Data-driven Approximation for Joint Pricing and Inventory-Control

Van-Anh Truong  
School of ORIE, Cornell University, Ithaca, NY 14853, truong@orie.cornell.edu,

Robin O. Roundy  
School of ORIE, Cornell University, Ithaca, NY 14853, robin@orie.cornell.edu,

Subject classifications: decision analysis: theory; dynamic programming; inventory/production: approximations/heuristics

Area of review: revenue management
1. Abstract

We propose a new data-driven algorithm called ICP for computing a joint inventory and pricing policy for a multi-stage, single-product system under independent demands. We obtain a polynomial bound for the number of samples required to compute the optimal policy to any degree of accuracy with high certainty. Our approach is truly parameter-free, in the sense that we make neither parametric assumptions about the form of the objective function, nor distributional assumptions about the underlying randomness. In particular, we do not use an a priori model of the demand as a function of the price. We allow capacity and minimum-order-quantity constraints on the inventory ordered in each period under stronger assumptions.

2. Introduction

The joint inventory-control and pricing problem over a finite horizon is an important problem at the interface of Operations and Marketing. It has received much interest in the academic literature from the early development of Inventory Theory (see Whitin (1955)). In all of the literature, inventory and pricing decisions are made assuming full knowledge of the distribution of the demand as a function of the price. In practice, however, the price-demand relationship is not known a priori. Even when demand data have been collected, the selection of the most appropriate model for the relationship remains subject to debate.

In this paper, we develop the first non-parametric approach to the joint inventory and pricing problem in the presence of stochastic demands whose relationship to the price is unknown. In an algorithm called the ICP Algorithm, we apply a new stochