q_s + \sum_{j \in (s,t-L]} u_j))^+ - (D_{[s,t]} - (X_s + \sum_{j \in [s,t-L]} u_j))^+. \tag{3}$$

Now using the equalities, $NI_t = X_s + Q_s + \sum_{j \in (s,t-L]} Q_j - D_{[s,t)}$ (for each $s \leq t - L$) and $u_j = Q_j + \bar{Q}_j$ (for each $j = s, \ldots, t - L$), we conclude that equation (3) can be written as

$$(D_t - NI_t - \sum_{j \in (s,t-L]} \bar{Q}_j)^+ - (D_t - NI_t - \sum_{j \in [s,t-L]} \bar{Q}_j)^+. \tag{4}$$

To see why (3) (hence (4)) above holds, observe that the inequality

$$(D_{[s,t]} - (X_s + Q_s + \sum_{j \in (s,t-L]} u_j))^+ > \bar{Q}_s$$

is equivalent to the inequality

$$(D_{[s,t]} - (X_s + \sum_{j \in [s,t-L]} u_j))^+ > 0.$$

Next we describe several properties of the parameters W_{st}. Clearly, if $\bar{Q}_s = 0$ (i.e. $Q_s = u_s$), then $W_{st} = 0$ for each $t \geq s + L$. It is also readily verified from (4) above that if $W_{st} > 0$ for some $s \leq t - L$, then we have $W_{jt} = \bar{Q}_j$ for each $j \in (s, t - L]$.

Let $\bar{\Pi}_s^P$ (for $s = 1 - L, \ldots, T - L$) be equal to the overall forced backlogging cost in periods $s + L, \ldots, T$ associated with period s, i.e., let $\bar{\Pi}_s^P = \sum_{t=s+L}^{T} p_t W_{st}^P$ (we again assume that $D_j = d_j$ with probability 1 for each $j \leq 0$). Let $u_{-L} = \infty,\ q_{-L} = 0$ and $\bar{q}_{-L} = \infty$, and also define, for each $t = 1, \ldots, T$,

$$W_{-L,t} := (D_{[1-L,t]} - (x_{1-L} + \sum_{j \in [1-L, t-L]} u_j))^+ = (D_t - NI_t - \sum_{j \in [1-L, t-L]} \bar{Q}_j)^+,$$

and $\bar{\Pi}_{-L}^P = \bar{\Pi}_{-L} := \sum_{t=1}^{T} p_t W_{-L,t}$. The last definition of $\bar{\Pi}_{-L}$ is meant to account for forced backlogging penalty cost which is independent of *any* decision, and is forced by the demands on *any* feasible policy. It is now readily verified that, for each $t = 1, \ldots, T$, we have $\bar{\Pi}_{t-L} = \sum_{j=-L}^{t-L} W_{jt} = (D_t - NI_t)^+$ (the sum $\sum_{j=-L}^{t-L} W_{jt}$ is telescopic). This implies the following theorem.

Theorem 2.5.1 *Let P be a non-anticipatory policy. Then the cost of the policy P is $\mathcal{C}(P) := \sum_{t=-L}^{0} \bar{\Pi}_t^P + H_{(-\infty,T]} + \sum_{t=1}^{T-L} (H_t^P + \bar{\Pi}_t^P)$.*

To provide more intuition, we demonstrate the new backlogging cost accounting approach through a simple example. Suppose that the order capacity is 5 in all periods, $L = 0$ and $\alpha = 1$. Assume that the inventory position at the beginning of period 3 was $x_3 = 3$, and that we have ordered $q_3 = 3,\ q_4 = 5,\ q_5 = 4$ and $q_6 = 2$ units in periods 3, 4, 5 and 6, respectively. Now say that the demands were $d_3 = 3,\ d_4 = 3,\ d_5 = 5$ and $d_6 = 11$ in periods 3, 4, 5 and 6, respectively. In particular, the accumulated demand over periods $[3, 6]$, $d_{[3,6]}$, is equal to 22. This implies that in period 6 we had a shortage of 5 units, each of which incurred a penalty cost of p_t at the end of period 6. Out of these 5 units of shortage at the end of period 6, we associate a backlogging penalty of 3 units of shortage with period 6 (the unused slack capacity in this period is 3), a penalty of 1 unit of shortage with period 5 (the unused slack capacity in this period is 1), no cost is associated with period 4 since we ordered up to capacity, and finally the penalty of 1 units of shortage is associated with period 3 ($d_{[3,6]} - (3+3+5+5+5) = 1$). In other words, $w_{36} = 1,\ w_{46} = 0,\ w_{56} = 1$ and $w_{66} = 3$. This example illustrates how we backtrack the 'source' of each unit of shortage and its corresponding backlogging penalty cost incurred in period t, and associate it as forced backlogging penalty cost to past periods. If $L > 0$, then we start the backtracking in period $t - L$, since only orders in periods earlier than $t - L + 1$ could have arrived by time t.

The intuition is that once a shortage is incurred in period t, it is allocated to past periods $s \leq t - L$ in which the orders were below the available capacity. More specifically, the shortage and the resulting backlogging cost in period t are charged to periods $s \leq t-L$ with positive unused slack capacity going backward in time from period $t-L$. Each period $s \leq t - L$, can be charged with a backlogging penalty cost in period t for up to $\bar{q}_s$ units, the unused slack capacity in period s.

Note that the first two terms of $\mathcal{C}(P)$ in Theorem 2.5.1, $\sum_{t=-L}^{0} \bar{\Pi}_t^P$ and $H_{(-\infty,T]}$, are independent of any decision we make and are common to all feasible policies. Recall that $\sum_{t=-L}^{0} \bar{\Pi}_t^P$ represents the forced backlogging penalty that is forced on any feasible policy. Since these two terms are also non-negative, we omit them from the analysis. This does not impact our approximation results. From now on we will write the cost of a feasible policy P as $\mathcal{C}(P) = \sum_{t=1}^{T-L} (H_t^P + \bar{\Pi}_t^P)$.

Finally, observe that for uncapacitated models with $u_s = \infty$ for each s (and hence $\bar{q}_s = \infty$), our backlogging cost accounting approach is in fact identical to the traditional backlogging accounting scheme discussed above. This implies that the cost accounting scheme proposed in this section is a generalization of the one discussed in Section 2.3

above. Therefore, the proceeding discussion in Section 2.6 below is also a generalization of the corresponding algorithm and analysis in Section 2.4 above.

2.6 Dual-Balancing Policy - Capacitated Model

In this section, we describe a new policy for the capacitated periodic-review stochastic inventory control problem. Similar to Section 2.4 above, we call it a *dual-balancing policy*. We shall show that this policy has a worst-case performance guarantee of 2, i.e., for each instance of the problem, the expected cost of the policy is at most twice the expected cost of an optimal policy. As in Section 2.4 we still assume, without loss of generality, that for each $t = 1, \ldots, T$, $c_t = 0$ and $h_t,\ p_t \geq 0$, and that fractional orders are allowed.

The dual-balancing policy presented in this section is based on a balancing idea similar to the one used in Section 2.4 for the uncapacitated model. In the uncapacitated case, the dual-balancing policy is balancing, in each period s and conditioned on the observed information set f_s, the expected marginal holding cost of the units ordered in the period against the expected (traditional) backlogging penalty cost in period $s + L$, a lead time ahead of s. As we have already seen that this approach does not work in the case where there is a capacity constraint on the size of the order in period s. For once, the order size q'_s that balances these two costs may not be reachable when $q'_s > u_s$.

In turn, we consider the marginal backlogging penalty cost accounting and the corresponding cost it associates with period s as described in Section 2.5 above. Conditioned on the observed information set f_s, we now balance the expected marginal holding cost of the units ordered in period s against the expected *marginal* backlogging penalty costs associated with period s. We will again use superscript B to refer to the dual-balancing policy. For each period $s = 1, \ldots, T - L$, conditioning on the observed information set f_s, let again $l_s^B(q_s^B)$ be the expected holding cost incurred over $[s,T]$ by the units ordered by the dual-balancing policy in period s. That is, $l_s^B(q_s^B) := E[H_s^B(q_s^B)|f_s]$. In addition, let $\bar{\pi}_s^B(q_s^B)$ be the expected backlogging penalty cost associated with period s by the modified marginal backlogging penalty cost accounting, again conditioned on the observed information set f_s. More precisely, $\bar{\pi}_s^B := E[\bar{\Pi}_s^B(q_s^B)|f_s]$. Recall that in Section 2.5 we have defined $\bar{\Pi}_s^B = \sum_{t=s+L}^{T} p_t W_{st}^B$ where,

$$W_{st}^B = \min\{\bar{Q}_s^B,\ (D_{[s,t]} - (X_s^B + Q_s^B + \sum_{j\in(s,t]} u_j))^+\} =$$
$$(D_{[s,t]} - (X_s^B + Q_s^B + \sum_{j\in(s,t-L]} u_j))^+ - (D_{[s,t]} - (X_s^B + \sum_{j\in[s,t-L]} u_j))^+.$$

Since if we condition on f_s, x_s^B the inventory position at the beginning of period s, is known deterministically, it is clear that $l_s^B(q_s^B)$ and $\bar{\pi}_s^B(q_s^B)$ are both indeed functions of q_s^B, the number of units ordered in period s.

Since fractional orders are allowed, the functions $l_s^B(q_s^B)$ and $\bar{\pi}_s^B(q_s^B)$ are continuous. In each period $s = 1,\dots,T-L$, given the observed information set f_s, the dual-balancing policy will order $q_s^B = q_s' \leq u_s$ units such that the expected marginal ordering and holding cost incurred by these units over $[s,T]$ is equal to the expected marginal backlogging penalty cost associated with period s. In other words, we order q_s' units such that $l_s^B(q_s') = E[H_s^B(q_s')|f_s] = \bar{\pi}_s^B(q_s') = E[\bar{\Pi}_s^B(q_s')|f_s]$. Next we show that this policy is well-defined. It is readily verified that $l_s^B(q_s^B)$ is a convex increasing function of q_s^B that is equal 0 for $q_s^B = 0$ and that is going to ∞ as q_s^B goes to ∞. Similarly, one can verify that $\bar{\pi}_s^B(q_s^B)$ is a decreasing convex function of q_s^B that has a non-negative value at $q_s^B = 0$ and that is equal to 0 for $q_s^B = u_s$ (in this case there is no unused slack capacity at s and $\bar{q}_s^B = 0$). Our assumption that these functions are continuous implies that q_s' as defined above always exists. Moreover, q_s' is the again the minimizer of the function $g_s(q_s^B) := \max\{l_s^B(q_s^B), \bar{\pi}_s^B(q_s^B)\}$, which is a convex function of q_s^B being a maximum of two convex functions. Hence, in each period s, we again need to solve a convex minimization problem of a single variable. In particular, if for each $j \geq s$, $D_{[s,j]}$ has any of the distributions that are commonly used in inventory theory, then it is extremely easy to evaluate the functions $l_s^B(q_s^B)$ and $\bar{\pi}_s^B(q_s^B)$. More generally, the complexity of the algorithm is again of order T (the number of time periods) times the complexity of solving the single variable convex minimization defined above. The complexity of this minimization problem can vary depending on the level of information we assume on the demand distributions and their characteristics. In all of the common scenarios there exist straightforward methods to solve this problem efficiently. In particular, q'_s lies in the intersection of two monotone convex functions, which suggests that bi-section methods can be effective in computing q'_s.

We note that as in the uncapacitated case, the dual-balancing policy for the capacitated model is not a state-dependent base stock policy. However, it can be computed in an *on-line* manner, i.e., computing the policy action in period s does not require any knowledge on the future decisions to be made in the next periods.

2.6.1 Analysis

Next we shall show that, for each instance of the problem, the expected cost of the dual-balancing policy described above is at most twice the expected cost of an optimal policy. The analysis will follow along the lines of the analysis in the uncapacitated case discussed in Section 2.4 above.

Using the marginal cost accounting scheme discussed in Section 2.5, the expected cost of the dual-balancing policy can be expressed as $E[\mathcal{C}(B)] = \sum_{t=1}^{T-L} E[H_t^B + \bar{\Pi}_t^B]$. For each $t = 1, \ldots, T-L$, let Z_t be again the *random balanced cost* by the dual-balancing policy in period t, i.e., $Z_t = E[H_t^B|\mathcal{F}_t] = E[H_t^B|\mathcal{F}_t]$. Note that Z_t is a function of the observed information set in period t. By the construction of the dual-balancing policy, we again know that, with probability 1, $E[H_t^B|\mathcal{F}_t] = E[\bar{\Pi}_t^B|\mathcal{F}_t]$, for each period $t = 1, \ldots, T-L$. The following lemma is analogous to Lemma 2.4.1 and follows by identical arguments.

Lemma 2.6.1 *The expected cost of the dual-balancing policy is equal to twice the expected sum of the Z_t variables, i.e., $E[\mathcal{C}(B)] = 2\sum_{t=1}^{T-L} E[Z_t]$.*

As in Section 2.4, let $\mathcal{T}_H$ be again the set of periods in which the optimal policy had more inventory than the dual-balancing policy, i.e., the set of periods t such that $Y_t^B < Y_t^{OPT}$. Similarly, let $\mathcal{T}_\Pi$ be again the set of periods in which the dual-balancing had at least as much inventory as OPT, i.e., the set of periods t such that $Y_t^B \geq Y_t^{OPT}$. It is readily verified that Lemma 2.4.2 is still valid, i.e., $\sum_{t\in\mathcal{T}_H} H_t^B \leq H^{OPT}$ with probability 1 (where H^{OPT} is again the holding cost incurred by the optimal policy over the entire horizon).

It is left to show that, with probability 1, the marginal backlogging penalty cost of the dual-balancing policy associated with periods $t \in \mathcal{T}_\Pi$ is at most the overall backlogging penalty incurred by OPT, denoted again by Π^{OPT}. This is done in the next lemma which is analogous to Lemma 2.4.3.

Lemma 2.6.2 *For each realization $f_T \in \mathcal{F}_T$, the marginal backlogging penalty cost of the dual-balancing policy associated with all periods $t \in \mathcal{T}_\Pi$ is at most the overall backlogging penalty incurred by OPT, denoted by Π^{OPT}, i.e., $\sum_{t\in\mathcal{T}_\Pi} \bar{\Pi}_t^B \leq \Pi^{OPT}$ with probability 1.*

Proof : The marginal backlogging penalty cost associated with each period $s \in \mathcal{T}_\Pi$ is equal to

$$\sum_{s\in\mathcal{T}_\Pi}\sum_{t:t\geq s+L} p_t W^B_{st} = \sum_t p_t \sum_{s\in\mathcal{T}_\Pi:s\leq t-L} W^B_{st}.$$

Therefore, it is sufficient to show that for each $t = 1,\ldots,T$, the traditional backlogging penalty cost incurred by OPT in that period is at least as much as the forced backlogging penalty costs incurred by the dual-balancing policy in period t as a result of decisions made in periods $\{s \in \mathcal{T}_\Pi : s \leq t-L\}$. In other words, it is sufficient to show that for each $t = 1,\ldots,T$, we have

$$(D_t - NI^{OPT}_t)^+ \geq \sum_{s\in\mathcal{T}_\Pi:s\leq t-L} W^B_{st},$$

with probability 1.

Consider now a specific realization $f_T \in \mathcal{F}_T$ and some period $t = 1,\ldots,T$. If there is no period in $\{s \in \mathcal{T}_\Pi : s \leq t\}$ with $w^B_{st} > 0$, then there is nothing to prove. Assume that such a period s exists, and let s_l and s_e be the latest and the earliest periods in $\{s \in \mathcal{T}_\Pi : s \leq t-L\}$ with $w^B_{st} > 0$, respectively (it is possible that $s_l = s_e$). We note again that here we abuse the notation and consider the set $\mathcal{T}_\Pi$ as the realized set of periods according to the specific realization f_T. In particular, s_e and s_l are the respective realizations of random variables S_e and S_l. We have already seen (in the discussion in Section 2.5) that for each $s \in (s_e, s_l]$ we have $w^B_{st} = \bar{q}^B_s$, and

$$w^B_{s_e,t} \leq d_{[s_e,t-L]} - (x_{s_e} + q^B_{s_e} + \sum_{j\in(s_e,t-L]} u_j).$$

Indeed,

$$d_t - ni^{OPT}_t = d_t - (y^{OPT}_{s_l} + \sum_{j\in(s_l,t-L]} q^{OPT}_j - d_{[s_l,t)}) \geq$$

$$d_{[s_l,t]} - (y^B_{s_l} + \sum_{j\in(s_l,t-L]} u_j) = d_{[s_l,t]} - (y^B_{s_e} + \sum_{j\in(s_e,s_l]} q^B_j - d_{[s_e,s_l)} + \sum_{j\in(s_l,t-L]} u_j) =$$

$$d_{[s_e,t]} - (x^B_{s_e} + q^B_{s_e} + \sum_{(s_e,t-L]} u_j) + \sum_{j\in(s_e,s_l]} \bar{q}^B_j = \sum_{j\in[s_e,s_l]} w^B_{st} \geq \sum_{j\in[s_e,s_l]\cap\mathcal{T}_\Pi} w^B_{st}.$$

The first equality is based again on the identity that for each feasible policy and for each $s \le t$, we have $NI_t = Y_s + \sum_{j \in (s,t-L]} Q_j - D_{[s,t)}$, applied to OPT and periods $s_l \le t$. The first inequality follows from the assumption that $s_l \in \mathcal{T}_{\Pi}$ and so $y_{s_l}^{OPT} \le y_{s_l}^{B}$, and from the capacity constraints that imply $q_j^{OPT} \le u_j$. The second equality follows from the identity (for each $s \le s'$) $Y_{s'} = Y_s + \sum_{j \in (s,s']} Q_j - D_{[s,s')}$ applied to the dual-balancing policy and periods $s_e \le s_l$. The last equality is achieved by adding and subtracting $\sum_{j \in (s_e,s_l]} \bar{q}_j^B$ and from the identity that $u_j = Q_j + \bar{Q}_j$. The proof then follows. ■

As a corollary of Lemmas 2.6.1, 2.4.2 and 2.6.2 we get the following theorem. The proof is identical to the one of Theorem 2.4.4.

Theorem 2.6.3 *The dual-balancing policy for the capacitated model has a worst-case performance guarantee of 2, i.e., for each instance of the capacitated periodic-review stochastic inventory control problem, the expected cost of the dual-balancing policy is at most twice the expected cost of an optimal solution, i.e.,* $E[\mathcal{C}(B)] \le 2E[\mathcal{C}(OPT)]$.

2.7 Dual-Balancing Policies - Extensions

In this Section, we discuss several important extensions of the algorithms discussed above in Sections 2.4 and 2.6. We first discuss the extension of the algorithms and their worst-case analysis for the case where the demand distributions are integer-valued and the orders are allowed to be only integers. Then we discuss the extension to models with stochastic lead times, again under the assumption of non-crossing orders. Finally, we discuss the details of the cost transformation that enables us to consider models with positive per-unit ordering cost. This cost transformation may also improve the worst-case guarantees and the typical performance of the algorithms in many scenarios.

2.7.1 Integer-Valued Demands

We now discuss the case in which the demands are integer-valued random variables, and the order in each period is also restricted to an integer. In this case, in each period s, the functions $l_s^B(q_s^B)$ and $\bar{\pi}_s^B(q_s^B)$ are originally defined only for integer values of q_s^B. We define these functions for any value of q_s^B by interpolating piecewise linear extensions of the integer values. It is clear that these extended functions preserve the properties of convexity and monotonicity discussed in the previous (continuous) case. However, it is still possible (and even likely) that the value q'_s that balances the functions l_s^B and $\bar{\pi}_s^B$ is not an integer. Instead we consider the two consecutive integers q_s^1 and $q_s^2 := q_s^1 + 1$ such that $q_s^1 < q'_s < q_s^2$. In particular, $q'_s := \lambda q_s^1 + (1-\lambda)q_s^2$ for some $0 < \lambda < 1$. In periods, we now order q_s^1 units with probability λ and q_s^2 units with probability $1-\lambda$. This constructs what we call a *randomized dual-balancing policy*.

Observe that now at the beginning of time period s the order quantity of the dual-balancing policy is still a random variable $Q_s^B = Q'_s$ with support consists of two points $\{q_s^1, q_s^2\} = \{q_s^1(f_s), q_s^2(f_s)\}$ which is a function of the observed information set f_s. We would like to show that this policy admits the same performance guarantee of 2. For each $t = 1, \ldots, T-L$, let Z_t be again the random balanced cost of the dual-balancing policy in period t. Focus now on some period s. For a given observed information set $f_s \mathcal{F}_s$ we have for some $0 \leq \lambda = \lambda(f_s) \leq 1$,

$$z_s = E[H_s^B(Q'_s)|f_s] = \lambda E[H_s^B(q_s^1)|f_s] + (1-\lambda)E[H_s^B(q_s^2)|f_s] =$$
$$E[H_s^B(\lambda q_s^1 + (1-\lambda)q_s^2)|f_s],$$

and

$$z_s = E[\bar{\Pi}_s^B(Q'_s)|f_s] = \lambda E[\bar{\Pi}_s^B(q_s^1)|f_s] + (1-\lambda)E[\bar{\Pi}_s^B(q_s^2)|f_s] = E[\bar{\Pi}_s^B(\lambda q_s^1 + (1-\lambda)q_s^2)|f_s].$$

The second equality (in each of the two expressions above) follows from the fact that we consider piecewise linear functions. By the definition of the algorithm we also have,

$$\lambda E[H_s^B(q_s^1)|f_s] + (1-\lambda)E[H_s^B(q_s^2)|f_s] = \lambda E[\bar{\Pi}_s^B(q_s^1)|f_s] + (1-\lambda)E[\bar{\Pi}_s^B(q_s^2)|f_s].$$

It is now readily seen that, for each period s and each $f_s \in \mathcal{F}_s$, we again have $E[H_s^B(Q'_s) + \bar{\Pi}_s^B(Q'_s)|f_s] = 2z_s$, i.e., $E[H_s^B(Q'_s) + \bar{\Pi}_s^B(Q'_s)|\mathcal{F}_s] = 2Z_s$. This implies that Lemmas 2.4.1 and 2.6.1 are still valid.

Now define the sets $\mathcal{T}_H$ and $\mathcal{T}_{\bar{\Pi}}$ in the following way. Let $\mathcal{T}_H = \{t : X_t^B + Q_t^2 \leq Y_t^{OPT}\}$, and $\mathcal{T}_{\bar{\Pi}} = \{t : X_t^B + Q_t^2 > Y_t^{OPT}\}$. Observe that for each period s, conditioned on some $f_s \in \mathcal{F}_s$, we know deterministically x_s^B, q_2^B and, if the optimal policy is deterministic, we also know y_s^{OPT}. Therefore, we know whether $s \in T_H$ or $s \in \mathcal{T}_{\Pi}$. If the optimal policy is also a randomized policy, we condition not only on f_s but also on the decision made by the optimal policy in period s. Moreover, if $s \in T_H$, then, with probability 1, $Y_s^B \leq Y_s^{OPT}$, and if $s \in \mathcal{T}_{\bar{\Pi}}$, then, with probability 1, $Y_s^B \geq Y_s^{OPT}$. This implies that Lemmas 2.4.2, 2.4.3 and 2.6.2 are also still valid. The following theorem is now established (the proof is identical to that of Theorem 2.4.4 above).

Theorem 2.7.1 *The randomized dual-balancing policy has a worst-case performance guarantee of 2, i.e., for each instance of the capacitated (uncapacitated) periodic-review stochastic inventory control problem, the expected cost of the randomized dual-balancing policy is at most twice the expected cost of an optimal solution, i.e,* $E[\mathcal{C}(B)] \leq 2E[\mathcal{C}(OPT)]$.

2.7.2 Stochastic Lead Times

In this section, we consider the more general model, where the lead time of an order placed in period s is some nonnegative integer-valued random variable L_s (we again deviate from the convention in that l_s has been already used to denote the expected marginal holding cost in period s). We assume that the random variables $L_1, \ldots, L_T$ are correlated, and in particular, that $s + L_s \leq t + L_t$ for each $s \leq t$. In other words, we assume that any order placed at time s will arrive no later than any other order placed after period s. This is a very common assumption in the inventory literature, usually described as ”no order crossing”.

Next we describe how to extend the dual-balancing policy and the analysis of the worst-case expected performance to this more general setting. The following discussion addresses the capacitated model, but as we have already seen, this is a generalization that captures the uncapacitated model as well. For each $t = 1, \ldots, T$, let S_t be the latest period for which an order placed in that period arrives at or before time t. In other words, $S_t := \max\{s : s + L_s \leq t\}$. Now modify the definition of the random variables W_{st} (for each $s \leq t$) to be

$$W_{st} := \min\{\mathbb{1}(s \leq S_t)\bar{Q}_s,\ (D([_{s,t]} - (X_s + Q_s + \sum_{j \in (s, S_t]} u_j))^+\}.$$

Similarly to the discussion in Section 2.5 above, we can write

$$W_{st} = \mathbb{1}(s \leq S_t)[(D_{[s,t]} - (X_s + Q_s + \sum_{j \in (s, S_t]} u_j))^+ - (D_{[s,t]} - (X_s + \sum_{j \in [s, S_t]} u_j))^+],$$

and

$$W_{st} = \mathbb{1}(s \leq S_t)[(D_t - NI_t - \sum_{j \in (s, S_t]} \bar{Q}_j)^+ - (D_t - NI_t - \sum_{j \in [s, S_t]} \bar{Q}_j)^+].$$

We again define the marginal backlogging cost in period s as $\bar{\Pi}_s = \sum_{t \geq s} p_t W_{st}$. Now modify the definition of $H_s(Q_s)$ so that now

$$H_s(Q_s) = \sum_{j=s+L_s}^{T} h_j(Q_s \; - \; (D_{[s,j]} - X_s)^+)^+.$$

It is straightforward to check that we can still express the cost of each feasible policy P as $\mathcal{C}(P) = \sum_t (H_t + \bar{\Pi}_t)$. In each period, we again balance the conditional expected marginal holding cost against the conditional expected marginal backlogging cost. It is readily verified that the same analysis described in Subsection 2.6.1 above is still valid. Note that the modified functions l_s and $\bar{\pi}_s$ can be significantly harder to evaluate.

Theorem 2.7.2 *The dual-balancing policy provides a performance guarantee of 2 for the uncapacitated and the capacitated variants of the periodic-review stochastic inventory control problem with stochastic lead times and non-crossing orders.*

2.7.3 Cost Transformation

In this section, we discuss in detail the cost transformation that enables us to assume, without loss of generality that, for each period $t = 1, \ldots, T$, we have $c_t = 0$ and $h_t,\ p_t \geq 0$. Consider any instance of the problem with cost parameters that imply no speculative motivation for holding inventory or backorders (as discussed in Section 2.2). We use a simple standard transformation of the cost parameters (see [56]) to construct an equivalent instance, with the property that for each period $t = 1, \ldots, T$, we have $c_t = 0$ and $h_t,\ p_t \geq 0$. The modified instance has the same set of optimal policies. Applying the dual-balancing policy to that instance will provide a policy that is feasible and also has a performance guarantee of at most 2 with respect to the original problem. We shall also show that this cost transformation can improve the performance guarantee of the dual-balancing policy in cases where the ordering cost is the dominant part of the overall cost. In practice this is often the case.

We now describe the transformation for the case with no lead time ($L = 0$) and $\alpha = 1$; the extension to the case of arbitrary lead time is straightforward. Recall that any feasible policy P satisfies, for each $t = 1, \dots, T$, $Q_t = NI_t - NI_{t-1} + D_t$ (for ease of notation we omit the superscript P). Using these equations we can express the ordering cost in each period t as $c_t(NI_t - NI_{t-1} + D_t)$. Now replace NI_t with $NI_t^+ - NI_t^-$, its respective positive and negative parts.

This leads to the following transformation of cost parameters. We let $\hat{c}_t := 0,\ \hat{h}_t := h_t + c_t - c_{t+1}$ $(c_{T+1} = 0)$ and $\hat{p}_t := p_t - c_t + c_{t+1}$. Note that the assumptions on the cost parameters $c_t,\ h_t$, and p_t discussed in Section 2.2, and in particular, the assumption that there is no speculative motivation for holding inventory or backorders, imply that $\hat{h}_t$ and $\hat{b}_t$ above are non-negative ($t = 1, \dots, T$). Observe that the parameters $\hat{h}_t$ and $\hat{b}_t$ will still be non-negative even if the parameters $c_t,\ h_t$, and p_t are negative and as long and the above assumption holds. Moreover, this enables us to incorporate into the model a negative salvage cost at the end of the planning horizon (after the cost transformation we will have non-negative cost parameters). It is readily verified that the induced problem is equivalent to the original one. More specifically, for each realization of the demands, the cost of each feasible policy P in the modified input decreases by exactly $\sum_{t=1}^{T} c_t d_t$ (compared to its cost in the original input). Therefore, any optimal policy for the modified input is also optimal for the original input.

Now apply the dual-balancing policy to the modified problem. We have seen that the assumptions on $c_t,\ h_t$ and p_t ensure that $\hat{h}_t$ and $\hat{p}_t$ are non-negative and hence the analysis presented above is valid. Let opt and $\bar{opt}$ be the optimal expected cost of the original and modified inputs, respectively. Clearly, $opt = \bar{opt} + E[\sum_{t=1}^{T} c_t D_t]$. Now the expected cost of the dual-balancing policy in the modified input is at most $2\bar{opt}$. Its cost in the original input is then at most $2\bar{opt} + E[\sum_{t=1}^{T} c_t D_t] = 2opt - E[\sum_{t=1}^{T} c_t D_t]$. This implies that if $E[\sum_{t=1}^{T} c_t D_t]$ is a large fraction of *opt*, then the performance guarantee of the expected cost of the dual-balancing policy might be significantly better than 2. For example, in case $E[\sum_{t=1}^{T} c_t D_t] \geq 0.5opt$ we can conclude that the expected cost of the dual-balancing policy would be at most $1.5opt$. It is indeed the case in many real life problems that a major fraction of the total cost is due the ordering cost. The intuition of the above transformation is that $\sum_{t=1}^{T} c_t D_t$ is a cost that any feasible policy must pay. As a result, we treat it as an invariant in the cost of any policy and apply the approximation algorithm to the rest of the cost.

In the case where we have a lead time L, we use the equations $Q_t := NI_{t+L} - NI_{t+L-1} + D_{t+L}$, for each $t = 1, \ldots, T-L$, to get the same cost transformation. The transformation for $\alpha > 1$ is also straight forward.

2.8 A Class of Myopic Policies for the Uncapacitated Model

No computationally tractable procedure is known for finding the optimal base-stock inventory levels for the periodic-review inventory control problem with correlated demands. As a result, various simpler heuristics have been considered in the literature. For the uncapacitated model, many researchers have considered a *myopic policy*. In the myopic policy, we follow a base-stock policy $\{R^{my}(f_t) : \ f_t \in \mathcal{F}_t\}$. For each period t and possible information set in period t, the target inventory level $R^{my}(f_t)$ is computed as the minimizer of a one-period problem. Specifically, in period $s = 1, \ldots, T-L$ we focus only on minimizing the expected immediate cost that is going to be incurred in this period (or in $s+L$ in the presence of a lead time L). In other words, the target inventory level $R^{my}(f_s)$ minimizes the expected ordering, holding and backlogging costs in period $s+L$, while ignoring the cost over the rest of the horizon (i.e., the cost over $(s+L, T+1]$). This optimization problem has been proven to be convex and hence easy to solve (see, for example, [56]). It is then possible to implement the myopic policy on-line, where in each period s, we compute the base-stock level based on the current observed information set f_s. For each period t and each $f_t \in \mathcal{F}_t$, the myopic base-stock level provides an upper bound on the optimal base-stock level (see [56] for a proof). The intuition is that the myopic policy underestimates the holding cost, since it considers only the one-period holding cost. Therefore, it always orders more units than the optimal policy. Clearly, this policy might not be optimal in general, though in many cases it seems to perform extremely well. Under rather strong conditions it might even be optimal (see [55, 21, 22, 31]). A natural question to ask is whether the myopic policy yields a constant performance guarantee for the uncapacitated periodic-review inventory control problem, i.e., its expected cost is always bounded by some constant times the optimal expected cost.

Next, we provide a negative answer to this question. We show that the expected cost of the myopic policy can be arbitrarily more expensive than the expected optimal cost, even for the case when the demands are independent and the costs are stationary. The example that we construct provides important intuition concerning the cases for which the myopic policy performs poorly. In addition, we describe an extended class of myopic policies that generalizes the myopic policy discussed above. It is interesting that this class of policies also provides a lower bound on the optimal base-stock levels. As was shown in [23], these lower and upper bounds combined with the balancing idea lead to improved balancing policies. The improved balancing policies have a worst-case performance guarantee of 2, and they seem to perform significantly better in practice.

2.8.1 Myopic Policy - Bad example

Consider the following set of instances parameterized by T, the number of periods. We have a per-unit ordering cost of $c = 0$, a per-unit holding cost $h = 1$ and a unit backlogging penalty $p = 2$. The demands are specified as follows, $D_1 \in \{0, 1\}$ with probability 0.5 for 0 and 1, respectively. For $t = 2, \ldots, T-1,\ D_t := 0$ with probability 1, and $D_T := 1$ with probability 1. The lead time is considered to equal 0, and $\alpha = 1$.

It is easy to verify that the myopic policy will order 1 unit in period 1 and that this will result an expected cost of $0.5T$. On the other hand, if we do not order in period 1, then the expected cost is 1. This implies that as T becomes larger the expected cost of the myopic policy is $\Omega(T)$ times as expensive as the expected cost of the optimal policy.

The above example indicates that the myopic policy may perform poorly in cases where the demand from period to period can vary a lot, and forecasts can go down. There are indeed many real-life situations, when this is exactly the case, including new markets, volatile markets or end-of-life products.

2.8.2 A Class of Myopic Policies

As we mentioned before, by considering only the one-period problem, the myopic policy described above underestimates the actual holding cost that each unit ordered in period t is going to incur. This results in base-stock levels that are higher than the optimal base-stock levels.

We now describe an alternative myopic base-stock policy that we call *a minimizing policy*. Recall the functions $l_s^P(q_s),\ \pi_s^P(q_s)$ defined in Section 2.4 for each period $s = 1, \ldots, T-L$, where $q_s \geq 0$. Since at each period s we know x_s, we can equivalently write $l_s^P(y_s - x_s),\ \pi_s^P(y_s - x_s)$, where $y_s \geq x_s$. We now consider in each period s the problem: minimize $(l_s^P(y_s - x_s) + b_s^P(y_s - x_s))$ subject to $y_s \geq x_s$, i.e., minimize the expected ordering and holding costs incurred by the units ordered in period s over $[s, T]$ and the backlogging cost incurred in period $s + L$, conditioned on some $f_s \in \mathcal{F}_s$. We have already seen that this function is convex in y_s. Observe that $l_s^P(y_s - x_s) - l_s^P(y_s)$ and $\pi_s^P(y_s - x_s) - \pi_s^P(y_s)$ do not depend on y_s for $y_s \geq x_s$. This gives rise to the following equivalent one-period problem: $\min_{y_s \geq x_s}(l_s^P(y_s) + \pi_s^P(y_s))$. That is, both problems have the same minimizer. It is also clear that the new minimization problem is also convex in y_s and is easy to solve, in many cases as easy as the one-period problem solved by the myopic policy described above. We note that the function we minimize was used by Chan and Muckstadt [5].

For each $t = 1, \ldots, T$ and $f_t \in \mathcal{F}_t$, let $R^M(f_t)$ be the smallest base-stock level resulting from the minimizing policy in period t, for a given observed information set f_t. We now show that for each period t and $f_t \in \mathcal{F}_t$, we have $R^M(f_t) \leq R^{OPT}(f_t)$, where $R^{OPT}(f_t)$ is the optimal base-stock level.

Theorem 2.8.1 *For each period t and $f_t \in \mathcal{F}_t$, we have $R^M(f_t) \leq R^{OPT}(f_t)$.*

Proof : Recall the dynamic programming based framework described in Section 2.3. Observe that for each state (x_t, f_t), we know that $R^{OPT}(f_t)$ is the optimal base-stock level that results from the optimal solution for the corresponding subproblem defined over the interval $[t, T]$. It is enough to show that the optimal solution for each such problem must be at least $R^M(f_t)$.

Assume otherwise, i.e., $R^{OPT}(f_s) < R^M(f_s)$ for some period s and for all optimal policies. Consider now the base-stock policy P with base-stock level $R^P(f_s) = R^M(f_s)$ for period s, and $R^P(f_t) := R^{OPT}(f_t)$ for each $t = s + 1, \ldots, T$ and $f_t \in \mathcal{F}_t$. We will show that P, starting from period s with observed information set f_s, has an expected cost that is smaller than the expected cost of the optimal solution. From Section 2.3 we know that the expected cost of each policy P can be expressed as $\sum_{t=s}^{T-L} E[H_t^P + \Pi_t^P]$. Now by the definition of $R^M(f_s)$ we know that

$$E[(H_s^P + \Pi_s^P)|f_s] < E[(H_s^{OPT} + \Pi_s^{OPT})|f_s].$$

Moreover, for each $t \in (s, T]$, the inventory position Y_t^P will always be at least Y_t^{OPT}, and therefore $E[\Pi_t^P|f_s] \leq E[\Pi_t^{OPT}|f_s]$. It is also clear that in each period $t \in (s, T]$, the Q_t^P units ordered by policy P in period t will always be a subset of the units ordered by OPT in this period. Therefore, for each $t = s+1, \ldots, T$, we have that $E[H_t^P|f_s] \leq E[H_t^{OPT}|f_s]$. This concludes the proof. ■

We now define a generalization that captures the myopic policy and the minimization policy as two special cases. For each $t = 1, \ldots, T-L$, we define a sequence of one-period problems for each $k_t = 0, \ldots, T-t$, each generates a corresponding base-stock level. Given k, we define the one-period problem that aims to minimize the expected ordering and holding cost incurred by the units ordered in period t over the interval $[t, t+L+k_t]$, and the expected backlogging cost in period $t+L$. In other words, the parameter k_t defines the length of the horizon considered in the one-period problem being solved in period t. For each sequence of $k_1, \ldots, k_T$, we get a corresponding *k-minimizing* policy. It is clear that if $k_t = 0$ for each t, we get the myopic policy and if $k_t = T-t$ we get the minimizing policy. Note again that the myopic and the minimizing policies provide an upper bound and lower bound, respectively, on the optimal base-stock levels.

2.9 Bounds and Improved Policies for the Capacitated Model

In this section, we consider two semi-myopic policies for the capacitated model that are easy to compute in an on-line manner. These policies provide respective lower bounds and upper bounds on the inventory levels of an optimal policy denoted by y_t^{OPT}, for each period $t = 1, \ldots, T$. We again believe that these bounds can be used effectively to improve existing inventory control policies for the capacitated model discussed in this work and other capacitated stochastic inventory models. Moreover, as in [23], we shall show that these policies provide bounds that are strong in the following sense. Each policy, that for some period t and some state f_t, has inventory level outside the range defined by the respective lower and upper bounds can be improved. In particular, there is another (modified) policy that in period t and state f_t admits an inventory level within the specified range, with expected cost no greater than the expected cost of the original policy. In other words, any policy that violates these respective bounds is dominated by another policy. We then follow [23] and construct an *improved dual-balancing policy* that incorporates these bounds. This policy has also a performance guarantee of 2 and as the computational study for the uncapacitated model in [23] suggests, we expect that it will have a better typical performance.

The policies we consider are called *lower-myopic* and *upper-myopic* respectively. In the lower-myopic policy, in each period s, conditioning on the observed information set f_s, we minimize the *sum* of the expected marginal holding cost of the units ordered in that period and the traditional expected backlogging costs. That is, in each period s, we minimize

$$l_s(q_s) + E[p_{s+L}(D_{[s,s+L]} - (x_s + q_s))^+ | f_s],$$

under the constraint $0 \leq q_s \leq u_s$. This is a convex function of q_s. This policy is identical to the minimizing policy described in Section 2.8 above for the uncapacitated model. Thus, it admits a base-stock policy. However, in the capacitated model it is possible that the actual minimizer will not be reachable. In this case we order up to capacity, and this provides a modified base-stock policy. In this section, we extend and generalize the proof of Theorem 2.8.1 to the capacitated model. In the upper-myopic policy, in each period s, again conditioning on f_s, we minimize the sum of the expected period holding cost and the expected marginal backlogging cost. Thus, we minimize

$$\bar{\pi}_s(q_s) + E[h_{s+L}(x_s + q_s - D_{[s,s+L]})^+ | f_s],$$

subject to $0 \leq q_s \leq u_s$, which is also convex in q_s. We shall show that this policy provides upper bounds on the inventory levels of an optimal policy. However, unlike the lower-myopic policy above, this policy is not a modified base-stock policy, since the minimizer depends on x_s, the inventory position at the beginning of period s. To the best of our knowledge this is a new way for deriving upper bounds on the inventory levels of an optimal policy in the capacitated model. We note that it is not clear whether the classical myopic policy (discussed above), where we minimize the expected period cost, provides any bounds for capacitated models. Another similar open question is how the policy, that in each period, minimizes the sum of the expected marginal holding cost and expected marginal backlogging cost, is related to an optimal policy.

Let Y_t^{LM} and Y_t^{UM} be the respective inventory position (after orders are placed) of the lower-myopic and the upper-myopic policies in period $t = 1, \ldots, T$. Further assume that Y_t^{LM} is always the respective smallest minimizer, and Y_t^{UM} is always the respective largest minimizer, resulting from the single-period problems defined above. Note that the inventory position levels depend on the specific state f_t, but for ease of notation we omit the indication of the state. The two semi-myopic policies described above can be implemented in an on-line manner, i.e., they can be computed independent of the action control in future periods. We shall show that for each evolution f_T these two policies provide lower and upper bounds on the inventory levels of an optimal policy, i.e., $Y_t^{LM} \leq Y_t^{OPT} \leq Y_t^{UM}$, with probability 1, for each $t = 1, \ldots, T$. Moreover, we shall show that, each non-dominated policy P, must have $Y_t^{LM} \leq Y_t^P \leq Y_t^{UM}$, for each

$t = 1, \ldots, T$.

The next two lemmas show that for each policy P that has, for some period s and state f_s, inventory position $y_s^P \notin [y_s^{LM}, y_s^{UM}]$, can be improved by a modified policy P' with $y_s^{P'} \in [y_s^{LM}, y_s^{UM}]$ and expected cost at most the expected cost of P. The proofs are similar to the ones in [23]. For the sake of simplicity, we consider a model with no lead time (the extensions to the case with $L > 0$ are straightforward).

Lemma 2.9.1 *Consider a feasible policy P, and suppose that for some period s and information set f_s, we have $y_s^P < y_s^{LM}$. Further assume that s is the earliest such period. Then the policy P' that follows P until period $s-1$, then orders up to y_s^{LM} in period s and again imitates P over the interval $(s, T]$, has expected cost smaller than the expected cost of P.*

Proof : Since P' follows P over $[1, s)$, we conclude that they incur exactly the same cost over that interval, and that they have the same inventory position $x_s \leq y_s^P < y_s^{LM}$. Moreover, s is assumed to be the first period with $y_s^P < y_s^{LM}$. Thus, P' can indeed order up to y_s^{LM}. Now over $(s, T]$ P' imitates P, that is, it orders nothing if $X_j^{P'} \geq Y_j^P$ and orders up to Y_j^P otherwise (for each $j \in (s, T]$). Moreover, the policy P' has ordered $q_s^{P'}$ units in period s. Consider the overall expected marginal holding cost of these units and the expected (traditional) backlogging cost incurred by policy P' in period s. By the definition of $q_s^{P'}$, it is clear that this is smaller than the expected marginal holding cost and expected (traditional) backlogging cost incurred by the policy P in period s. For each period $j \in (s, T]$, we know that $Y_j^{P'} \geq Y_j^P$ and $Q_j^{P'} \leq Q_j^P$, with probability 1. This implies that the backlogging incurred by policy P' over that interval is no greater than the backlogging cost incurred by policy P, and similarly, the marginal holding cost policy P' incurs over that interval is no greater than the respective marginal holding cost of policy P. The proof then follows. ∎

Lemma 2.9.2 *Consider a feasible policy P, and suppose that for some period s and information set f_s, we have $y_s^P > y_s^{UM}$. Further assume that s is the earliest such period. Then the policy P' that follows P until period $s-1$, then orders up to y_s^{UM} in period s and again imitates P over the interval $(s, T]$, has expected cost smaller than the expected cost of P.*

Proof : By identical arguments to the ones in Lemma 2.9.1, we conclude that P' and P incur the same cost over $[1, s)$ and that they have the same inventory position $x_s \leq y_s^{UM} < y_s^P$. The first inequality follows from the fact that s in the first period in which P has more inventory than the upper-myopic policy. Thus, P' can order up to Y_s^{UM}, and assume that it orders $q_s^{p'}$ units. Consider the overall expected marginal backlogging cost and expected period holding cost incurred in period s by policy P' (i.e., $\bar{\pi}_s^{P'}(q_s^{p'})$ defined in Section 2.6 above). By the definition of $q_s^{P'}$, we conclude that this is smaller than the respective expected cost incurred by policy P in period s. Now over $(s, T]$, P' again tries to imitate P, i.e., for each $j \in (s, T]$, it will order up to Y_j^P or up to the capacity u_j. Now let S' be the earliest (random) period after period s in which P' has reached $Y_{S'}^P$. Clearly, over $(S', T]$ the policies P' and P are again identical and hence, incur the same cost. Observe that, for each $j \in (s, S']$, we have $Y_j^{P'} \leq Y_j^P$ and $\bar{Q}_j^{P'} \leq \bar{Q}_j^P$, with probability 1. This implies that the expected holding cost and the expected marginal backlogging penalty incurred by policy P' over that interval are each no greater than the respective expected cost incurred by policy P. The proof then follows. ∎

Lemmas 2.9.1 and 2.9.2 imply the following corollary.

Corollary 2.9.3 *For each optimal policy and each evolution f_T, the lower-myopic and upper-myopic policies provide respective lower and upper bounds on the inventory levels of the optimal policy, i.e., $Y_t^{LM} \leq Y_t^{OPT} \leq Y_t^{UM}$, with probability 1, for each $t = 1, \ldots, T$.*

Now consider the *improved dual-balancing policy*. In each period s we still consider balancing expected marginal holding cost l_s against the expected marginal backlogging cost $\bar{\pi}_s$, and compute q'_s as described in Section 2.6 above. However, in each case, where the original dual-balancing rule orders up to a level below Y_s^{LM} or above Y_s^{UM}, we fix this decision by instead increasing the order up to Y_s^{LM} (or up to capacity) or decreasing it down to Y_s^{UM}, respectively. We next prove the following theorem.

Theorem 2.9.4 *The improved dual-balancing policy has a performance guarantee of 2.*

Proof : Observe that the in the improved dual-balancing policy it is not true anymore that, in each period t, the expected marginal holding cost is equal to the expected marginal backlogging cost. Now let Z_t be the maximum among the expected marginal holding cost and expected marginal backlogging cost, i.e.,

$$Z_t = \max\{E[H_t^B(Q'_t)|\mathcal{F}_t], E[\bar{\Pi}_t^B(Q'_t)|\mathcal{F}_t]\}.$$

Similar to Lemma 2.6.1, we now conclude that $E[\mathcal{C}(B)] \leq 2\sum_t E[Z_t]$.

Next we slightly modify the definition of the sets $\mathcal{T}_H$ and $\mathcal{T}_\Pi$ defined originally in Section 2.4 above. Now the set $\mathcal{T}_H$ will consist of periods t with $Y_t^B < Y_t^{OPT}$ and also periods with $Y_t^B = Y^{OPT} = Y_t^{LM} < Y_t^{UM}$. The set $\mathcal{T}_\Pi$ will consist of all other periods. Observe that the arguments used to prove Lemmas 2.4.2 and 2.6.2 are still valid. It is then sufficient to show that, for each $t \in \mathcal{T}_H$, we have $E[H_t^B(Q'_t)|\mathcal{F}_t] = Z_t$, and, for each $t \in \mathcal{T}_\Pi$, we have $Z_t = E[\Pi_t^B(Q'_t)|\mathcal{F}_t]\}$. This will imply that the arguments in the proof of Theorem 2.6.3 are still valid and the performance guarantee of the policy then follows.

Assume now that for some $t \in \mathcal{T}_H$ and some $f_t \in \mathcal{F}_t$, we have $E[H_t^B(q_t')|f_t] < z_t$. However, this can happen only if in that period the balancing rule was not followed and instead we have ordered only up to y_t^{UM}. This leads to contradiction since by Corollary 2.9.3, we know that $y_t^{OPT} \leq y_t^{UM}$.

Similarly, assume that for $t \in \mathcal{T}_\Pi$ and some $f_t \in \mathcal{F}_t$, we have $E[\Pi_t^B(Q_t')|\mathcal{F}_t] < z_t$. This can happen only if in that period the balancing rule was not followed and the order was increased up y_t^{LM} or up to capacity. However, this again leads to contradiction since by Corollary 2.9.3, we know that $y_t^B \leq y_t^{LM} \leq y_t^{OPT}$. This concludes the proof. ■

2.10 The Stochastic Lot-Sizing Problem

In this section, we change the previous model and in addition to the per-unit ordering cost, consider a fixed ordering cost K that is incurred in each period t with positive order (i.e., when $Q_t > 0$). For ease of notation, we will assume again, without loss of generality, that $c_t = 0$. We call this model the *stochastic lot-sizing problem*. The goal is again to find a policy that minimizes the expected discounted overall ordering, holding and backlogging costs. Naturally, this model is more complicated. Here we will assume that $L = 0$, $\alpha = 1$ and that in each period $t = 1, \ldots, T$, the conditional joint distribution I_t of $(D_t, \ldots, D_T)$ is such that the demand D_t is known deterministically (i.e., with probability 1). The underlying assumption here is that at the beginning of period t our forecast for the demand in that period is sufficiently accurate, so that we can assume it is given deterministically. A primary example is make-to-order systems.

As noted in Section 2.1, for many settings it is known that the optimal solution can be described as a set $\{(s_t, S_t) = (s_t(f_t), S_t(f_t))\}_t$. In each period t place an order if and only if the current inventory level is below s_t. If we place an order in period t, we will increase the inventory level up to S_t. We next describe a policy which we call the *triple-balancing policy* and denote by TB, and analyze its worst-case expected performance. Specifically, we show that its expected cost is at most 3 times the expected cost of the optimal solution. We note that in this case the policy and its analysis are identical for discrete and continuous demands.

2.10.1 The Triple-Balancing Policy

The policy follows two rules that specify when to place an order and how many units to order once an order is placed:

Rule 1: When to order. At the beginning of period s, we let s^* be the period in which the triple-balancing policy has last placed an order, i.e., s^* is the latest order placed so far. Thus, $s^* < s$, where $s^* = 0$ if no order has been placed yet. We place an order in period s if and only if, by not placing it in period s, the accumulated backlogging cost over the interval $(s^*, s]$ exceeds K. If we place an order, we update s^* and set it equal to s. Observe that since, in each period s, the conditional joint distribution I_s is such that D_s is known deterministically, this procedure is well-defined.

Rule 2: How much to order If we place an order in period $s < T$, then we focus on the holding cost incurred by the units ordered in s over the interval $[s, T]$, again using marginal cost accounting. We then order q_s^B units such that

$q_s^B := \max\{q_s : E[H_s^B(q_s)|f_s] \leq K\}$, where again $f_s \in \mathcal{F}_s$ is the current information set. That is, we order the maximum number of units as long as the conditional expectation of the holding cost that these units will incur over $[s, T]$, as seen from time period s, is at most K. In case $s = T$, we just order enough to cover all current back orders and the demand d_T. Observe that q_s^B must always be large enough to cover all of the backlogged units of demand over $(s^*, s]$. Hence, at the end of a period s in which an order was placed, there are no unsatisfied units of demand. We note that since for each $f_s \in \mathcal{F}_s$, the function $E[H_s^B(q_s)|f_s]$ is convex in q_s, it is relatively easy to compute q_s^B.

This concludes the description of the algorithm. Next we describe the analysis of the worst-case expected performance.

2.10.2 Analysis

Let N be the random variable of the number of orders placed by the triple-balancing policy. We next define a sequence of random variables $S_0, \ldots, S_{T+1}$. We let $S_0 = 0,\ S_{T+1} = T + 1$, and let S_i (for $i = 1, \ldots, T$) be the time period in which the i^{th} order of the triple-balancing policy was placed, or $T + 1$ if $N < i$ (i.e., the triple-balancing policy has placed fewer than i orders). Observe that $S_1, \ldots, S_T$ are random variables, which induce a partition of the time horizon. Consequently, we let Z_i, for each $i = 0, \ldots, T$, be the following random variable. If $S_i < T$, then Z_i is equal to the holding cost that the triple balancing policy incurs over $[S_i, S_{i+1})$ (denoted by H_i) plus the backlogging and ordering costs it incurs over $(S_i, S_{i+1}]$. If $S_i \geq T$, then $Z_i = 0$. Similarly, we define the set of variables $Z'_0, \ldots, Z'_{T+1}$ with respect to the cost of OPT over the corresponding intervals induced by the orders of the triple-balancing policy. It is clear that $\mathcal{C}(B) = \sum_{i=0}^{T} Z_i \cdot 1(S_i < T)$ and $C(OPT) = \sum_{i=0}^{T} Z'_i \cdot 1(S_i < T)$. We first develop a lower bound on the expected cost of OPT using the expectation of the random variable N.

Lemma 2.10.1 *For each instance of the stochastic lot-sizing problem with correlated demand the expected cost of an optimal policy OPT is at least $KE[N]$.*

Proof : We have already observed that $\mathcal{C}(OPT) = \sum_{i=0}^{T} Z_i' \cdot 1(S_i < T)$. Using again the linearity of expectation and conditional expectation, we can write,

$$E[\mathcal{C}(OPT)] = \sum_{i=0}^{T} E[1(S_i < T)E[Z_i'|S_i, \mathcal{F}_{S_i}]] \geq$$
$$\sum_{i=0}^{T} E[1(S_i < T)E[Z_i' \cdot 1(S_{i+1} \leq T)|S_i, \mathcal{F}_{S_i}]]$$

Next we show that for each $i = 0, \ldots, T$, we have that,

$$E[Z_i' \cdot 1(S_{i+1} \leq T)|S_i, \mathcal{F}_{S_i}] \geq K \cdot Pr(S_{i+1} \leq T|S_i, \mathcal{F}_{S_i})$$

Conditioned on some $S_i = s_i$ and $f_{s_i} \in \mathcal{F}_{s_i}$, we know d_{s_i} (where s_i is the realization of S_i). As a result, we also know the inventory levels of OPT and the triple-balancing policy at the end of period s_i deterministically. Therefore, exactly one of the following 2 cases must apply:

Case 1: At the end of period s_i, the inventory level of OPT is at most the inventory level of the triple-balancing policy, i.e., $y_{s_i}^{OPT} \leq y_{s_i}^{TB}$. Now either OPT places an order over $(s_i, S_{i+1}]$ and hence incurs a cost of at least K over this interval, or it does not; then, unless s_i is the last order of the triple-balancing policy, it will incur backlogging cost of at least K.

Case 2: At the end of period s_i, the inventory level of OPT is strictly larger than the inventory level of the triple-balancing policy, i.e., $y_{s_i}^{OPT} > y_{s_i}^{TB}$. However, by the construction of the triple-balancing policy, we know that if OPT has more physical inventory, then the expected holding cost it will incur over $[s_i, S_{i+1})$ is at least K.

We conclude that in both cases, $Z_i' \cdot 1(S_{i+1} \leq T)|s_i, f_{s_i})] \geq K \cdot 1(S_{i+1} \leq T|s_i, f_{s_i})$. Taking expectation we have $E[Z_i' \cdot 1(S_{i+1} \leq T)|S_i, \mathcal{F}_{S_i}] \geq K \cdot Pr(S_{i+1} \leq T|S_i, \mathcal{F}_{S_i})$.

This implies that

$$E[\mathcal{C}(OPT)] \geq K \cdot E[\sum_{i=0}^{T} 1(S_i < T) \cdot Pr(S_{i+1} \leq T | S_i, F_{S_i})] =$$

$$K \cdot E[\sum_{i=0}^{T} E[1(S_i < T) \cdot 1(S_{i+1} \leq T) | S_i, F_{S_i}]] = K \cdot E[N].$$

■

To finish the analysis we next show that the expected difference between the cost of the triple-balancing policy (denoted by TB) and the cost of the optimal policy is at most $2KE[N]$.

Lemma 2.10.2 *For each instance of the problem, we have* $E[\mathcal{C}(TB) - \mathcal{C}(OPT)] \leq 2KE[N]$.

Proof : Clearly,

$$E[\mathcal{C}(TB) - \mathcal{C}(OPT)] = E[\sum_{i=0}^{T} (Z_i - Z_i') \cdot 1(S_i < T)] =$$

$$\sum_{i=0}^{T} E[1(S_i < T) \cdot E[(Z_i - Z_i') | S_i, \mathcal{F}_{S_i}]].$$

We next bound $E[(Z_i - Z_i') | S_i, \mathcal{F}_{S_i}]$ for each $i = 0, \ldots, T$. For $i = 0$, it is clear that the holding costs that the TB policy and OPT incur over $[s_0, S_1)$ are identical (this cost is due initial inventory that exists at the beginning of the horizon). Also observe that the backlogging and ordering costs of the TB policy over $(S_0, S_1]$ are at most K if $S_1 = T + 1$ and at most $2K$ otherwise. In the latter case, we conclude that OPT either placed an order on the interval $(S_0, S_1]$ or incurred backlogging cost of at least K. Hence, $Z_0 - Z_0' \leq K$.

For each $i = 1, \ldots, T$, we condition on some s_i and $f_{s_i} \in \mathcal{F}_{s_i}$. We then know what are $y_{s_i}^{TB}$ and $y_{s_i}^{OPT}$ deterministically. We now claim that:

$$(Z_i - Z_i')|s_i, f_{s_i} \leq 1(y_{s_i}^{TB} \leq y_{s_i}^{OPT}) \cdot (K + 1(S_{i+1} \leq T|s_i, f_{s_i}) \cdot K) +$$
$$1(y_{s_i}^{TB} > y_{s_i}^{OPT}) \cdot (H_i|s_i, f_{s_i} + 1(S_{i+1} \leq T|s_i, f_{s_i}) \cdot K).$$

In first case where $y_{s_i}^{TB} \leq y_{s_i}^{OPT}$ we know that OPT will incur over $[s_i, S_{i+1})$ at least as much holding cost as the TB policy. By the construction of the algorithm we know that the TB policy will not incur more than K backlogging cost and will place at most one order over $(s_i, S_{i+1}]$. In the second case where $y_{s_i}^{TB} > y_{s_i}^{OPT}$ we know that the ordering cost and backlogging costs of OPT over $(s_i, S_{i+1}]$ are at least K, which is more than the backlogging cost the TB policy incurs on that interval. In addition, TB will incur holding cost $H_i|s_i, f_{s_i}$ over $[s_i, S_{i+1})$ and will place at most one order over $(s_i, S_{i+1}]$. Taking expectation of both sides we conclude that:

$$E[(Z_i - Z_i')|S_i, \mathcal{F}_{S_i}] \leq E[1(y_{s_i}^{TB} \leq y_{S_i}^{OPT}) \cdot (K + 1(S_{i+1} \leq T) \cdot K)|S_i, \mathcal{F}_{s_i}] +$$
$$E[1(y_{S_i}^{TB} > y_{S_i}^{OPT}) \cdot (H_i + 1(S_{i+1} \leq T) \cdot K)|S_i, \mathcal{F}_{S_i}] \leq E[K + 1(S_{i+1} \leq T)|S_i, \mathcal{F}_{S_i}].$$

The last inequality is by the construction of the algorithm ($E[H_i|s_i, f_{s_i}] \leq K]$ for each $S_i = s_i$ and $f_{s_i} \in \mathcal{F}_{s_i}$). This implies that for each $i = 2, \ldots, T$, we have

$$E[(Z_i - Z_i') \cdot 1(S_i < T)] = E[1(S_i < T) \cdot E[Z_i - Z_i'|S_i, \mathcal{F}_{S_i}]] \leq$$
$$E[K + 1(S_{i+1} \leq T)].$$

Finally, we have that:

$$E[\sum_{i=0}^{T}(Z_i - Z_i') \cdot 1(S_i < T)] \leq$$
$$K + K \cdot E[\sum_{i=1}^{T} 1(S_i < T)] + K \cdot E[\sum_{i=1}^{T+1} 1(S_i < T) \cdot 1(S_{i+1} \leq T)] =$$
$$K + K \cdot E[N] + K \cdot (E[N] - 1) = 2KE[N].$$

■

As a corollary of Lemmas 2.10.1 and 2.10.2, we get the following theorem.

Theorem 2.10.3 *For each instance of the stochastic lot-sizing problem, the expected cost of the triple-balancing policy is at most 3 times the expected cost of an optimal policy.*

2.11 Conclusions

In this chapter we have proposed a new approach for devising provably good policies for stochastic inventory control models with time dependent and correlated demand. These models are known to be hard, in the sense that computing optimal policies is usually intractable and in many cases even computing a good policies is a challenging task. In turn, our approach leads to policies that are simple computationally and conceptually and provide constant performance guarantees on the worst-case expected behavior of these policies.

We note that all of the results described in the chapter can be extended under rather mild conditions to the counterpart models with infinite horizon, where the goal is to minimize the expected average or discounted cost.

We think it would be an interesting challenge to extend the ideas introduced in this work to more complicated inventory models, such as multi-echelon and/or multi-item models. These issues will be addressed in future work.

It would also be important to establish a more rigorous analysis of the computational hardness of these models. As far as we know there does not exist any rigorous proof of that kind.

Chapter 3

Near-Optimal Sample-based Policies for Single-Period and Multiperiod Newsvendor Models

3.1 Introduction

In this chapter, we address two fundamental models in stochastic inventory theory, the *single-period newsvendor model* and the *multiperiod newsvendor model*, under the assumption that the explicit demand distributions are not known and that the only information available is a set of independent samples drawn from the true distributions. Under the assumption that the demand distributions are specified explicitly, these models are well-studied and usually easy to solve. However, in most real-life scenarios, the true demand distributions are not available or they are too complex to work with. Usually, the information that is available comes from historical data, simulation setting and from forecasting and market analysis of future trends in the demands. Thus, we believe that a *sample-driven* algorithmic framework is very attractive, both in practice and in theory. In this chapter, we shall describe how to compute *sample-based policies*, that is, policies that are computed based only on observed samples of the demands without *any* access to, or assumptions on, the true demand distributions. Moreover, we shall prove that the quality (expected cost) of these policies is very close to that of the optimal policies, which have *full* access to the explicit demand distributions.

In the single-period newsvendor model, a random demand for a single commodity occurs in a single period. At the beginning of the period, *before* the the actual demand is observed, we decide how many units of the commodity to order. Next, the actual demand is observed and is satisfied to the maximum extent possible from the units that were ordered. At the end of the period, a per-unit *holding cost* is incurred for each unused unit of the commodity and a per-unit *lost-sales* penalty cost is incurred for each unmet unit of demand. The goal is to minimize the total expected cost. This model is usually easy to solve *if* the demand distribution is specified explicitly. However, we are not aware of any optimization algorithm with analytical error bounds in the case where only samples are available.

For the newsvendor model, we take one of the most common approaches to stochastic optimization models that is also used in practice, that is, we solve the *sample average approximation* (SAA) counterpart. The original objective function is the expectation of some random function taken with respect to the true underlying probability distributions. Instead, in the SAA counterpart the objective function is the average value over finitely many independent samples that are drawn from the probability distributions either by means of *Monte Carlo* sampling or based on available historical data (see [45] for details). In the newsvendor model the samples will be drawn from the (true) demand distribution and the objective value of each order level will be computed as the average of its cost with respect to each one of the samples of demand. The SAA counterpart of the newsvendor problem is extremely easy to solve.

We also provide a novel analysis regarding the number of samples required to guarantee that, with a specified *confidence probability*, the expected cost of an optimal solution to the SAA counterpart has a small specified relative error. Here small relative error means that the ratio between the expected cost of the optimal solution to the SAA, with respect to the *original* objective function, and the optimal expected cost (of the original model) is very close to 1. The upper bounds that we establish on the number of samples required are general, easy to compute and apply to *any* demand distribution with finite mean. In particular, neither the algorithm nor its analysis require *any* other assumption on the demand distribution. The bounds depend on the specified confidence probability and the relative error mentioned above, as well as on the ratio between the per-unit holding and lost-sales penalty costs. However, they are *completely independent* of the demand distribution. Conversely, our results imply what kind of guarantees one can hope for, given historical data with fixed size. The analysis has two novel aspects. First, it is not based on approximating the objective function and its value, but on using first-order information, that is, on stochastically evaluating one-sided derivatives. This is motivated by the fact that the newsvendor cost function is convex and hence, optimal solutions can be characterized in a compact way through first-order information. The second novel aspect is that we establish a connection between first-order information and bounds on the the relative error of the objective value. Moreover, the one-sided derivatives of the newsvendor cost function are nicely bounded. Thus, the well-known Hoeffding inequality [20] implies that they can be estimated accurately with a bounded number of samples.

In the multi-period newsvendor model, there is a *sequence* of independent (not necessarily identical distributed) random demands for a single commodity over a discrete planning horizon of a finite number of periods that need to be satisfied. At the beginning of each period we can place an order for any number of units and this order is assumed to arrive after a (fixed) *lead time* of several periods. Only then do we observe the actual demand in the period. Excess inventory at the end of a period is carried to the next period incurring a per-unit holding cost. Symmetrically, each unit of unsatisfied demand is carried to the next period incurring a per-unit *backlogging penalty* cost. The goal is to find an ordering policy with minimum total expected cost. The multi-period model can be formulated as a tractable dynamic program, where at each stage we minimize a single-variable convex function. Thus, the optimal policies can be efficiently computed, again *if* the demand distributions are specified explicitly (see [56] for details).

As was pointed in [48], the SAA counterparts for multistage stochastic models seem to be very hard to solve in general. Instead of solving the SAA counter part of the multiperiod model, we propose a dynamic programming framework that is significantly different algorithmically from previous sample-based algorithms. The approximate policy is computed in stages backward in time via a dynamic programming approach. The main challenge here arises from the fact that in a backward dynamic programming framework, the optimal solution in each stage heavily depends on the solutions already computed in the previous stages of the algorithm. Therefore, the algorithm maintains a *shadow dynamic program* that 'imitates' the *exact dynamic program* which would have been used to compute the exact optimal policy, if the explicit demand distributions were known. That is, in each stage, we consider a subproblem that is similar to the corresponding subproblem in the exact dynamic program. However, this subproblem is defined with respect to the approximate solutions already computed by the algorithm in the previous stages, instead of the optimal solutions that define the corresponding subproblem in the exact dynamic program. The algorithm is carefully designed to maintain (with high probability) the convexity of each one of the subproblems that are being solved throughout the execution of the algorithm. Thus, in each stage there is again a single-variable convex minimization problem that is solved approximately. As in the newsvendor case, first-order information is used in the optimization. To do so we use some general structural properties of these functions to establish a key lemma (Lemma 3.3.6 below) that again relates first-order information of these functions to relative error with respect to their optimal objective value. We believe that this lemma will have additional applications in approximating other classes of stochastic dynamic programs. Since the one-sided derivatives of these functions are again nicely bounded, the Hoeffding inequality again implies that they can be estimated using only bounded number of samples. The analysis indicates that the relative error of the approximation procedure in each stage of the algorithm is carefully controlled, which leads to policies that, with high probability, have overall small relative error. The upper bounds on the number of samples required are again easy to compute and independent of the specific demand distributions. In particular, they grow as a polynomial in the number of periods. To the best of our knowledge, this is the first result of this kind for multistage stochastic models and for stochastic dynamic programs. In particular, the existing approaches to approximating stochastic dynamic programs do not admit constant worst-case guarantees of the kind discussed in this work (see [50]).

We note that the bounds on the number of samples established in this work should be considered only as upper bounds. It is very likely that in practice a significantly smaller number will be sufficient. In addition, in cases where there is more information on the demand distributions (e.g., moments), it might be possible to establish tighter theoretical bounds.

We believe that this work sets the foundations for additional sample-based algorithms for stochastic inventory models and stochastic dynamic programs with analyzed performance guarantees. In particular, it seems very likely, that the same algorithms and analysis as described in this chapter will be applicable to a multi-period model where there exists Markov modulated demand process.

We next relate our work to the existing literature. In this chapter, we consider the *black box* model that was used by Gupta, Pál, Ravi and Sinha [19]. More specifically, we assume that the information about the demands is available through a black box that can generate, on request, independent samples from the demand distributions.

The problem of unavailable or complex probability distributions within the context of stochastic optimization models has been extensively addressed in the literature. Solving the sample average approximation (SAA) is an approach that has been explored both theoretically and in practice; this approach immediately raises several natural questions. First, is it possible to efficiently solve this model? Secondly, what is the quality of an optimal solution to the SAA model with respect to the original objective function and how does it depend on the number of samples that define it? Thirdly, how is the set of optimal solutions of the SAA model related to the set of optimal solutions of the initial problem?

The latter two questions have been addressed in several recent papers. Kleywegt, Shapiro and Homem-De-Mello [28] have considered the SAA in a general setting of two-stage discrete stochastic optimization models (see [41] for discussion on two-stage stochastic models). They have shown that, for a sufficiently large number of samples, the optimal value of the SAA problem converges to the optimal value of the original problem with probability 1. They have also used large-deviation results to show that the *additive* error of an optimal solution to the SAA model (i.e., the difference between its objective value and the optimal objective value of the original problem) converges to zero with probability 1 for a sufficiently large number of samples. However, their bounds on the number of samples required, depend on the variability of the objective function. Hence, these bounds might be hard to compute and they can be very loose. Sahpiro, Homem-De-Mello and Kim [47, 46] have again focused on two-stage stochastic models and considered the probability of the event that an optimal solution to the SAA model is in fact an optimal solution to the original problem. Under the assumption that the probability distributions have finite support and that the original problem has a unique optimal solution, they have again used large-deviation results to show that this probability converges to 1 exponentially fast as the number of samples grows. Swamy and Shmoys [54] have considered a class of two-stage stochastic linear programs and bound the number of samples required to guarantee that, with specified high confidence probability, the optimal solution to the corresponding SAA model has a small specified relative error. Like ours, their bounds are easy to compute and do not depend on the specific underlying probability distributions. Charikar, Chekuri and Pál [6] have proposed a simpler proof and extended the class of problems to which this result applies. However, both of these results do not seem to capture the models we consider in this work. In particular, the newsvendor model admits a non-linear objective function and the multiperiod version is by nature a multistage stochastic problem. For multistage stochastic linear programs Swamy and Shmoys [54] have shown that the SAA model is still effective in providing a good solution to the original problem, but the bounds on the number of samples and the running time of the algorithms grow exponentially with the number of stages.

Infinitesimal perturbation analysis is a sample-based technique that has been extensively explored in the context of solving stochastic supply chain models (see [17], [18] and [27] for several examples). Usually, this method works as follows. First, the performance measures of interest (e.g., inventory level in each period, the number of backlogged units in each period, etc) are expressed as a function of the demands and the decision parameters. Then a sample-path approach is used to stochastically evaluate the derivatives of the different performance measures as a function of the decision parameters. More specifically, independent samples are drawn from the demand distributions, where each sample consists of a sequence (a path) of demands over the planning horizon. The cost of each set of decision parameters is then evaluated by averaging over all the sampled paths. The samples are drawn only once and the recursion is used to efficiently evaluate the partial derivatives of the relevant performance measures as a function of the decision parameters. In most cases it is possible to show that these gradient estimators are *consistent estimators*, i.e., they converge to the real value of the derivative with probability 1 for a sufficiently large number of samples. This gradient estimation technique is usually incorporated into *gradient search* methods (see [18] for details). The IPA-based methods seem very effective in practice for computing policies that can be characterized in a compact way with a relative small set of decision parameters. However, to the best of our knowledge, there is no analysis regarding the number of samples required to guarantee, with high confidence probability, a solution with small relative (or additive) error.

The *robust* or the *min-max* optimization approach is yet another way to address the uncertainty regarding the exact demand distributions in supply chain models, see for example, Bersimas and Thiele [2] and Gallego, Ryan and Simchi-Levi [15] (this approach was applied to many other stochastic optimization models). In this approach, instead of fitting to the data a unique distribution we allow a *family* of distributions. The assumption is that the true distribution of the demand belongs to that family of distributions. The specification of the relevant family of distributions can be based, for example, on moments fitting or on an uncertainty set that defines only the support of the distribution. The cost of each policy is usually evaluated with respect to the worst distribution within the specified family and the goal is to find the policy with the best worst case. In particular, we mention the work Scarf in [44], of Gallego and Moon [13] and of Perakis and Roels [38] on the maximization variant of the newsvendor problem. We note that under these approaches we no longer consider the original objective, that is, minimizing the expected cost. Moreover, the resulted solution can be very conservative. Hence, the main issue in applying these approaches is how to define the family of allowed distributions or the uncertainty set in a way that captures the uncertainty but does not lead to conservative solutions.

The rest of the chapter is organized as follows. In Section 3.2 we discuss the single-period newsvendor model, then in Section 3.3 we proceed to discuss the multiperiod model. Then in Section 3.4 we consider the case of approximating myopic policies. Finally, in Section 3.5 we provide a proof for a general multidimensional version of Lemma 3.3.6.

3.2 Newsvendor Problem

In this section, we consider the minimization variant of the classical single-period newsvendor problem. A random demand D for a single commodity occurs in a single period. At the beginning of the period, before the the demand is observed, we decide how many units of the commodity to order to satisfy the (random) demand D. The y units ordered arrive instantaneously and only then is the actual demand d (the realization of D) observed. If too many units were ordered, i.e., $y > d$, a per-unit holding cost $h > 0$ is incurred for each unit of excess inventory, i.e., the overall cost is $h(y - d)$. On the other hand, if not enough units were ordered, i.e., $d > y$, a per-unit lost-sales penalty $b > 0$ is incurred for each unit of unsatisfied demand, i.e., the overall cost is $b(d - y)$. The goal is to minimize the expected cost $C(y) = E[h(y - D)^+ + b(D - y)^+]$ (where $x^+ = \max(x, 0)$ and the expectation is taken with respect to the random demand D). Note that, without loss of generality, the per-unit ordering cost is assumed to be equal to 0. This assumption will be discussed at the end of this section. The newsvendor problem is a well-studied model and much is known about the properties of its objective function C and its optimal solutions. For the sake of completeness, we next discuss some of these properties, highlighting the ones used in this section.

3.2.1 Optimality Conditions

Observe that for each realization d of D, the function $h(y-d)^+ + b(d-y)^+$ is a (piecewise linear) convex function of y. Therefore, the cost function $C(y)$ is also a convex function of y, since it is the expectation of a convex function. The general characterization of optimal solutions to convex minimization problems (i.e., of a global minimizer) relies on the notion of a *subgradient* (for details, the reader is referred to [40]).

Definition 3.2.1 *Given a function $f : \mathbb{R}^m \mapsto \mathbb{R}$, a vector $w \in \mathbb{R}^m$ is called a* subgradient *of f at $u \in \mathbb{R}^m$ if the inequality $f(v) \geq f(u) + w \cdot (v - u)$ holds for each $v \in \mathbb{R}^m$.*

Note that a function f might not have a subgradient at u or might not have a unique subgradient. Let $\partial f(u)$ denote the set of subgradients of f at u (where again, $\partial f(u)$ can be empty). It is readily verified that $\partial f(u)$ is a convex set. For a finite convex function f, the set $\partial f(u)$ is nonempty for each $u \in \mathbb{R}^m$, and it consists of a unique subgradient if and only if f is differentiable at u.

Throughout this chapter, we are going to use the following characterization of a global minimum of a convex function.

Theorem 3.2.2 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be a convex function. Then $u \in \mathbb{R}^m$ is a global minimizer (i.e., $f(u) \leq f(v)$ for each $v \in \mathbb{R}^m$) if and only if $\bar{0} \in \partial f(u)$ (where $\bar{0}$ is the vector of all zeros).*

The objective function, $C(y)$, considered in the newsvendor problem is a single-variable finite convex function. Therefore, the set $\partial C(u)$ is an interval. It is readily verified that the endpoints of this interval are the right-hand and left-hand derivatives of

C at u, denoted by $C^r(u)$ and $C^l(u)$, respectively. Observe that since C is convex, it is always true that $C^l(u) \le C^r(u)$. Indeed, if C is differentiable at u, then $C^r(u) = C^l(u) = C'(u)$ and $\partial C(u)$ is a singleton. It is also clear that $0 \in \partial C(u)$ if and only if $C^r(u) \ge 0$ and $C^l(u) \le 0$. This provides an optimality criterion equivalent to Theorem 3.2.2 above.

For the newsvendor model, it is easy to derive explicit expressions for the right-hand and left-hand derivatives of C. Using a standard dominated convergence theorem (see [3]), the order of integration (expectation) and the limit (derivatives) can be interchanged, and the one-sided derivatives of C can be expressed explicitly. We get that $C^r(y) = -b + (b+h)F(y)$, where $F(y) := Pr(D \le y)$ is the CDF of D, and that $C^l = -b + (h+b)Pr(D < y)$. The right-hand and the left-hand derivatives are equal at all the continuity points of F. In particular, if F is continuous, then C is continuously differentiable with $C'(y) = -b + (b+h)F(y)$. It is therefore relatively easy to check whether the latter optimality criterion is satisfied (i.e., whether $C^r(u) \ge 0$ and $C^l(u) \le 0$).

Moreover, if the distribution of the demand D is given explicitly, then it is usually easy to compute an optimal solution. More specifically, this optimal solution, denoted by y^*, is the $\frac{b}{b+h}$ quantile of the distribution of D, i.e., $y^* = \inf\{y : F(y) \ge \frac{b}{b+h}\}$. Observe that the definition of y^* is equivalent to the optimality criterion above, i.e., it is equivalent to the inequalities $C^r(y^*) \ge 0$ and $C^l(y^*) \le 0$. Finally, we note that all of the above is valid for any demand distribution D with $E[|D|] < \infty$, including cases when negative demand is allowed. It is clear that in the case where $E[|D|] = \infty$, the problem is not well-defined, because any ordering policy will incur infinite expected cost.

3.2.2 Sample Average Approximation

In most real-life scenarios, the demand distribution is not known and the only thing available is data from past periods. Consider a model where instead of an explicitly specified demand distribution there is a black box that generates independent samples of the demand drawn from the true distribution of D. Assuming that the demands in all periods are independent and identically distributed (i.i.d) random variables, distributed according to D, this will correspond to available data from past periods or to samples coming from a simulation procedure or from a marketing experiment that can be replicated. Note that there is no assumption on the actual demand distribution. In particular, there is no parametric assumption, and there are no assumptions on the existence of higher moments (beyond the necessary assumption that $E[|D|] < \infty$). Under these assumptions, a natural question that arises is how many samples from the black box or equivalently, what is the size of past data that are required to be able to find a provably good solution to the original newsvendor problem. By a provably good solution, we mean a solution with expected cost at most $(1 + \epsilon')C(y^*)$ for a specified $0 < \epsilon'$, where $C(y^*)$ is the optimal expected cost under *full knowledge* of the demand distribution D.

Our approach is based on the natural and common idea of solving the *sample average approximation (SAA)* counterpart of the problem. Suppose that we have N independent samples of the demand D, denoted by $d^1, \ldots, d^N$. The SAA counterpart is defined in the following way. Instead of using the demand distribution of D that is not available, we assume that each one of the samples of D occurs with the same probability of $\frac{1}{N}$. Now define the newsvendor problem with respect to the induced *empirical* distribution. In other words, the problem is defined as

$$\min_{y \geq 0} \hat{C}(y) := \frac{1}{N} \sum_{i=1}^{N} (h(y - d^i)^+ + b(d^i - y)^+).$$

Throughout the chapter we use *hat* to distinguish between deterministic functions such as C above, that are defined by taking expectations with respect to the underlying demand distributions, and their SAA counterparts which are (like $\hat{C}$) random variables because they are functions of the *random samples* being drawn from the demand distributions. More generally, the symbol *hat* will be used to denote any quantity that depends on the random samples from the demand distributions. In addition, unless stated otherwise, all expectations are taken with respect to the true underlying demand distributions.

Let $\hat{Y} = \hat{Y}(N)$ denote the optimal solution to the SAA counterpart. Note again that $\hat{Y}$ is a random variable that is dependent on the specific N (independent) samples of D. Clearly, for each given N samples of the demand D, $\hat{y}$ (the realization of $\hat{Y}$) is defined to be the $\frac{b}{b+h}$ quantile of the sample, i.e., $\hat{y} = \inf\{y : \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(d^i \leq y) \geq \frac{b}{b+h}\}$ (where $\mathbb{1}(d^i \leq y)$ is the respective indicator function which is equal to 1 exactly when $d^i \leq y$). It follows immediately that $\hat{y} = \min_{1 \leq j \leq N}\{d^j : \frac{1}{N}\sum_{i=1}^{N} \mathbb{1}(d^i \leq d^j) \geq \frac{b}{b+h}\}$. Hence, given the demand samples $d^1, \ldots, d^N$, the optimal solution to the SAA counterpart, $\hat{y}$, can be computed very efficiently by finding the $\frac{b}{b+h}$ quantile of the samples.

This makes the SAA counterpart very attractive to solve. However, the natural question is how the SAA counterpart is related to the original problem as a function of the number of samples N. Consider any specified *accuracy level* $\epsilon' > 0$ and a *confidence level* $1 - \delta'$ (where $0 < \delta' < 1$). We will show that there exists a number of samples $N = N(\epsilon', \delta', h, b)$ such that, with probability at least $1 - \delta'$, the optimal solution to the SAA counterpart defined on N samples, has expected cost $C(\hat{Y})$ at most $(1 + \epsilon')C(y^*)$. Note that we compare the expected cost of $\hat{y}$ (the realization of $\hat{Y}$) to the optimal expected cost, where there is full access to the true distribution of D. As we will show, the number N of required samples is polynomial in $\frac{1}{\epsilon'}$, $\log(\frac{1}{\delta'})$ and is also dependent on the minimum among the values $\frac{b}{b+h}$ and $\frac{h}{b+h}$ (that define the optimal solution y^* above).

However, it is completely *independent* of the demand distribution D.

The first issue that needs to be addressed is how many samples are required to make it very likely (i.e., with high probability) that $\hat{y}$, the realization of $\hat{Y}$, is 'close' to the real $\frac{b}{b+h}$ quantile (and the optimal solution) y^*. Here 'close' does not mean necessarily that $|y^* - \hat{y}|$ is small but that $F(\hat{y}) = Pr(D \leq \hat{y})$ is 'close' to $F(y^*)$. Recall, that $F(y) := Pr(D \leq y)$ (for each $y \in \mathbb{R}$), and let $\bar{F}(y) := Pr(D \geq y) = 1 - F(y) + Pr(D = y)$ (where here we depart from the traditional notation). Observe that by the definition of y^* as the $\frac{b}{b+h}$ quantile of D, $F(y^*) \geq \frac{b}{b+h}$ and $\bar{F}(y^*) \geq \frac{h}{b+h}$. The following definition provides a precise notion of what we mean by 'close' above.

Definition 3.2.3 *Let $\hat{y}$ be some realization of $\hat{Y}$ and let $\epsilon > 0$. We will say that $\hat{y}$ is* ϵ-accurate *if $F(\hat{y}) \geq \frac{b}{b+h} - \epsilon$ **and** $\bar{F}(\hat{y}) \geq \frac{h}{b+h} - \epsilon$.*

Once again, the above definition can be related to a statement on the right-hand and left-hand derivatives of C at $\hat{y}$. Observe that $Pr(D < y) = 1 - \bar{F}(y)$. It is straightforward to verify that if $\hat{y}$ is $\epsilon-$accurate, then $C^r(\hat{y}) \geq -\epsilon(b+h)$ and $C^l(\hat{y}) \leq \epsilon(b+h)$. This implies that there exists a subgradient $r \in \partial C(\hat{y})$ such that $|r| \leq \epsilon(b+h)$. Intuitively, this implies that, for ϵ sufficiently small, 0 is 'almost' a subgradient at $\hat{y}$, and hence $\hat{y}$ is 'close' to being optimal. This intuitive observation is made rigorous in Lemma 3.2.6 and Theorem 3.2.7 below.

Next we wish to establish upper bounds on the number of samples N required in order to guarantee that $\hat{y}$, the realization of $\hat{Y}$, is $\epsilon-$accurate with high probability (for each specified $\epsilon > 0$ and confidence probability $1-\delta$). In doing that we use the following variant of the well-known Hoeffding inequality.

Theorem 3.2.4 *(Hoeffding Inequality [20]). Let $X^1, \ldots, X^N$ be i.i.d. random variables such that $X^1 \in [\alpha, \beta]$ (i.e., $Pr(X^1 \in [\alpha, \beta]) = 1$) for some $\alpha < \beta$. Then, for each*

$\epsilon > 0$, we have, $Pr(\frac{1}{N}\sum_{i=1}^{N} X^i - E[X^1] \geq \epsilon) \leq e^{-2\epsilon^2 N/(\beta-\alpha)^2}$.

Observe again that each sequence of N samples $d^1, \ldots, d^N$ induces an empirical distribution of the demand that can be used to define empirical counterparts to F and $\bar{F}$ defined above. More specifically, consider the samples $d^1, \ldots, d^N$ as coming from N i.i.d random variables $D^1, \ldots, D^N$ all distributed according to the random variable D. For each $y \in \mathbb{R}$ and for each $i = 1, \ldots, N$, let $X^i = X^i(y) = \mathbb{1}(D^i \leq y)$. Given a sequence of N samples, $d^1, \ldots, d^N$, the empirical distribution function is defined as $\hat{F}(y) := \hat{Pr}(D \leq y) := \frac{1}{N}\sum_{i=1}^{N} x^i$ (where x^i is the realization of X^i). Since $x^i \in \{0, 1\}$, $\hat{F}(y)$ is well-defined. Similarly, for each $i = 1, \ldots, N$, let $Z^i = Z^i(y) = \mathbb{1}(D^i \geq y)$, and for each sequence $d^1, \ldots, d^N$, define $\hat{\bar{F}}(y) := \hat{Pr}(D \geq y) = \frac{1}{N}\sum_{i=1}^{N} z^i$ (where again, z^i is the realization of Z^i). Note that $\hat{F}$ and $\hat{\bar{F}}$ are random because they depend on the specific N samples $d^1, \ldots, d^N$.

Observe that $X^1, \ldots, X^N$ (as well as respectively $Z^1, \ldots, Z^N$) are i.i.d random variables. Moreover, $X^i \in [0, 1]$ (respectively, $Z^i \in [0, 1]$), and $E[X^i] = Pr(D \leq y) = F(y)$ ($E[Z^i] = Pr(D \geq y) = \bar{F}(y)$). The Hoeffding inequality then implies that, for each fixed y, and for a sufficiently large N, the random empirical distribution function $\hat{F}(y)$ will be, with high probability, close in its value to $F(y)$, i.e., it implies a bound on the probability that $\hat{F}(y) - F(y) \geq \epsilon$. Similarly, it provides a bound, for each fixed y, on the probability that $\hat{\bar{F}}(y) - \bar{F}(y) \geq \epsilon$.

Using the above idea, we get the following lemma.

Lemma 3.2.5 *For each $\epsilon > 0$ and $0 < \delta < 1$, if the number of samples is $N \geq N(\epsilon, \delta) = \frac{1}{2}\frac{1}{\epsilon^2}\log(\frac{2}{\delta})$, then $\hat{Y}$, the $\frac{b}{b+h}$ quantile of the sample is $\epsilon-$accurate with probability at least $1 - \delta$.*

Proof : First, we bound the probability of the event $B := [F(\hat{Y}) < \frac{b}{b+h} - \epsilon]$ and show that it is at most $\frac{\delta}{2}$. Note that since $\hat{Y}$ is a random variable, we can not apply the Hoeffding inequality directly to $\hat{Y}$. Let q be the $(\frac{b}{b+h} - \epsilon)$ quantile of D, i.e., $q = \inf\{y : F(y) \geq \frac{b}{b+h} - \epsilon\}$. Note that since F is right-continuous, $F(q) \geq \frac{b}{b+h} - \epsilon$. Clearly, B is the event $[\hat{Y} \in (-\infty, q)]$. Now let $\tau_k \downarrow 0$ be a nonnegative, monotone decreasing sequence. Define, for each $k = 1, 2, \ldots$ the event $B_k = [\hat{F}(q - \tau_k) \geq \frac{b}{b+h}]$. Since $\hat{F}$ is monotone increasing, we conclude that $B_k \subseteq B_{k+1}$ and that $B_k \uparrow \bar{B}$ (where $\bar{B}$ is the limit set), which implies that $Pr(B_k) \uparrow Pr(\bar{B})$. Since $\hat{F}(\hat{Y}) \geq \frac{b}{b+h}$ with probability 1, and $\hat{F}$ is monotone increasing, the event B can be written as $[\hat{Y} < q] \cap [\hat{F}(y) \geq \frac{b}{b+h} : \forall \hat{Y} \leq y < q]$. However, each event $[\hat{F}(y) \geq \frac{b}{b+h} : \forall \hat{y} \leq y < q]$ is contained in all the events B_k for k sufficiently large. This implies that $B \subseteq \bar{B}$. However, for each k we have $F(q - \tau_k) < \frac{b}{b+h} - \epsilon$. Therefore, by the Hoeffding inequality, the probability of the event B_k is at most $\exp^{-2\epsilon^2 N}$. By the choice of N above, we conclude that $Pr(B_k) \leq \frac{\delta}{2}$, which also implies that $Pr(\bar{B}) \leq \frac{\delta}{2}$. This concludes the first part of the proof.

Next we bound the probability on the event $L := [\bar{F}(\hat{Y}) < \frac{h}{b+h} - \epsilon]$. Similar to the previous case, now let $q := \sup\{y : \bar{F}(y) \geq \frac{h}{b+h} - \epsilon\}$. Since $\bar{F}$ is left-continuous, we know that $\bar{F}(q) \geq \frac{h}{b+h} - \epsilon$. Since $\hat{\bar{F}}(y)$ is monotone decreasing in y and $\hat{\bar{F}}(\hat{Y}) \geq \frac{h}{b+h}$ with probability 1, we can write L as $[\hat{Y} \in (q, \infty)]$ or as $[\hat{Y} > q] \cap [\hat{\bar{F}}(y) \geq \frac{h}{b+h} : \forall q < y \leq \hat{Y}]$. For $k = 1, 2, \ldots$, define a sequence of events $L_k := [\hat{\bar{F}}(q + \tau_k) \geq \frac{h}{b+h}]$. It is again readily verified, that since $\hat{\bar{F}}$ is monotone decreasing we have $L_k \uparrow \bar{L}$ (where $\bar{L}$ is the limit set), which implies $Pr(L_k) \uparrow Pr(\bar{L})$. Since, for each k, we have that $\bar{F}(q + \tau_k) < \frac{h}{b+h} - \epsilon$, we use again Hoeffding inequality and the chosen N to conclude that $Pr(L_k) \leq \frac{\delta}{2}$, which implies that $Pr(\bar{L}) \leq \frac{\delta}{2}$. Finally, by an argument similar to the one above, we observe that $L \subseteq \bar{L}$, which implies that $Pr(L) \leq \frac{\delta}{2}$.

Now consider the event $[\hat{Y}$ is not $\epsilon-$accurate$]$. It is readily verified that this is a subset of the union of B and L above, and hence has probability at most δ. This concludes the proof of the lemma. ■

Lemma 3.2.5 provides bounds on the number of samples that are required for the $\frac{b}{b+h}$ sample quantile $\hat{Y}$ to be $\epsilon-$accurate with high probability. The next natural question is whether an $\epsilon-$accurate $\hat{y}$ is sufficient to guarantee a bounded relative error in the cost function. In other words, does the fact that $\hat{y}$ is $\epsilon-$accurate imply that it has a good objective value relative to the minimum value of the original cost function C (recall, that the function C is computed under full knowledge of the distribution of D). In the next lemma we show that if $\hat{y}$ is $\epsilon-$accurate, then the difference between its expected cost $C(\hat{y})$ and the optimal expected cost $C(y^*)$ is at most ϵ times $(b+h)|\hat{y}-y^*|$. Moreover, in this case we also provide a lower bound on the optimal cost.

Lemma 3.2.6 *Let $\epsilon > 0$ and assume that $\hat{y}$ is $\epsilon-$accurate. Then:*

(i) $C(\hat{y}) - C(y^*) \leq \epsilon(b+h)|\hat{y}-y^*|$.

(ii) $C(y^*) \geq (\frac{hb}{b+h} - \epsilon \max(b,h))|\hat{y}-y^*|$.

Proof : Suppose $\hat{y}$ is $\epsilon-$accurate. Clearly, either $\hat{y} \geq y^*$ or $\hat{y} < y^*$. Suppose first that $\hat{y} \geq y^*$. We will obtain an upper bound on the difference $C(\hat{y}) - C(y^*)$. Clearly, if the realized demand d is within $(-\infty, \hat{y})$, then the difference between the costs incurred by $\hat{y}$ and y^*, respectively, is at most $h(\hat{y}-y^*)$. On the other hand, if d falls within $[\hat{y}, \infty)$, then y^* has higher cost than $\hat{y}$, by exactly $b(\hat{y}-y^*)$. Now since $\hat{y}$ is assumed to be $\epsilon-$accurate, we know that

$$Pr([D \in [\hat{y}, \infty)]) = Pr(D \geq \hat{y}) = \bar{F}(\hat{y}) \geq \frac{h}{b+h} - \epsilon.$$

We also know that

$$Pr([D \in [0, \hat{y})]) = Pr(D < \hat{y}) = 1 - \bar{F}(\hat{y}) \leq 1 - (\frac{h}{b+h} - \epsilon) = \frac{b}{b+h} + \epsilon.$$

This implies that

$$C(\hat{y}) - C(y^*) \leq h(\frac{b}{b+h} + \epsilon)(\hat{y} - y^*) - b(\frac{h}{b+h} - \epsilon)(\hat{y} - y^*) = \epsilon(b+h)(\hat{y} - y^*).$$

Similarly, if $\hat{y} < y^*$, then for each realization $d \in (\hat{y}, \infty)$ the difference between the costs of $\hat{y}$ and y^*, respectively, is at most $b(y^* - \hat{y})$, and if $d \in (-\infty, \hat{y}]$, then the cost of y^* exceeds the cost of $\hat{y}$ by exactly $h(y^* - \hat{y})$. Since $\hat{y}$ is assumed to be $\epsilon-$accurate, we know that

$$Pr(D \leq \hat{y}) = F(\hat{y}) \geq \frac{b}{b+h} - \epsilon,$$

which also implies that

$$Pr(D > \hat{y}) = 1 - F(\hat{y}) \leq \frac{h}{b+h} + \epsilon.$$

We conclude that

$$C(\hat{y}) - C(y^*) \leq b(\frac{h}{b+h} + \epsilon)(y^* - \hat{y}) - h(\frac{b}{b+h} - \epsilon)(y^* - \hat{y}) = \epsilon(b+h)(y^* - \hat{y}).$$

The proof of part (i) then follows.

The above arguments also imply that if $\hat{y} \geq y^*$ then $C(y^*) \geq E[\mathbb{1}(D \geq \hat{y})b(\hat{y} - y^*)] = \bar{F}(\hat{y})(\hat{y} - y^*)$. We conclude that $C(y^*)$ is at least $b(\frac{h}{b+h} - \epsilon)(\hat{y} - y^*)$. Similarly, in the case $\hat{y} < y^*$, we conclude that $C(y^*)$ is at least $E[\mathbb{1}(D \leq \hat{y})h(y^* - \hat{y})] \geq h(\frac{b}{b+h} - \epsilon)(y^* - \hat{y})$. In other words, $C(y^*) \geq (\frac{hb}{b+h} - \epsilon \max(b, h))|\hat{y} - y^*|$. This concludes the proof of the lemma. ■

We note that there are examples in which the two inequalities in Lemma 3.2.6 above are simultaneously tight. Using Lemmas 3.2.5 and 3.2.6, we can prove the following theorem. As a convention throughout the chapter, we use ϵ' and δ' with reference to the objective value of the solution (in this case, $C(\hat{y})$). We use ϵ and δ with respect to $F(\hat{y})$ and $\bar{F}(\hat{y})$ or, equivalently, to $C^r(\hat{y})$ and $C^l(\hat{y})$.

Theorem 3.2.7 *Consider a newsvendor problem specified by a per-unit holding cost $h > 0$, a per-unit backlogging penalty $b > 0$ and demand distribution D with $E[D] < \infty$. Let $0 < \epsilon' \leq 1$ be a specified accuracy level and $1-\delta'$ (for $0 < \delta' < 1$) be a specified confidence level. Suppose that $N \geq \frac{9}{2\epsilon'^2}(\frac{\min(b,h)}{h+b})^{-2}\log(\frac{2}{\delta'})$ and the SAA counterpart is solved with respect to N i.i.d samples of D. Let $\hat{Y}$ be the optimal solution to the SAA counterpart and $\hat{y}$ denote its realization. Then, with probability at least $1-\delta'$, the expected cost of $\hat{Y}$ is at most $1+\epsilon'$ times the expected cost of an optimal solution to the specified newsvendor problem denoted by y^*. In other words, $C(\hat{Y}) \leq (1+\epsilon')C(y^*)$ with probability at least $1-\delta'$.*

Proof : Let $\epsilon = \frac{\epsilon'}{3}\frac{\min(b,h)}{b+h}$ and $\delta = \delta'$. Suppose that the SAA counterpart is solved for $N \geq \frac{1}{2}\frac{1}{\epsilon^2}\log(\frac{2}{\delta})$ i.i.d. samples of the demand, and that $\hat{y}$ is again the realized optimal solution to the sample-based problem. By lemma 3.2.5, we know that for the specified N there is a probability of at least $1-\delta$ that $\hat{y}$ is $\epsilon-$accurate. Now by Lemma 3.2.6 we know that in this case $C(\hat{y}) - C(y^*) \leq \epsilon(b+h)|\hat{y}-y^*|$ and that $C(y^*) \geq (\frac{hb}{b+h} - \epsilon\max(b,h))|\hat{y}-y^*|$. It is then sufficient to show that $\epsilon(b+h) \leq \epsilon'(\frac{hb}{b+h} - \epsilon\max(b,h))$. Indeed,

$$\epsilon(b+h) \leq (2+\epsilon')\epsilon\max(b,h) - \epsilon'\epsilon\max(b,h) =$$
$$\frac{(2+\epsilon')\epsilon'}{3}\frac{\max(b,h)\min(b,h)}{b+h} - \epsilon'\epsilon\max(b,h) \leq \epsilon'(\frac{hb}{b+h} - \epsilon\max(b,h)).$$

In the first equality we just substitute $\epsilon = \frac{\epsilon'}{3}\frac{\min(b,h)}{b+h}$. The second inequality follows from the assumption that $\epsilon' \leq 1$. We now conclude that, with probability at least $1-\delta'$, we have $C(\hat{Y}) - C(y^*) \leq \epsilon' C(y^*)$ from which the proof of the theorem follows. ■

We note again that the required number of samples does not depend on the demand distribution D. On the other hand, N depends on the square of the reciprocal of $\frac{\min(b,h)}{b+h}$. This implies that the required number of samples N might be large in cases where $\frac{b}{b+h}$ is very close either to 1 or 0. Since the optimal solution y^* is the $\frac{b}{b+h}$ quantile of D, this is consistent with the well-known fact that in order to approximate an extreme (very high or very low) quantile one needs many samples. The intuitive explanation is that if, for example, $\frac{b}{b+h}$ is close to 1, it will take many samples before we see the event $[D > y^*]$. We also note that the bound above is insensitive to scaling of the parameters h and b. It is important to keep in mind that these are only upper bounds on the number of samples required. It is likely that in many cases a significantly lower number of samples will suffice. Moreover, under additional assumptions on the demand distribution it might be possible to get improved bounds.

Finally, the above result holds for newsvendor models with positive per-unit ordering cost as long as $E[D] \geq 0$. Suppose that the per-unit ordering cost is some $c > 0$ (i.e., if y units are ordered a cost of cy is incurred). Clearly, we can assume, without loss of generality, that $c < b$ since otherwise the optimal solution is to order nothing. Consider now a modified newsvendor problem with holding cost and penalty cost parameters $\bar{h} = h + c > 0$ and $\bar{b} = b - c > 0$, respectively. It is readily verified that the modified cost function $\bar{C}(y) = E[\bar{h}(y - D)^+ + \bar{b}(D - y)^+]$ is such that $C(y) = \bar{C}(y) + cE[D]$ and hence the two problems are equivalent. Moreover, if $E[D] \geq 0$ and if the solution $\hat{y}$ guarantees a $1 + \epsilon'$ accuracy level for the modified problem, then it does so also with respect to the original problem, since the cost of each feasible solution is increased by the same positive constant $cE[D]$. Observe that this still allows negative demand.

3.3 MultiPeriod Model

In this section, we consider the multi-period extension of the newsvendor problem, called the *multi-period newsvendor problem*. The goal now is to satisfy a *sequence* of random demands for a single commodity over a planning horizon of T discrete periods (indexed by $t = 1, \dots, T$) with minimum expected cost. The random demand in period t is denoted by D_t. We assume that $D_1, \dots, D_T$ are independent but not necessarily identically distributed.

Each feasible policy P makes decisions in T stages, one decision at the beginning of each period, specifying the number of units to be ordered in that period. Let $Q_t \geq 0$ denote the size of the order in period t. This order is assumed to arrive instantaneously and only then is the demand in period t observed (d_t will denote the realization of D_t). At the end of this section, we discuss the extension to the case where there a positive *lead time* of several periods until the order arrives. For each period $t = 1, \dots, T$, let X_t be the *net inventory* at the beginning of the period. If the net inventory X_t is positive, it corresponds to physical inventory that is left from pervious periods (i.e., from periods $1, \dots, t-1$), and if the net inventory is negative it corresponds to unsatisfied units of demand from previous periods. The dynamics of the model are captured through the equation $X_t = X_{t-1} + Q_{t-1} - D_{t-1}$ (for each $t = 2, \dots, T$). Costs are incurred in the following way. At the end of period t, consider the net inventory x_{t+1} (the realization of X_{t+1}). If $x_{t+1} > 0$, i.e., there are excess units in inventory, then a per-unit holding cost $h_t > 0$ is incurred for each unit in inventory, leading to a total cost of $h_t x_{t+1}$ (the parameter h_t is the per unit cost for carrying one unit of inventory from period t to $t+1$). If, on the other hand, $x_{t+1} < 0$, i.e., there are units of unsatisfied demand, then a per-unit *backlogging penalty cost* $b_t > 0$ is incurred for each unit of unsatisfied demand, and the total cost is $-b_t x_{t+1}$. In particular, all of the unsatisfied units of demand will stay in the system until they are satisfied. That is, b_t plays a role symmetric to that of h_t and can be viewed as the per-unit cost for carrying one unit of shortage from period t to $t+1$. We again assume that the per-unit ordering cost in each period is equal to 0. At the end of this section, we shall relax this assumption. The goal is to find an ordering policy that minimizes the overall expected holding and backlogging cost.

The decision of how many units to order in period t can be equivalently described as the level $Y_t \geq X_t$ to which the net inventory is raised (where clearly $Q_t = Y_t - X_t \geq 0$). Thus, the multi-period model can be viewed as consisting of a sequence of *constrained* newsvendor problems, one in each period. The newsvendor problem in period t is defined with respect to $D_t,\ h_t$ and b_t, under the constraint that $y_t \geq x_t$ (where again x_t and y_t are the respective realizations of X_t and Y_t). However, these newsvendor problems are linked together. More specifically, the decision in period t may constrain the newsvendor problems in future periods since it may impact the net inventory in these periods. Thus, myopically minimizing the expected newsvendor cost in period t is, in general, not optimal with respect to the total cost over the entire horizon. This makes the multi-period model significantly more complicated. Nevertheless, given full access to the demand distributions $D_1, \ldots, D_t$, this model can be solved to optimality by means of dynamic programming. The multi-period model is again well-studied. We present a summary of the main known results regarding the structure of optimal policies, emphasizing those facts that will be essential for our results. This serves as a background for the proceeding discussion about the sample-based algorithm and its analysis.

3.3.1 Optimal Policies

It is a well-known fact that in the multi-period model described above, the class of *base-stock policies* is optimal. A base-stock policy is characterized by a set of target inventory (base-stock) levels associated with each period t and each possible state of the system in period t. At the beginning of each period t, a base-stock policy aims to keep the inventory level as close as possible to the relevant target level. Thus, if the inventory level at the beginning of the period is below the target level, then the base-stock policy will order up to the target level. If, on the other hand, the inventory level at the beginning of the period is higher than the target, then no order is placed.

Optimal base-stock policy has two important properties. First, the optimal base-stock level in period t is *independent* of all the decisions made (i.e., orders placed) prior to period t. In particular, it is independent of X_t. Second, its optimality is conditioned on the execution of an optimal base-stock policy in the future periods $t+1,\ldots,T$. As a result, optimal base-stock policies can be computed using dynamic programming, where the optimal base-stock levels are computed by a backward recursion from period T to period 1. The main problem is that the state space in each period might be very large, which makes the relevant dynamic program intractable. However, in the model discussed here, the demands in different periods are assumed to be independent, and the corresponding dynamic program is therefore usually easy to solve, again *if* we have full access to the demand distributions. In particular, an optimal base-stock policy in this model consists of T base-stock levels, one for each period.

Next, we present a dynamic programming formulation of the model discussed above and highlight the facts that are most relevant to the proceeding discussion. In the following subsection, we shall show how to use a similar dynamic programming framework to construct a sample-based policy that approximates an optimal base-stock policy.

Let $C_t(y_t)$ be the newsvendor cost associated with period t (for $t=1,\ldots,T$) as a function of the inventory level y_t after ordering, i.e.,

$$C_t(y_t) = E[h_t(y_t - D_t)^+ + b_t(D_t - y_t)^+].$$

For each $t=1,\ldots,T$, let $V_t(x_t)$ be the optimal (minimum) expected cost over the interval $[t,T]$ assuming that the inventory level at the beginning of period t is x_t and that optimal decisions are going to be made over the entire horizon $(t,T]$. Also let $U_t(y_t)$ be the expected cost over the horizon $[t,T]$ given that the inventory level in period t was raised to y_t (after the order in period t was placed) and that an optimal policy is followed over the interval $(t, T]$. Clearly, $U_T(y_T) = C_T(y_T)$ and $V_T(x_T) = \min_{y_T \geq x_T} C_T(y_T)$. Now for each $t = 1, \ldots, T-1$,

$$U_t(y_t) = C_t(y_t) + E[V_{t+1}(y_t - D_t)]. \tag{1}$$

We can now write, for each $t = 1, \ldots, T$,

$$V_t(x_t) = \min_{y_t \geq x_t} U_t(y_t). \tag{2}$$

Observe that the optimal expected cost V_t has two parts, the newsvendor (or the period) cost, C_t, and the expected future cost, $E[V_{t+1}(y_t - D_t)]$ (where the expectation is taken with respect to D_t). The decision in period t effects the future cost since it effects the inventory level at the beginning of the next period.

The above dynamic program provides a correct formulation of the model discussed above (see [56] for a detailed discussion). The goal is to compute $V_1(x_1)$, where x_1 is the inventory level at the beginning of the horizon, which is given as an input. The following fact provides insight with regard to why this formulation is indeed correct and to why base-stock policies are optimal.

Fact 3.3.1 *Let $f : \mathbb{R} \mapsto \mathbb{R}$, be a real-valued convex function with a minimizer r (i.e., $f(r) \leq f(y)$ for each $y \in \mathbb{R}$). Then the following holds:*

(i) The function $w(x) = \min_{y \geq x} f(y)$ is convex in x.

(ii) For each $x \leq r$, we have $w(x) = f(r)$, and for each $x > r$, we have $w(x) = f(x)$.

Using Fact 3.3.1 above, it is straightforward to show that, for each $t = 1, \ldots, T$, the function $U_t(y_t)$ is convex and attains a minimum, and that the function $V_t(x_t)$ is convex.

The proof is done by induction over the periods, as follows. The claim is clearly true for $t = T$ since U_T is just a newsvendor cost function and $V_T(x_T) = \min_{y_T \geq x_T} U_T(y_T)$. Suppose now that the claim is true for $t+1, \ldots, T$ (for some $t < T$). From (1), it is readily verified that U_t is convex since it is a sum of two convex functions. It attains a minimum because $\lim_{y_t \to \infty} U_t(y_t) = \infty$ and $\lim_{y_t \to -\infty} U_t(y_t) = \infty$. The convexity of V_t follows again from Fact 3.3.1 above. This also implies that base-stock policies are indeed optimal. Moreover, if the demand distributions are explicitly specified, it is usually straightforward to recursively compute optimal base-stock levels $R_1, \ldots, R_T$, since they are simply minimizers of the functions $U_1, \ldots, U_T$, respectively. More specifically, if the demand distributions are known explicitly, we can compute R_T, which is a minimizer of a newsvendor cost function, then recursively define U_{T-1} and solve for its minimizer R_{T-1} and so on. In particular, if the minimizers $R_{t+1}, \ldots, R_T$ were already computed, then $U_t(y_t)$ is a convex function of a single variable and hence it is relatively easy to compute its minimizer. Throughout the chapter we assume, without loss of generality, that for each $t = 1, \ldots, T$, the optimal base-stock level in period t is denoted by R_t and that this is the *smallest* minimizer of U_t (in case it has more than one minimizer). The minimizer R_t of U_t can then be viewed as the best policy in period t conditioning on the fact that the optimal base-stock policy $R_{t+1}, \ldots, R_T$ will be executed over $[t+1, T]$.

By applying Fact 3.3.1 above to V_{t+1} and U_{t+1}, it is readily verified that the function U_t can be expressed as, $U_t(y_t) =$

$$C_t(y_t) + E[\mathbb{1}(y_t - D_t \leq R_{t+1})U_{t+1}(R_{t+1}) + \mathbb{1}(y_t - D_t > R_{t+1})(U_{t+1}(y_t - D_t)]. \quad (3)$$

Clearly this is a continuous function of y_t. As in the newsvendor model, one can derive explicit expressions for the right-hand and left-hand derivatives of the functions $U_1, \ldots, U_T$, as follows. Assume first that all the demand distributions are continuous, which implies that the functions $U_1, \ldots, U_T$ are all continuously differentiable. The derivative of $U_T(y_T)$ is clearly $U_T'(y_T) = C_T' = -b_T + (h_T + b_T)F_T(y_T)$, where F_T is the CDF of D_T. Now consider the function $U_t(y_t)$ for some $t < T$. Using again the dominated convergence theorem, one can change the order of expectation and integration to get

$$U_t'(y_t) = C_t'(y_t) + E[V_{t+1}'(y_t - D_t)]. \tag{4}$$

However, by Fact 3.3.1 and (3) above, the derivative $V_{t+1}'(x_{t+1})$ is equal to 0 for each $x_{t+1} \leq R_{t+1}$ and is equal $U_{t+1}'(x_{t+1})$ for each $x_{t+1} > R_{t+1}$ (where again R_{t+1} is a minimizer of U_{t+1}). This implies that

$$E[V_{t+1}'(y_t - D_t)] = E[\mathbb{1}(y_t - D_t > R_{t+1})U_{t+1}'(y_t - D_t)]. \tag{5}$$

Applying this argument recursively, we obtain

$$U_t'(y_t) = C_t'(y_t) + E[\sum_{j=t+1}^{T} \mathbb{1}(A_{jt}(y_t))C_j'(y_t - D_{[t,j)})], \tag{6}$$

where $D_{[t,j)}$ is the accumulated demand over the interval $[t,j)$ (i.e., $D_{[t,j)} = \sum_{k=t}^{j-1} D_k$), and $A_{jt}(y_t)$ is the event that for each $k \in (t,j]$ the inequality $y_t - D_{[t,k)} > R_k$ holds. Observe that $y_t - D_{[t,k)}$ is the inventory level at the beginning of period k, assuming that we order up to y_t in period t and do not order in any of the periods $t+1, \ldots, k-1$. If $y_t - D_{[t,k)} \leq R_k$, then the optimal base-stock level in period k is reachable, and the decision made in period t does not have any impact on the future cost over the interval $[k,T]$. However, if $y_t - D_{[t,s)} > R_s$ for each $s = t+1, \ldots, k$, then the optimal base-stock level in period k is not reachable due to the decision made in period t, and the derivative $C_k'(y_t - D_{[t,k)})$ accounts for that impact on the cost in period k. The derivative of U_t consists of a sum of derivatives of newsvendor cost functions multiplied by the respective indicator functions.

For general (independent) demand distributions, the functions $U_1, \ldots, U_t$ might not be differentiable, but similar arguments can be used to derive explicit expressions for the right-hand and left-hand derivatives of U_t, denoted by U_t^r and U_t^l, respectively. This is done simply by replacing C_j' by C_j^r and C_j^l (see Section 3.2 above), respectively, in the above expression of U_t' (for each $j = t, \ldots, T$). In addition, in the right-hand derivative the events $A_{jt}(y_t)$ are defined with respect to weak inequalities. This also provides an optimality criterion for finding a minimizer R_t of U_t, namely, $U_t^r(R_t) \geq 0$ and $U_t^l(R_t) \leq 0$. If the demand distributions are explicitly given, it is usually easy to evaluate the one-sided derivatives of U_t. This suggests the following approach for solving the dynamic program presented above. In each stage, compute R_t such that $0 \in \partial U_t(R_t)$, by considering the respective one-sided derivatives of U_t. In the next subsection, we shall use a similar algorithmic approach, but with respect to an approximate base-stock policy and under the assumption that the only information about the demand distributions is available through a black box.

3.3.2 Approximate Base-Stock Levels

In order to exactly solve the dynamic program described above, it is essential to know the demand distributions. However, as mentioned before, in most real-life scenarios these distributions are either not available or are too complicated to work with directly. Instead we shall consider this model again under the assumption that the only information available is through a black box that on request can generate independent samples from the demand distributions $D_1, \ldots, D_T$. As in the newsvendor model discussed in Section 3.2, the goal is to find a policy with expected cost close to the expected cost of an optimal policy that is assumed to have full access to the demand distributions. In particular, we shall describe a sample-based algorithm that, for each specified accuracy level ϵ' and confidence level δ', computes a base-stock policy such that with probability at least $1-\delta'$, the expected cost of the policy is at most $1+\epsilon'$ times the expected cost of an optimal policy. Throughout the chapter, we use $R_1, \ldots, R_T$ to denote the *minimal optimal base-stock-level*, i.e., the optimal base-stock policy , in which, for each $t = 1, \ldots, T$, the base-stock level R_t is the smallest minimizer of U_t defined above. Next we provide an overview of the algorithm and its analysis.

An overview of the algorithm and its analysis. First we note again that our approach departs from the SAA method or the IPA methods discussed in Sections 3.1 and 3.2 above. Instead, it is based on a dynamic programming framework. That is, the base-stock levels of the policy are computed using a backward recursion. In particular, the base-stock level in period t, denoted by $\tilde{R}_t$, is computed based on the previously computed base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$. If $T = 1$, then this is reduced to solving the SAA of the single-period newsvendor model already discussed in Section 3.2. However, if $T > 1$ and the base-stock levels are approximated recursively, then the issue of convexity needs to be carefully addressed. It is no longer clear, whether each subproblem is still convex, and whether base-stock policies are still optimal. More specifically, assume that some (approximate) base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t+1, T]$, not necessarily an optimal one, was already computed in previous stages of the algorithm. Now let $\tilde{U}_t(y_t)$ be the expected cost over $[t, T]$ of a policy that orders up to y_t in period t and then follows the base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over $[t+1, T]$ (as before, expectations are taken with respect to the underlying demand distributions $D_1, \ldots, D_T$). Let $\tilde{V}_t(x_t)$ be the minimum expected cost over $[t, T]$ over all ordering policies in period t, given that the inventory level at the beginning of the period is x_t and that the policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is followed over $[t+1, T]$. Clearly, $\tilde{V}_t(x_t) = \min_{y_t \geq x_t} \tilde{U}_t(y_t)$. The functions $\tilde{U}_t$ and $\tilde{V}_t$ play analogous roles to those of U_t and V_t, respectively, but are defined with respect to $\tilde{R}_{t+1},\ldots,\tilde{R}_T$ instead of $R_{t+1},\ldots,R_T$. They define a *shadow dynamic program* to the one described above, that is based on the functions U_t and V_t. From now on, we will distinguish functions and objects that are defined with respect to the approximate policy $\tilde{R}_1,\ldots,\tilde{R}_T$ by adding the *tilde* sign above them. The convexity of U_t and V_t and the optimality of base-stock policies are heavily based on the optimality of $R_{t+1},\ldots,R_T$ (using Fact 3.3.1 above). Since the approximate policy $\tilde{R}_{t+1},\ldots,\tilde{R}_T$ is not necessarily optimal, the functions $\tilde{U}_t$ and $\tilde{V}_t$ might not be convex, and hence a base-stock policy in period t might not be optimal any more. In order to keep the subproblem (i.e., the function $\tilde{U}_t$) in each stage tractable, the algorithm is going to maintain (with high probability) an invariant under which the convexity of $\tilde{U}_t$ and $\tilde{V}_t$ and the optimality of base-stock policies are preserved (see Definition 3.3.2 and the key Lemma 3.3.3, where we establish the resulting convexity of the functions $\tilde{U}_t$ and $\tilde{V}_t$). Assuming that $\tilde{U}_t$ and $\tilde{V}_t$ are indeed convex, it would be natural to compute the smallest minimizer of $\tilde{U}_t$, denoted by $\bar{R}_t$. However, this also requires full access to the explicit demand distributions. Instead, the algorithm takes the following approach. In each stage $t = T,\ldots,1$, the algorithm uses a sample-based procedure to compute a base-stock level $\tilde{R}_t$ that, with high probability, has the following two properties. First, the base-stock level $\tilde{R}_t$ is a good approximation of the minimizer $\bar{R}_t$, in that $\tilde{U}_t(\tilde{R}_t)$ is close to the minimum value $\tilde{U}_t(\bar{R}_t)$, i.e., it has a small relative error. Second, $\tilde{R}_t$ is, with high probability, greater or equal than $\bar{R}_t$. It is this latter property that preserves the invariant of the algorithm, and in particular, preserves the convexity of $\tilde{U}_{t-1}$ and $\tilde{V}_{t-1}$ in the next stage.

The justification for this approach is given in the second key Lemma 3.3.4, where it is shown that the properties of $\tilde{R}_t,\ldots,\tilde{R}_T$ also guarantee that small errors relative to $\tilde{U}_t(\bar{R}_t),\ldots,\tilde{U}_T(\bar{R}_T)$, respectively, accumulate but have impact only on the expected cost over $[t, T]$ and do not propagate to the interval $[1, t)$. Thus, applying this approach recursively leads to a base-stock policy for the entire horizon with expected cost close to the optimal expected cost. Analogous to the newsvendor cost function, the functions $\tilde{U}_1, \ldots, \tilde{U}_T$ also have similar explicit expressions for the one-sided derivatives that are also bounded, and hence can be estimated accurately with samples. However, in order to compute such $\tilde{R}_t$ in each stage, it is essential to establish an explicit connection between first order information, i.e., information about the value of the one-sided derivatives of $\tilde{U}_t$ at a certain point, and the bounded error this guarantees relative to $\tilde{U}_t(\tilde{R}_t)$. This is done in Lemma 3.3.6 below which plays a similarly key role to Lemma 3.2.6 in the previous section. Finally, in Lemma 3.3.7, Corollaries 3.3.8 and 3.3.9, and Lemma 3.3.10, it is shown how the one-sided derivatives of $\tilde{U}_t$ can be estimated using samples in order to compute $\tilde{R}_t$ that, with high probability, maintains the two required properties.

Next we discuss one of the key ideas underlying the algorithm, that is, the invariant that preserves the convexity of the functions $\tilde{U}_t$ and $\tilde{V}_t$ above and the optimality of a base-stock policy in period t. In the case where there exists an optimal ordering policy in period t which is a base-stock policy (i.e., $\tilde{U}_t$ is convex), let again $\bar{R}_t = \bar{R}_t | \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ be the smallest minimizer of $\tilde{U}_t$, i.e., the smallest optimal base-stock level in period t, given that the policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is followed in periods $t+1, \ldots, T$. If the optimal ordering policy in period t given $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ is not a base-stock policy, we say that $\bar{R}_t$ does not exist.

Definition 3.3.2 *A base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ for the interval $[t+1, T]$ is called an* upper base-stock policy *if, for each $j = t+1, \ldots, T$, such that $\bar{R}_j$ exists, the inequality $\tilde{R}_j \geq \bar{R}_j$ holds.*

The algorithm is going to preserve this invariant by computing in each stage $t =$

$T, \ldots, 1$ an $\tilde{R}_t$ such that with high probability, $\tilde{R}_t \geq \bar{R}_t$. In the next two lemmas we shall show several important structural properties of upper base-stock policies. In the first of these lemmas, we shall show that, for each $j = 1, \ldots, T$, assuming that the base-stock policy $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$ is followed over $[j+1, T]$, there exists an optimal ordering policy in period j which is a base-stock policy, and that the convexity of the functions $\tilde{U}_j$ and $\tilde{V}_j$ is preserved.

Lemma 3.3.3 *Let $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ be an upper base-stock policy over $[t+1, T]$. For each $j = t, \ldots, T$, let $\tilde{U}_j(y_j)$ be the expected cost over $[j, T]$ of a policy that orders up to y_j in period j, and then follows $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$, and let $\tilde{V}_j(x_j)$ be the minimum expected cost over $[j, T]$ given that at the beginning of period j the inventory level is x_j and that over $[j+1, T]$ the base-stock policy $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$ is followed, i.e., $\tilde{V}_j(x_j) = \min_{y_j \geq x_j} \tilde{U}_j(y_j)$. Then, for each period $j = t, \ldots, T$,*

(i) The functions $\tilde{U}_j$ and $\tilde{V}_j$ are convex and $\tilde{U}_j$ attains a minimum.

(ii) Given that over $[j+1, T]$ we follow the base-stock policy $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$, there exists an optimal ordering policy in period j which is a base-stock policy, i.e., $\bar{R}_j$ does exist.

Proof : For each $k = t+1, \ldots, T$, let $\tilde{B}_k(x_k)$ be the expected cost of the base-stock policy $\tilde{R}_k, \ldots, \tilde{R}_T$ over the interval $[k, T]$ given that there are x_k units in inventory at the beginning of period k. Thus, for each $t < T$, $\tilde{U}_t(y_t) = C_t(y_t) + E[\tilde{B}_{t+1}(y_t - D_t)]$.

The proof follows by induction on $j = T, \ldots, t$. For $j = T$, observe that $\tilde{U}_T = U_T = C_T$ and $\tilde{V}_T(x_T) = V_T(x_T) = \min_{y_T \geq x_T} \tilde{U}_T(y_T)$, which implies that both $\tilde{U}_T$ and $\tilde{V}_T$ are convex, $\tilde{U}_T$ attains a minimum and $\bar{R}_T = R_T$ is indeed an optimal base-stock policy in period T. In particular, $\bar{R}_T$ is the smallest minimizer of $\tilde{U}_T$. Now assume that the claim is true for $j > t$, i.e., for each of the periods $j, \ldots, T$. In particular, the functions $\tilde{U}_j, \ldots, \tilde{U}_T$ are convex, $\bar{R}_j, \ldots, \bar{R}_T$ are their respective smallest minimizers, and the functions $\tilde{V}_j, \ldots, \tilde{V}_T$ are convex. Consider now the function $\tilde{U}_{j-1}(y_{j-1}) = C_{j-1}(y_{j-1})+ E[\tilde{B}_j(y_{j-1} - D_{j-1})]$. Since C_{j-1} is convex, it is sufficient to show that $\tilde{B}_j$ is convex. By induction, $\bar{R}_j$ is a minimizer of $\tilde{U}_j$ and $\tilde{R}_j \geq \bar{R}_j$. Hence, the function $B_j(x_j)$ can be expressed as $B_j(x_j) = \max\{\tilde{V}_j(x_j), \tilde{U}_j(\tilde{R}_j)\}$, which implies that it is indeed convex, since it is the maximum of two convex functions. It is straightforward to see that $\tilde{U}_{j-1}$ is convex and has a minimizer, where again its smallest minimizer is denoted by $\bar{R}_{j-1}$. By Fact 3.3.1, we conclude that $\tilde{V}_{j-1}(x_{j-1})$ is also convex and that there exists an optimal base-stock policy in period $j-1$ (again assuming that the policy $\tilde{R}_j, \ldots \tilde{R}_T$ is followed over $[j, T]$). ■

The next key lemma considers the case where an upper base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t+1, T]$ is known to provide a good solution for that interval. More specifically, for each $j = t+1, \ldots, T$, the expected cost of the base-stock policy $\tilde{R}_j, \ldots, \tilde{R}_T$ is assumed to be close to optimal over the interval $[j+1, T]$, i.e., $\tilde{U}_j(\tilde{R}_j) \leq \alpha_j U_j(R_j)$ for some $\alpha_j \geq 1$. We shall show that this gives rise to a good policy over the entire horizon $[1, T]$.

Lemma 3.3.4 *For some $t < T$, let $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ be an upper base-stock policy for the interval $[t+1, T]$ and consider again the function $\tilde{U}_t(y_t)$ and its smallest minimizer $\bar{R}_t$. Furthermore, assume that for each $j = t+1, \ldots, T$, the cost of the base-stock policy $\tilde{R}_j, \ldots, \tilde{R}_T$ is at most α_j times the optimal expected cost over that interval (where $\alpha_j \geq 1$), i.e., $\tilde{U}_j(\tilde{R}_j) \leq \alpha_j U_j(R_j)$. Let $\alpha = \max_j \alpha_j$ and consider the minimal optimal base-stock policy over the interval $[t, T]$, denoted again by $R_t, \ldots, R_T$. Then the expected cost of the base-stock policy $\bar{R}_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t, T]$ is at most α times the expected cost of an optimal base-stock policy over that interval.*

Proof : Recall that for each $j = t+1, \ldots, T$ the function $U_j(y_j)$ is defined with respect to the optimal base-stock levels $R_{j+1}, \ldots, R_T$, and the function $\tilde{U}_j(y_j)$ is defined with respect to the base-stock levels $\tilde{R}_{j+1}, \ldots, \tilde{R}_T$.

Suppose first that under the assumptions in the lemma, the following *structural claim* is true. For each $j > t$, consider the interval $[j, T]$. Then the expected cost of the policy $R_j, \tilde{R}_{j+1}, \ldots, \tilde{R}_T$ over that interval is at most α times the respective (optimal) expected cost of the policy $R_j, R_{j+1}, \ldots, R_T$.

Now consider the *modified* base-stock policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$. This policy consists of the optimal base-stock level R_t in period t, followed by $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the rest of the interval. Since in period t the policy is identical to the optimal base-stock policy $R_t, \ldots, R_T$, it is clear that the two policies incur the same cost in period t (for each possible realization of D_t). Moreover, at the beginning of period $t+1$ both policies will have the same inventory level X_{t+1}. Let $\tilde{t} \geq t+1$ be the first period after t in which either the modified or the optimal policy placed an order. Observe that $\tilde{t}$ is a random variable. Over the (possibly empty) interval $[t+1, \tilde{t})$ neither the modified policy nor the optimal policy ordered. This implies that over that interval they had exactly the same inventory level, and therefore, they have incurred exactly the same cost. Now in the period $\tilde{t}$ exactly one of two cases applies. If the modified policy has placed an order, then by our assumption, it is clear that the expected cost that the modified policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ incurs over the interval $[\tilde{t}, T]$ is at most α times the expected cost of the optimal base-stock policy $R_t, \ldots, R_T$ over that interval. Now consider the case in which the optimal policy has placed an order in $\tilde{t}$ and the modified policy did not. At the beginning of period $\tilde{t}$, the inventory level of both policies is the same and is equal to $X_{t+1} - D_{[t+1,\tilde{t})}$. By our assumption we have

$$R_{\tilde{t}} \geq X_{t+1} - D_{[t+1,\tilde{t})} > \tilde{R}_{\tilde{t}} \geq \bar{R}_{\tilde{t}},$$

where again $\bar{R}_{\tilde{t}}$ is the smallest minimizer of $\tilde{U}_{\tilde{t}}$, i.e., the best policy in $\tilde{t}$ given that the policy $\tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$ is followed over $[\tilde{t}+1, T]$. Note that $\tilde{R}_{\tilde{t}} \geq \bar{R}_{\tilde{t}}$ by the definition of an upper base-stock policy. Since $\tilde{R}_T \geq \bar{R}_T = R_T$ we conclude that if the above case applies, then, with probability 1, $\tilde{t} < T$. In particular, over the interval $[\tilde{t}, T]$, the optimal policy executed the policy $R_{\tilde{t}}, R_{\tilde{t}+1}, \ldots, R_T$. By the structural claim above, we also know that the expected cost of $R_{\tilde{t}}, \tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$ over $[\tilde{t}, T]$ is at most α times the expected cost of the optimal policy $R_{\tilde{t}}, R_{\tilde{t}+1}, \ldots, R_T$ over that interval. However, observe that $X_{t+1} - D_{[t+1,\tilde{t})}$ is closer to $\bar{R}_{\tilde{t}}$ than $R_{\tilde{t}}$ (see inequality above). This implies that the modified policy has expected cost over the interval $[\tilde{t}, T]$ no greater than the expected cost of the policy $R_{\tilde{t}}, \tilde{R}_{\tilde{t}+1}, \ldots, \tilde{R}_T$, i.e., at most α times the optimal expected cost of that interval.

It follows that indeed the policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ has total expected cost over $[t, T]$ at most α times the optimal expected cost over that interval. Now consider the policy $\bar{R}_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$, where the base-stock level R_t is replaced by $\bar{R}_t$. By Lemma 3.3.3 we know that given an upper base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ over the interval $[t+1, T]$ the optimal ordering policy in period t is a base-stock policy, and $\bar{R}_t$ is the smallest optimal base-stock level. Therefore, this policy has even a lower expected cost than the modified policy $R_t, \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ discussed above.

It is left to show that the structural claim is indeed valid. It is readily verified that the claim is true for $j = T$ and $j = T - 1$. The proof of the induction step is done by arguments identical to the arguments used above. The proof then follows. ■

Consider now an upper base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ such that for each $t = 1, \ldots, T$, the base-stock level $\tilde{R}_t$ is a good approximation of $\bar{R}_t = \bar{R}_t | \tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (by Lemma 3.3.3 above we know that $\bar{R}_t$ is well-defined). More specifically, for each $t = 1, \ldots, T$, we have $\tilde{R}_t \geq \bar{R}_t$ and $\tilde{U}_t(\tilde{R}_t) \leq (1 + \epsilon'_t)\tilde{U}_t(\bar{R}_t)$ (for some specified $0 \leq \epsilon'_t$), where

$\tilde{U}_t(y_t)$ is again defined with respect to $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (see above) and $\bar{R}_t$ is its smallest minimizer. Using Lemmas 3.3.3 and 3.3.4, it is straightforward to verify (by backward induction on the periods) that, for each $s = 1, \ldots, T$, the expected cost of the base-stock policy $\tilde{R}_s, \ldots, \tilde{R}_T$ has expected cost at most $\prod_{j=s}^{T}(1 + \epsilon'_j)$ times the optimal expected cost over the interval $[s, T]$.

For $s = T$, the claim is trivially true. Now assume it is true for some $s > 1$. Applying Lemma 3.3.4 above with $t = s - 1$ and $\alpha = \prod_{j=s}^{T}(1 + \epsilon'_j)$ we conclude that the policy $\bar{R}_{s-1}, \tilde{R}_s, \ldots, \tilde{R}_T$ has expected cost at most $\alpha = \prod_{j=s}^{T}(1 + \epsilon'_j)$ times the optimal expected cost over the interval $[s - 1, T]$. Now by the definition of $\tilde{R}_{s-1}$, we conclude that the policy $\tilde{R}_{s-1}, \tilde{R}_s, \ldots, \tilde{R}_T$ has expected cost at most $(1 + \epsilon'_{s-1})\alpha$ times the optimal expected cost over $[s - 1, T]$ from which the claim follows. In particular, this implies that the expected cost of the base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ over the entire horizon $[1, T]$ is at most $\prod_{t=1}^{T}(1+\epsilon'_t)$ times the optimal expected cost. In other words, the properties of an upper base-stock policy guarantee that errors over the interval $[t + 1, T]$ do not propagate to the interval $[1, t]$.

To compute such a base-stock policy it is sufficient to compute recursively, for each $t = T, \ldots, 1$, a base-stock level $\tilde{R}_t$ with the following two properties. To preserve the invariant of an upper base-stock level policy, it is required that $\tilde{R}_t \geq \bar{R}_t$. In addition, $\tilde{R}_t$ is required to have a small relative error with respect to $\tilde{U}_t(y_t)$ and its minimizer $\bar{R}_t$, i.e., $\tilde{U}_t(\tilde{R}_t) \leq (1 + \epsilon'_t)\tilde{U}_t(\bar{R}_t)$. Recall that, if the invariant of an upper base-stock policy is preserved, then the function $\tilde{U}_t$ is convex with (one-sided) derivatives as given in (7) above. This suggests the same approach as before, i.e., use first-order information in order to find a point with objective value close to optimal. However, unlike the newsvendor cost function, the minimizer of $\tilde{U}_t$ is not a well-defined quantile of the distribution of D_t and it is less obvious how to establish a connection between the value of the one-sided derivatives of $\tilde{U}_t$ at some point y and the relative error of that point with respect to the minimum expected cost $\tilde{U}_t(\bar{R}_t)$. This is established in the next key lemma which has an analogous role to that of Lemma 3.2.6 above. Observe, that for each $t = 1, \ldots, T$, the function $\tilde{U}_t$ is bounded from below by the newsvendor cost C_t, i.e., for each y, we have $\tilde{U}_t(y) \geq C_t(y)$. Now $C_t(y) = E[h_t(y - D_t)^+ + b_t(D_t - y)^+]$ and for each fixed y the function $h_t(y - D_t)^+ + b_t(D_t - y)^+$ is convex in D_t. Applying Jensen's inequality we conclude that the inequality $C_t(y) \geq h_t(y - E[D_t])^+ + b_t(E[D_t] - y)^+$ holds for each y. For each $t = 1, \ldots, T$, the function $h_t(y - E[D_t])^+ + b_t(E[D_t] - y)^+$ is piecewise linear and convex with a minimum attained at $y = E[D_t]$ and equal to zero. Moreover, it provides a lower bound on $\tilde{U}_t(y)$. This structural property of the functions $\tilde{U}_1, \ldots, \tilde{U}_T$ can be used to establish an explicit connection between first order information and relative errors. The next lemma is specialized to the specific setting of the functions $\tilde{U}_1, \ldots, \tilde{U}_T$. In [30], we present a general version of this lemma that is valid in the multi-dimensional case. We believe that this structural lemma will have additional applications in different settings.

Before we state and prove the lemma, we introduce the following definition.

Definition 3.3.5 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be convex and finite. A point $y \in \mathbb{R}^m$ is called an $\epsilon-$point of f if there exists a subgradient r of the function f at y with Euclidean norm less than ϵ, i.e., there exists $r \in \partial f(y)$ with $||r||_2 \leq \epsilon$.*

Lemma 3.3.6 *Let $f : \mathbb{R} \mapsto \mathbb{R}$ be convex and finite with a minimum at y^*, i.e., $f(y^*) \leq f(y)$ for each $y \in \mathbb{R}$. Suppose that $\bar{f}(y) = h(y - d)^+ + b(d - y)^+$ (where $h,\ b > 0$) is a convex piecewise linear function with minimum equal to 0 at d, such that $f(y) \geq \bar{f}(y)$ for each $y \in \mathbb{R}$. Let $0 < \epsilon' \leq 1$ be the specified accuracy level, and let $\epsilon = \frac{\epsilon'}{3}\min(b, h)$. If $\hat{y}$ is an $\epsilon-$point of f, then $f(\hat{y}) \leq (1 + \epsilon')f(y^*)$.*

Proof : Let $\lambda = \min(b, h)$. Since $\hat{y}$ is an $\epsilon-$point of f, we conclude that there exists some $r \in \partial f(\hat{y})$ with $|r| \leq \epsilon$, and by Definition 3.2.1, $f(\hat{y}) - f(y^*) \leq \epsilon|\hat{y} - y^*|$. Now let $d_1 \leq d \leq d_2$ be the two points where $\bar{f}$ takes the minimum value of f, i.e., $\bar{f}(d_1) = \bar{f}(d_2) = f(y^*)$. Let $L_1 = d - d_1$ and $L_2 = d_2 - d$. Clearly, $f(y^*) = bL_1 = hL_2$, which implies that $f(y^*) \geq \frac{\lambda}{2}(L_1 + L_2)$. Moreover, since $f(y) \geq \bar{f}(y)$ for each y, we conclude that $y^* \in [d_1, d_2]$.

Now consider the point $\hat{y}$. Suppose first that $\hat{y} \in [d_1, d_2]$. This implies that $|\hat{y} - y^*| \leq L_1 + L_2$. We now get that

$$f(\hat{y}) - f(y^*) \leq \epsilon|\hat{y} - y^*| \leq \epsilon(L_1 + L_2) = \frac{\epsilon'}{3}\lambda(L_1 + L_2) \leq \frac{2}{3}\epsilon' f(y^*).$$

The equality is just a substitution of $\epsilon = \frac{\epsilon'}{3}\lambda$. The claim then follows.

Now assume that $\hat{y} \notin [d_1, d_2]$. Without loss of generality, assume $\hat{y} > d_2$ (a symmetric proof applies if $\hat{y} < d_1$). Let $x = \hat{y} - d_2$. Since f is convex, it is clear that d_2 is also an $\epsilon-$point of f. By the same arguments as used above we conclude that $f(d_2) - f(y^*) \leq \frac{2}{3}\epsilon' f(y^*)$ and that $f(\hat{y}) - f(d_2) \leq \epsilon|\hat{y} - d_2| = \epsilon x$. This implies that,

$$f(\hat{y}) - f(y^*) = f(\hat{y}) - f(d_2) + f(d_2) - f(y^*) \leq \epsilon x + \frac{2}{3}\epsilon' f(y^*).$$

It is then sufficient to show that $\epsilon x \leq \frac{1}{3}\epsilon' f(y^*)$. We first bound x from above. Now $\bar{f}(\hat{y}) = f(y^*) + hx \geq f(y^*) + \lambda x$. In addition, $f(\hat{y}) \leq f(d_2) + \epsilon x \leq (1 + \frac{2}{3}\epsilon')f(y^*) + \epsilon x$. However, $\bar{f}(\hat{y}) \leq f(\hat{y})$. We conclude that,

$$x \leq \frac{2}{3}\frac{\epsilon' f(y^*)}{\lambda - \epsilon} = \frac{2}{\lambda}\frac{\epsilon' f(y^*)}{(3 - \epsilon')} \leq \frac{\epsilon' f(y^*)}{\lambda}.$$

The equality is again just a substitution of $\epsilon = \frac{\epsilon'}{3}\lambda$. The last inequality is because $\epsilon' \leq 1$. However, this implies that $\epsilon x \leq \frac{\epsilon' f(y^*)}{3}$, from which the claim follows. ∎

Lemma 3.3.6 above establishes an explicit connection between $\epsilon-$points of the functions $\tilde{U}_t$ (for $t = 1, \ldots, T$) and the relative error they guarantee. We note that slightly tighter bounds can be proven using somewhat more involved algebra.

Since the demand distributions are not available, it is left to show how the one-sided derivatives of these functions can be estimated with high accuracy and high confidence probability using random samples.

We next derive expressions of the one-sided derivatives of $\tilde{U}_{j-1}$, similar to (6) above. Note that $\tilde{U}_{j-1}(y_{j-1}) =$

$$C_{j-1}(y_{j-1})+E[\mathbb{1}(y_{j-1}-D_{j-1} \leq \tilde{R}_j)\tilde{U}_j(\tilde{R}_j)+\mathbb{1}(y_{j-1}-D_{j-1} > \tilde{R}_j)\tilde{U}_j(y_{j-1}-D_{j-1})].$$

It is again readily verified that $\tilde{U}_{j-1}$ is a continuous function of y_{j-1}. If we take the right-hand derivative, and then apply this process recursively (similar to (4)-(6) above), we get that the right-hand derivative of $\tilde{U}_{j-1}$ is

$$\tilde{U}^r_{j-1}(y_{j-1}) = C^r_{j-1}(y_{j-1}) + E[\sum_{k=j}^{T} \mathbb{1}(\tilde{A}_{k,j-1}(y_{j-1}))C^r_k(y_{j-1} - D_{[j-1,k)})]. \tag{7}$$

The events $\tilde{A}_{k,j-1}(y_{j-1})$ are defined with respect to $\tilde{R}_j, \ldots, \tilde{R}_T$ instead of the respective $R_j, \ldots, R_T$ (see (6) above). We get a similar expression for the left-hand derivative of $\tilde{U}_{j-1}$ by replacing C^r_k by C^l_k for each $k = j-1, \ldots, T$. As in the case of U^r_t, the events $\tilde{A}_{jt}$ are defined with respect to weak inequalities. It is easy to verify that the right-hand and left-hand derivatives of the function $\tilde{U}_t$ are bounded between the $-(T-t+1)b_t$ and $(T-t+1)h_t$. The next lemma shows that for each y, there exist explicit computable bounded random variables with expectations equal to $\tilde{U}^r_t(y)$ and $\tilde{U}^l_t(y)$, respectively. This implies that the right-hand and the left-hand derivatives of the function $\tilde{U}_t$ can be evaluated stochastically with high accuracy and high probability (using again the Hoeffding inequality).

Lemma 3.3.7 *For each $t = 1, \ldots, T-1$, $j > t$ and y_t consider the random variable $\tilde{M}^r_{tj}(y_t) = \mathbb{1}(\tilde{A}_{jt}(y_t))(-b_j + (h_j + b_j)\mathbb{1}(D_j \leq y_t - D_{[t,j)}))$. Then $E[\tilde{M}^r_{tj}(y)] = E[\mathbb{1}(\tilde{A}_{jt}(y_t))C^r_j(y_t - D_{[t,j)})]$.*

Proof : First note again that the expectations above are taken with respect to the underlying demand distributions $D_t, \ldots, D_T$. In particular, the base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ that define the events $\tilde{A}_{jt}$ are assumed to be known deterministically. Using conditional expectations we can write

$$E[\tilde{M}^r_{tj}(y)] = E[E[\tilde{M}^r_{tj}(y)|D_{[t,j)}]] =$$
$$E[E[\mathbb{1}(\tilde{A}_{jt}(y_t))(-b_j + (h_j + b_j)\mathbb{1}(D_j \leq y_t - D_{[t,j)}))|D_{[t,j)}]] =$$
$$E[\mathbb{1}(\tilde{A}_{jt}(y_t))E[-b_j + (h_j + b_j)\mathbb{1}(D_j \leq y_t - D_{[t,j)})|D_{[t,j)}]] =$$
$$E[\mathbb{1}(\tilde{A}_{jt}(y_t))C^r_j(y_t - D_{[t,j)})].$$

We condition on $D_{[t,j)}$ and then the indicator $\mathbb{1}(\tilde{A}_{jt}(y_t))$ is known deterministically. In the last equality we use the definition of C^r_j and uncondition. The claim then follows. ■

Considering (7), we immediately get the following corollary.

Corollary 3.3.8 *For each $t = 1, \ldots, T$ and y_t, the right-hand derivative of $\tilde{U}_t$ is given by*
$\tilde{U}^r_t(y_t) = E[-b_t + (h_t + b_t)\mathbb{1}(D_t \leq y_t) + \sum_{j=t+1}^{T} \tilde{M}^r_{tj}(y_t)]$.

Analogously to the random variables $\tilde{M}^r_{tj}$, we define for each $t = 1, \ldots, T$, $j > t$ and y_t the random variable $\tilde{M}^l_{tj}(y_t)$ by replacing the indicator $\mathbb{1}(D_j \leq y_t - D_{[t,j)})$ in the definition of $\tilde{M}^r_{tj}$ by the indicator $\mathbb{1}(D_j < y_t - D_{[t,j)})$. We get the following corollary.

Corollary 3.3.9 *For each $t = 1, \ldots, T$ and y_t, the left-hand derivative of $\tilde{U}_t$ is given by*
$\tilde{U}^l_t(y_t) = E[-b_t + (h_t + b_t)\mathbb{1}(D_t < y_t) + \sum_{j=t+1}^{T} \tilde{M}^l_{tj}(y_t)]$.

Lemma 3.3.7 and Corollaries 3.3.8 and 3.3.9 imply that we can estimate the right-hand and left-hand derivatives of the functions $\tilde{U}_1, \ldots, \tilde{U}_T$ with bounded number of samples. For each $t = 1, \ldots, T$, take N_t samples from the demand distributions $D_t, \ldots, D_T$

that are *independent* of $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$, i.e., independent of the samples taken in previous stages. Let $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ be the respective right-hand and left-hand sample-based estimators of the one-sided derivatives of $\tilde{U}_t$. Note that since the base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ have already been computed based on independent samples from the previous stages, the function $\tilde{U}_t$ and its one-sided derivatives are considered to be deterministic. On the other hand, $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ are random and determined by the specific N_t samples that define them.

To simplify notation, we will assume from now on that $b_t = b > 0$ and $h_t = h > 0$, for each $t = 1, \ldots, T$. To evaluate the right-hand and left-hand derivatives of $\tilde{U}_t$ at a certain point y_t, consider each sample path $d_t^i, \ldots, d_T^i$ (for $i = 1, \ldots, N_t$), evaluate the random variables $-b + (h+b)\mathbb{1}(D_t \le y_t) + \sum_{j=t+1}^{T} \tilde{M}_{tj}^r(y_t)$ and $-b + (h+b)\mathbb{1}(D_t < y_t) + \sum_{j=t+1}^{T} \tilde{M}_{tj}^l(y_t)$, respectively, and then average over the N_t samples. Note that each of the variables $\tilde{M}_{tj}^r$ (respectively, $\tilde{M}_{tj}^l$) can take values only within $[-b, h]$, so $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ are also bounded. By arguments similar to Lemma 3.2.5 (using again the Hoeffding inequality), it easy to compute the number of samples N_t to guarantee that the minimizer is an ϵ_t-point of the function $\tilde{U}_t$ with probability at least $1 - \delta_t$ (for specified accuracy and confidence levels). However, as we have already seen that, in the multiperiod setting, it is also essential to preserve the invariant that $\tilde{R}_t \ge \bar{R}_t$ to ensure that the problem in the next stage is still convex. That is, we wish to find an ϵ_t-point $\tilde{R}_t$ of $\tilde{U}_t$ but with the additional property that $\tilde{R}_t \ge \bar{R}_t$, where again $\bar{R}_t$ is the smallest minimizer of $\tilde{U}_t$.

In turn, we compute $\tilde{R}_t$ in the following way. Given N_t samples, we let $\tilde{R}_t$ be the minimal point with sample right-hand derivative at least $\frac{\epsilon_t}{2}$ (i.e., $\hat{\tilde{U}}_t^r(\tilde{R}_t) \ge \frac{\epsilon_t}{2}$). That is, $\tilde{R}_t := \inf\{y : \hat{\tilde{U}}_t^r(y) \ge \frac{\epsilon_t}{2}\}$. First observe that $\tilde{R}_t$ is well-defined for each $0 < \epsilon_t \le 2h(T - t + 1)$, since the slope of $\tilde{U}_t$ varies from $-b(T - t + 1)$ to $h(T - t + 1)$.

By the definition of $\tilde{R}_t$, it is also clear that $\hat{\tilde{U}}^l_t(\tilde{R}_t) < \frac{\epsilon_t}{2}$. The next Lemma analyzes the required number of samples N_t to guarantee that, with probability at least $1 - \delta_t$, the point $\tilde{R}_t$ is both an ϵ_t-point of $\tilde{U}_t$ and satisfies $\tilde{R}_t \geq \bar{R}_t$.

Lemma 3.3.10 *For each $t = 1, \ldots, T$, define $\tilde{U}_t(y_t)$ be as above with respect to an upper base-stock policy $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$. Then $\tilde{U}_t$ is convex and attains a minimum, where $\bar{R}_t$ is its smallest minimizer. Let $0 < \epsilon_t$ and $0 < \delta_t < 1$ be the respective accuracy and confidence levels. Suppose we generate $N_t \geq 2((b + h)(T - t + 1))^2 \frac{1}{\epsilon_t^2} \log(\frac{2}{\delta_t})$ independent samples of the demands $D_t, \ldots, D_T$ that are also independent of the the base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$, and use them to compute $\tilde{R}_t = \inf\{y : \hat{U}^r_t(y) \geq \frac{\epsilon_t}{2}\}$. Then, with probability at least $1 - \delta_t$, the base-stock level $\tilde{R}_t$ is an ϵ_t-point of $\tilde{U}_t$, and $\tilde{R}_t \geq \bar{R}_t$.*

Proof : The proof follows along the lines of the proof of Lemma 3.2.5 above. First note again that $\tilde{R}_t$ is a random variable that depends on the specific N_t samples and that $\bar{R}_t$ is a deterministic quantity (the smallest minimizer of the function $\tilde{U}_t$ that is assumed to be known deterministically). Observe that the event $[\tilde{R}_t \geq \bar{R}_t] \cap [\tilde{R}_t$ is an ϵ_t- point$]$ contains the event $[\tilde{U}^r_t(\tilde{R}_t) \geq 0] \cap [\tilde{U}^l_t(\tilde{R}_t) \leq \epsilon_t]$. It is then sufficient to show that each of the events $[U^r_t(\tilde{R}_t) < 0]$ and $[U^l_t(\tilde{R}_t) > \epsilon_t]$ has probability at most $\frac{\delta_t}{2}$.

By the optimality and minimality of $\bar{R}_t$, we know that $\bar{R}_t = \inf\{y : \tilde{U}^r_t(y) \geq 0\}$ and that $\tilde{U}^r_t(\bar{R}_t) \geq 0$. This implies that the event $[\tilde{U}^r_t(\tilde{R}_t) < 0]$ is equivalent to the event $[\tilde{R}_t \in (-\infty, \bar{R}_t)]$. For a decreasing nonnegative sequence of numbers $\tau_k \downarrow 0$ define the monotone increasing sequence of events $B_k = [\hat{\tilde{U}}^r_t(\bar{R}_t - \tau_k) \geq \frac{\epsilon_t}{2}]$ (note again that $\hat{\tilde{U}}^r_t(y)$ is a random variable dependent on the specific N_t samples). By the monotonicity of $\hat{\tilde{U}}^r_t$, it is readily verified that $B_k \uparrow \bar{B}$ (where $\bar{B}$ is the limit set), and $[\tilde{R}_t \in (-\infty, \bar{R}_t)] \subseteq \bar{B}$. However, by the Hoeffding inequality (applied to the random samples of $-b + (h+b)\mathbb{1}(D_t \leq \bar{R}_t - \tau_k) + \sum_{j=t+1}^{T} \tilde{M}_{tj}^{r}(\bar{R}_t - \tau_k)$ defined above) and the specific choice of N_t, we conclude that, for each k, we have $Pr(B_k) \leq \frac{\delta_t}{2}$, which implies that $Pr([\tilde{R}_t \in (-\infty, \bar{R}_t)]) \leq Pr(\bar{B}) \leq \frac{\delta_t}{2}$.

Now consider again the function $\tilde{U}_t$, and let q be the maximal point with left-hand derivative at most ϵ_t, i.e., $q = \sup\{y : \tilde{U}_t^l(y) \leq \epsilon_t\}$. Since $\tilde{U}_t^l$ is left continuous, we conclude that $\tilde{U}^l(q) \leq \epsilon_t$. This implies that the event $[\tilde{U}_t^l(\tilde{R}_t) > \epsilon_t]$ is equivalent to the event $[\tilde{R}_t \in (q, \infty)]$. Define the monotone increasing sequence of events $L_k = [\hat{\tilde{U}}_t^l(q+\tau_k) \leq \frac{\epsilon_t}{2}]$. By the monotonicity of $\hat{\tilde{U}}_t^l$ it is clear that $L_k \uparrow \bar{L}$ (where $\bar{L}$ is the limit set), and that $[\tilde{U}_t^l(\tilde{R}_t) > \epsilon_t] \subseteq \bar{L}$. Using again the Hoeffding inequality (now applied to the random samples of $-b+(h+b)\mathbb{1}(D_t \leq \bar{R}_t - \tau_k) + \sum_{j=t+1}^{T} \tilde{M}_{tj}^{l}(\bar{R}_t - \tau_k))$ and the choice of N_t, we conclude that for each k, $Pr(L_k) \leq \frac{\delta_t}{2}$. This implies that $Pr([\tilde{U}_t^l(\tilde{R}_t) > \epsilon_t] \leq Pr(\bar{L}) \leq \frac{\delta_t}{2}$. It is now clear that $\tilde{R}_t \geq \bar{R}_t$ and that $\tilde{R}_t$ is and ϵ_t-point with probability at least $1 - \delta_t$. ■

Note that it is relatively easy to compute $\tilde{R}_t$ above. In particular, it is readily verified that the functions $\hat{\tilde{U}}_t^r$ and $\hat{\tilde{U}}_t^l$ change values in at most $(2(T-t)+1)N_t$ distinct points y_t. This and other properties enable us to compute $\tilde{R}_t$ in relatively efficient ways.

3.3.3 An Algorithm

Next we shall provide a detailed description of the algorithm and a complete analysis of the number of samples required. For ease of notation we will again assume that $h_t = h > 0$ and $b_t = b > 0$ for each $t = 1, \ldots, T$.

For a specified accuracy level ϵ' (where $0 < \epsilon' \leq \frac{1}{2}$) and confidence level δ' (where $0 < \delta' < 1$), let $\epsilon_t' = \frac{\epsilon'}{2T}$ and $\delta_t' = \frac{\delta'}{T}$. For each $t = 1, \ldots, T$, let $\epsilon_t = \frac{\epsilon_t'}{3}\min(b,h)$ and $\delta_t = \delta_t'$. The algorithm computes a base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ in the following way. In stage $t = T, \ldots, 1$, consider the function $\tilde{U}_t$ as defined above with respect to the previously computed base-stock levels $\tilde{R}_{t+1}, \ldots, \tilde{R}_T$ (where $\tilde{U}_T = C_T$). Use the black box to generate $N_t = 2((b+h)(T-t+1))^2 \frac{1}{\epsilon_t^2} \log(\frac{2}{\delta_t})$ independent samples of the demands $D_t, \ldots, D_T$ and compute $\tilde{R}_t = \inf\{y : \hat{\tilde{U}}_t^r(y) \geq \frac{\epsilon_t}{2}\}$. Note that in each stage, the algorithm is using an *additional* N_t samples that are independent of the samples used in previous stages. In the next theorem we show that the algorithm computes a base-stock policy that satisfies the required accuracy and confidence levels.

Theorem 3.3.11 *For each specified accuracy level ϵ' (where $0 < \epsilon' \leq \frac{1}{2}$) and confidence level δ' (where $0 < \delta' < 1$), the algorithm computes a base-stock policy $\tilde{R}_1, \ldots, \tilde{R}_T$ such that with probability at least $1 - \delta'$, the expected cost of the policy is at most $1 + \epsilon'$ times the optimal expected cost.*

Proof : For each $t = 1, \ldots, T$, let I_t be the event that $\tilde{R}_t, \ldots, \tilde{R}_T$ is an upper base-stock policy and that for each $j = t, \ldots, T$, $\tilde{R}_j$ is ϵ_j−point of $\tilde{U}_j$. In particular, by the choice of ϵ_j and Lemma 3.3.6, $\tilde{U}_j(\tilde{R}_j) \leq (1+\epsilon_j')\tilde{U}_j(\bar{R}_j)$, where again $\bar{R}_j$ is the smallest minimizer of $\tilde{U}_j$.

Lemma 3.3.4 implies that if I_1 occurs then $\tilde{R}_1, \ldots, \tilde{R}_T$ is an upper base-stock policy with expected cost at most $\prod_{t=1}^T (1 + \epsilon_t') = (1 + \frac{\epsilon'}{2T})^T \leq 1 + \epsilon'$ (for $\epsilon' \leq \frac{1}{2}$) times the optimal expected cost. It is then sufficient to show that I_1 occurs with probability at least $1 - \delta'$.

Clearly, $I_T \supseteq I_{T-1} \supseteq \cdots \supseteq I_1$. This implies that $Pr(I_1) = Pr(\cap_{t=1}^T I_t)$. It is then sufficient to show that $Pr([\cap_{t=1}^T I_t]^C) \leq \delta'$.

We first show that for each $t = 1, \ldots, T$ the event I_t occurs with positive probability. The proof is done by induction on $t = T, \ldots, 1$. For $t = T$ the claim follows trivially. Now assume that the claim is true for some $1 < t \leq T$ and consider the event I_{t-1}. We then have

$$Pr(I_{t-1}) = Pr(I_t \cap I_{t-1}) = Pr(I_t)Pr(I_{t-1}|I_t).$$

By induction we know that $Pr(I_t) > 0$, so conditioning on I_t is well-defined. Since the samples in each stage are independent of each other, Lemma 3.3.10 and the choice of N_t implies that $Pr(I_{t-1}|I_t) \geq 1 - \delta_t > 0$. The claim then follows.

Now observe that we can write $(\cap_{t=1}^{T} I_t)^C$ as

$$[\cap_{t=1}^{T} I_t]^C = I_T^C \cup [I_T \cap I_{T-1}^C] \cup [I_{T-1} \cap I_{T-2}^C] \cup \cdots \cup [I_2 \cap I_1^C].$$

However, because of the fact that the samples in each stage are independent of each other and the choice of N_t, for each $t = 1, \ldots, T$, Lemma 3.3.10 implies that $Pr(I_T^C) \leq \delta_T$ and $Pr(I_t \cap I_{t-1}^C) = Pr(I_t)Pr(I_{t-1}^C|I_t) \leq \delta_{t-1}$. We conclude that $Pr([\cap_{t=1}^{T} I_t]^C) \leq \sum_{t=1}^{T} \delta_t$ which implies that $Pr(I_1) \geq 1 - \sum_{t=1}^{T} \delta_t = 1 - \delta'$ as required. The proof then follows. ■

The next corollary provides upper bounds on the total number of samples needed from each of the random variables $D_1, \ldots, D_T$, denoted by $\mathcal{N}_t$.

Corollary 3.3.12 *For each $t = 1, \ldots, T$, specified accuracy level $0 < \epsilon' \leq \frac{1}{2}$ and confidence level $0 < \delta' < 1$, the algorithm requires at most $\mathcal{N}_t$ independent samples of D_t, where*

$$\mathcal{N}_t \geq 72\frac{T^2}{\epsilon'^2}(\frac{\min(b,h)}{h+b})^{-2}\log(\frac{2T}{\delta'})\sum_{j=1}^{t}(T-j+1)^2.$$

Observe that the number of samples required is increasing in the periods. In particular, it is of order $O(T^4)$ for the first period and increasing to order of $O(T^5)$ in the last period. The bounds are again independent of the demand distributions but do depend on the square of the reciprocal of $\frac{\min(b,h)}{b+h}$.

We note that the algorithm can be applied in the presence of a positive lead time, per-unit ordering costs and discount factor over time. The exact dynamic program described above can be extended in a straightforward way to capture these features in a way that preserves the convexity of the problem (see [56] for details). Similarly, the shadow dynamic program can still be used to construct an approximate base-stock policy with the same properties as above. Moreover, the bounds on the number of required samples stay very similar to the bounds established above.

3.4 Approximating Myopic Policies

In many cases, finding optimal policies can be computationally demanding, regardless of whether we have access to the demand distributions or not. As a result, researchers and practitioners have paid attention to *myopic policies*. In a myopic policy, we aim, in each period $t = 1, \ldots, T$, to minimize the expected cost (the newsvendor cost) in that period, ignoring the future costs. This provides what is called a *myopic base-stock policy*. As we have already mentioned, myopic policies may not be optimal in general. However, in many cases the myopic policy performs well, and in some cases it is even optimal. In this section, we shall describe a simple and very efficient sample-based procedure that computes a policy that, with high specified confidence probability, has expected cost very close to the expected cost of the myopic policy. In particular, if a myopic policy is optimal then the expected cost of the approximate policy is close to optimal. We let $R_1^m, \ldots, R_T^m$ denote the *minimal myopic policy*, where for each $t = 1, \ldots, T$, the base-stock level R_t^m is the smallest minimizer of $C_t(y)$ the newsvendor cost in that period.

The sample-based procedure is based on solving the newsvendor problems in each one of the periods independently. Consider each of the functions $C_1, \ldots, C_T$, and find, for each one of them, an approximate minimizer by means of solving the SAA counterpart. Let $\tilde{R}_t^m$ be the approximate solution in period t. However, in order to guarantee that the approximate policy has expected cost close the the expected myopic cost, it might not be sufficient to simply take $\tilde{R}_t^m$ to be the minimizer of the corresponding SAA model as discussed in Section 3.2. The problem is that if we approximate the exact myopic base-stock level from above, i.e., get $\tilde{R}_t^m \geq R_t^m$, we might impact the inventory level in the next period in a way that will incur high costs. In turn, we will approximate R_t^m from below, i.e., for each period t, we will compute $\tilde{R}_t^m$ that, with high probability, is an $\epsilon-$point with respect to C_t and is no greater than R_t^m.

The procedure is symmetric to the computation of $\tilde{R}_t$ in Section 3.3 above. Given N_t samples, consider the SAA counterpart and focus on the one-sided derivatives $\hat{C}_t^r$ and $\hat{C}_t^l$. We will compute $\tilde{R}_t^m$ as the maximum point with sample left-hand derivative value at most $-\frac{\epsilon_t}{2}$ (where $0 < \epsilon_t \leq b_t$, i.e., $\tilde{R}_t^m = \sup\{y : \hat{C}_t^l(y) \leq -\frac{\epsilon_t}{2}\}$. By a proof symmetric to the one of Lemma 3.3.10 above, we get the following lemma.

Lemma 3.4.1 *For each $t = 1, \ldots, T$, consider specified $0 < \epsilon_t \leq b_t$ and $0 < \delta_t < 1$. Further consider the SAA counterpart of C_t defined for $N_t \geq 2(b_t + h_t)^2 \frac{1}{\epsilon_t^2} \log(\frac{2}{\delta_t})$ samples. Compute $\tilde{R}_t^m = \sup\{y : \hat{C}_t^l(y) \leq -\frac{\epsilon_t}{2}\}$. Then, with probability at least $1 - \delta_t$, the base-stock level $\tilde{R}_t^m$ is an ϵ_t-point of C_t and $\tilde{R}_t^m \leq R_t^m$.*

Moreover, for each specified accuracy level $0 < \epsilon' \leq 1$ and confidence level $0 < \delta' < 1$, let $\epsilon_t' = \epsilon'$, $\epsilon_t = \frac{\min(b_t, h_t)}{3}\epsilon_t'$ and $\delta_t' = \delta_t = \frac{\delta'}{T}$. Now apply the above procedure for each of the periods $t = 1, \ldots, T$ for number of samples N_t as specified in Lemma 3.4.1 above, and compute an approximate policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$. We claim that, with probability at least $1 - \delta'$, this policy has expected cost at most $1 + \epsilon'$ times the expected cost of the myopic policy.

Theorem 3.4.2 *Consider the policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$ computed above. Then, with probability at least $1 - \delta'$, it has expected cost at most $1 + \epsilon'$ times the expected cost of the* *myopic policy.*

Proof : Let I be the event that, for each $t = 1, \ldots, T$, the base-stock level $\tilde{R}_t^m$ is ϵ_t-point with respect to C_t and $\tilde{R}_t^m \leq R_t^m$ (the corresponding myopic base-stock level). By the choice of N_t, it is readily verified that $Pr(I) \geq 1 - \delta'$. We claim that under the event I, the expected cost of the policy $\tilde{R}_1^m, \ldots, \tilde{R}_T^m$ is at most $1 + \epsilon'$ times the expected cost of the myopic policy.

For each $t = 1, \ldots, T$, let $\tilde{X}_t$ and X_t be the respective inventory levels of the approximate policy and of the myopic policy at the beginning of period t (note that these are two random variables). Then, $E[C_t(\tilde{R}_t^m \vee \tilde{X}_t)]$ and $E[C_t(R_t^m \vee X_t)]$ are the respective expected costs of the approximate policy and of the myopic policy in period t. It is sufficient to show that for each $t = 1, \ldots, T$, we have

$$E[C_t(\tilde{R}_t^m \vee \tilde{X}_t)] \leq (1 + \epsilon')E[C_t(R_t^m \vee X_t)].$$

Condition now on any realization of the demands $D_1, \ldots, D_{t-1}$ which results respective inventory levels $\tilde{x}_t$ and x_t (where these are the respective realizations of $\tilde{X}_t$ and X_t). Then one of the following cases applies:

Case 1. The base-stock level $\tilde{R}_t$ is reachable, i.e., $\tilde{x}_t \leq \tilde{R}_t^m$. The inequality then immediately follows by the fact that $\tilde{R}_t^m$ is an ϵ_t-point of C_t.

Case 2. The inventory level of the approximated policy, $\tilde{x}_t$, is between $\tilde{R}_t^m$ and R_t^m, i.e., $\tilde{R}_t^m \leq \tilde{x}_t \leq R_t^m$. It is readily verified that $\tilde{x}_t$ is also an ϵ_t-point of C_t which implies that the inequality still holds.

Case 3. Finally, consider the case where the inventory level of the approximate policy is above the myopic base-stock, i.e., $\tilde{x}_t > R_t^m$. Observe that under the event I, we know that, with probability 1, that $\tilde{x}_t \leq x_t$. In particular, this implies that $C_t(\tilde{x}_t) \leq C_t(x_t)$. This concludes the proof. ■

Observe that in this case the number of samples required from each demand distribution is significantly smaller and it is of order $O(\log(T))$ instead of $O(T^5)$.

3.5 General Structural Lemma

In this section, we provide a proof of a multi-dimensional version of the key structural Lemma 3.3.6 above. In many stochastic dynamic programs, one of the main challenges is to evaluate the future expected cost that is resulted by the decision being made in the current stage. This cost function is often very complex to evaluate. However, there are cases, where we can derive analytical expressions for gradients or subgradients and estimate them accurately using sample-based methods, similar to the one described above. In such cases, one of the natural issues to address is how to relate first-order information to relative errors. We believe that the following lemma provides effective tools to establish such relations for certain convex objective functions. This can lead to algorithms with rigorous analysis of their worst-case performance guarantees. More specifically, there are cases where we can derive piecewise linear functions that provide a lower bound on the real objective function value (see for example [22]). Piecewise linear approximations are also used in different heuristics for two-stage stochastic models (see, for example, [4]).

The following lemma indicates that in convex minimization models that admit a nonnegative piecewise linear function that lower bounds the original objective function value, there exists an explicit relation between first-order information and relative errors. Thus, we believe that this lemma will have applications in analyzing the worst-case performance of approximation algorithms for stochastic dynamic programs and stochastic two-stage models.

Lemma 3.5.1 *Let $f : \mathbb{R}^m \mapsto \mathbb{R}$ be convex and finite with a global minimizer denoted by y^*. Further assume that there exists a function $\bar{f} : \mathbb{R}^m \mapsto \mathbb{R}$ convex, nonnegative and piecewise linear, such that $\bar{f}(u) \leq f(u)$ for each $u \in \mathbb{R}^m$. Without loss of generality assume that $\bar{f}(u) = \lambda||u||_2$ for each $u \in \mathbb{R}^m$ and some $\lambda > 0$. Let $0 < \epsilon' \leq 1$ and $\epsilon = \frac{\epsilon'\lambda}{3}$. Then if $\hat{y} \in \mathbb{R}^m$ is an $\epsilon-$point (see definition 3.3.5 above) of f, its objective value $f(\hat{y})$ is at most $1+\epsilon'$ times the optimal objective value $f(y^*)$, i.e., $f(\hat{y}) \leq (1+\epsilon')f(y^*)$.*

Proof : The proof follows along the lines of those of Lemma 3.3.6 above. Let $L = \frac{f(y^*)}{\lambda}$ and $\mathcal{B}(0, L) \subseteq \mathbb{R}^m$ be a ball of radius L around the origin. It is straightforward to verify that $y^* \in \mathcal{B}(0, L)$. Now by definition we have $f(y^*) = \lambda L$ and by the fact that $\hat{y}$ is an $\epsilon-$point of f, we know that $f(\hat{y}) - f(y^*) \leq \epsilon||\hat{y} - y^*||_2$.

First consider the case where $\hat{y} \in \mathcal{B}(0, L)$. Then it is readily verified that $||\hat{y}-y^*||_2 \leq 2L$. However, this implies that

$$f(\hat{y}) - f(y^*) \leq \epsilon||\hat{y} - y^*||_2 \leq 2\epsilon L = \frac{2}{3}\epsilon'\lambda L = \frac{2}{3}\epsilon' f(y^*).$$

Now consider the case where $\hat{y} \notin \mathcal{B}(0, L)$, in particular, $||\hat{y}||_2 > L$. Let $d \in \mathcal{B}(0, L)$ be the point on the line between $\hat{y}$ and the origin with $||d||_2 = L$. Since f is convex it is readily verified that the point d is also an $\epsilon-$point of f. Then clearly, the point d satisfies $f(d) - f(y^*) \leq \frac{2}{3}\epsilon' f(y^*)$. Let $x = ||\hat{y} - d||_2$. The rest of the proof proceeds exactly like the proof of Lemma 3.3.6. ■

BIBLIOGRAPHY

[1] Y. Aviv and A. Federgruen. Stochastic inventory models with limited production capacity and varying parameters. *Probability in the Engineering and Information Sciences*, 11:107–135, 1997.

[2] D. Bertsimas and A. Thiele. A robust optimization approach to supply chain management. In *Proceedings of 14th IPCO*, pages 86–100, 2004.

[3] P. Billingsley. *Probability and Measure*. John Wiley & Sons, 1995. Third edition.

[4] J. Birge and S. W. Wallace. A separable piecewise linear upper bound for stochastic linear programs. *SIAM Jouranl of Control and Optimization*, 26:1–14, 1988.

[5] E. W. M. Chan. *Markov Chain Models for multi-echelon supply chains*. PhD thesis, School of OR&IE, Cornell University, Ithaca, NY, January 1999.

[6] M. Charikar, C. Chekuri, and M. Pál. Sampling bounds for stochastic optimization. Unpublished manuscript, 2005.

[7] F. Chen and J. Song. Optimal policies for multi-echelon inventory problems with Markov-modulated demand. *Operations Research*, 49:226–234, 2001.

[8] B. C. Dean, M. X. Goemans, and J. Vondák. Approximating the stochastic knapsack problem: The benefit of adaptivity. In *Proceedings of the 45th Annual IEEE Symposium on the Foundations of Computer Science, 2004.*, pages 208–217, 2004.

[9] L. Dong and H. L. Lee. Optimal policies and approximations for a serial multi-echelon inventory system with time-correlated demand. *Operations Research*, 51, 2003.

[10] S. Dye, L. Stougie, and A. Tomasgard. The stochastic single resource service provision problem. *Naval Research Logistics*, 50:869–887, 2003.

[11] N. Erkip, W. H. Hausman, and S. Nahmias. Optimal centralized ordering policies in multi-echelon inventory systems with correlated demands. *Mangement Science*, 36:381–392, 1990.

[12] A. Federgruen and P. Zipkin. An inventory model with limited production capacity and uncertain demands II. The discounted cost criterion. *Mathematics of Operations Research*, 11:208–215, 1986.

[13] G. Gallego and I. Moon. A min-max distribution newsboy problem: Review and extensions. *Journal of Operational Research Society*, 44:825–834, 1993.

[14] G. Gallego and Ö. Özer. Integrating replenishment decisions with advanced demand information. *Management Science*, 47:1344–1360, 2001.

[15] G. Gallego, J. K. Ryan, and D. Simchi-Levi. Minimax analysis for discrete finite horizon inventory models. *IIE Transactions*, pages 861–874, 2001.

[16] S. Gavireni and S. Tayur. An efficient procedure for non-stationary inventory control. *IIE Transactions*, 33:83–89, 2001.

[17] P. Glasserman and Y. C. Ho. *Gradient estimation via pertubation analysis*. Kluwer Academic Publishers, 1991.

[18] P. Glasserman and S. Tayur. Sensitivity analysis for base-stock levels in multiechelon production-inventory systems. *Management Science*, 41:263–282, 1995.

[19] A. Gupta, M. Pal, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 265–274, 2004.

[20] W. Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58:13–30, 1963.

[21] E. Ignall and A. F. Venott. Optimality of myopic inventory policies for several substitute products. *Managment Science*, 15:284–304, 1969.

[22] T. Iida and P. Zipkin. Approximate solutions of a dynamic forecast-inventory model. Working paper, 2001.

[23] P. Jackson, G. Hurley, R. Levi, R. O. Roundy, and D. B. Shmoys. A computaqtional and theoreticla study of stochatic inventory control policies in the MMFE model. In preparation, 2005.

[24] G. Janakiraman and J. A. Muckstadt. A decomposition approach for a capacitated singel-stage inventory system. Technical Report TR1360, ORIE Department, Cornell University, 2003.

[25] D. Joneja. The joint replenishment problem: new heuristics and worst case performance bounds. *Operations Research*, 38:723–771, 1990.

[26] R. Kapuscinski and S. Tayur. A capacitated production-inventory model with periodic demand. *Operations Research*, 46:899–911, 1998.

[27] R. Kapuscinski and S. Tayur. Optimal policies and simulation based optimization for capacitated production inventory systems. In *Quantative models for supply chain management*, chapter 2. Kluwer Academic Publisher, 1998.

[28] A. J. Kleywegt, A. Shapiro, and T. Homem-De-Mello. The sample average approximation method for stochastic discrete optimization. *SIAM Journal of Optimization*, 12:479–502, 2001.

[29] H. L. Lee, K. C. So, and C. S. Tang. The value of information sharing in two-level supply chains. *Mangement Science*, 46:626–643, 1999.

[30] R. Levi, R. Roundy, and D. B. Shmoys. Sample-based optimization algorithms for stochastic inventory control models. Full version, 2005.

[31] X. Lu, J. S. Song, and A. C. Regan. Inventory planning with forecast updates: approximate solutions and cost error bounds. Working paper, 2003.

[32] R. H. Möhring, F. J. Radermacher, and G. Weiss. Stochastic scheduling problems I: general strategies. *ZOR- Zeitschrift fur Operations Research*, 28:193–260, 1984.

[33] R. H. Möhring, F. J. Radermacher, and G. Weiss. Stochastic scheduling problems II: set strategies. *ZOR- Zeitschrift fur Operations Research*, 29:65–104, 1984.

[34] R. H. Möhring, A. Schulz, and M. Uetz. Approximation in stochastic scheduling: the power of LP-based priority policies. *Journal of the ACM (JACM)*, 46:924–942, 1999.

[35] A. Muharremoglu and J. N. Tsitsiklis. A single-unit decomposition approach to multi-echelon inventory systems. Working paper, 2001.

[36] Ö. Özer and G. Gallego. Integrating replenishment decisions with advance demand information. *Managment Science*, 47:1344–1360, 2001.

[37] Ö. Özer and W. Wei. Inventory control with limited capacity and advanced demand information. *Operations Research*, 52:988–1001, 2004.

[38] G. Perakis and G. Roels. The distribution-free newsvendor: Inventory management with limitted demand information. Unpublished manuscript, 2005.

[39] A. Referee. Referee report. Personal communication with the an anonymous referee, 2005.

[40] R. T. Rockafellar. *Convex Analysis*. Princeton University Press, 1972.

[41] A. Ruszczyński and A. Shapiro. Stochastic programming models. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in operations research and management science*, chapter 6. Elsevier, 2003.

[42] S. Axsäter. Simple solution procedures for a class of two-echelon inventory problems. *Operations Research*, 38:64–69, 1990.

[43] S. Axsäter and P. Lundell. In process safety stock. In *Proceedings of the 23rd IEEE Conference on Decision and Control*, pages 839–842, 1984.

[44] H. Scarf. A min-max solution to an inventory problem. In K. J. Arrow, S. Karlin, and H. Scarf, editors, *Studies in the mathematical theory of inventory and production*, chapter 12, pages 201–209. Stanford University Press, 1958.

[45] A. Shapiro. Monte Carlo sampling methods. In A. Ruszczyński and A. Shapiro, editors, *Stochastic Programming*, volume 10 of *Handbooks in operations research and management science*, chapter 6. Elsevier, 2003.

[46] A. Shapiro and T. Homem-De-Mello. On the rate of convergence of monte carlo approximations of stochastic programs. *SIAM Journal of Optimization*, 11:70–86, 2000.

[47] A. Shapiro, T. Homem-De-Mello, and J. Kim. Conditioning of convex piecewise linear stochastic programs. *Mathematical Programming*, 94:1–19, 2002.

[48] A. Shapiro and A. Nemirovski. On the complexity of stochastic programming problems. Unpublished manuscript, 2005.

[49] D. B. Shmoys and C. Swamy. Stochastic optimization is (almost) as easy as deterministic optimization. In *Proceedings of the 45th Annual IEEE Symposium on the Foundations of Computer Science, 2004.*, pages 228–237, 2004.

[50] J. Si, A. G. Barto, P. W. B, and D. W. II. *Handbookof learning and approximate dynamic programming*. Wiley&Sons, INC., Publications, 2004.

[51] E. A. Silver and H. C. Meal. A heuristic selecting lot-size requirements for the case of a deterministic time varying demand rate and discrete opportunities for replenishment. *Production and Inventory Management*, 14:64–74, 1973.

[52] J. Song and P. Zipkin. Inventory control in a fluctuating demand environment. *Operations Research*, 41:351–370, 1993.

[53] L. Stougie and M. H. van der Vlerk. Approximation in stochastic integer programming. Technical Report SOM Research Report 03A14, Eindhoven University of Technology, 2003.

[54] C. Swamy and D. B. Shmoys. Sampling-based approximation algorithms for multi-stage stochastic optimization. To appear in FOCS 2005, 2005.

[55] A. F. Veinott. Optimal policy for a multi-product, dynamic, non-stationary inventory problem. *Management Science*, 12:206–222, 1965.

[56] P. H. Zipkin. *Foundation of inventory management*. The McGraw-Hill Companies, Inc, 2000.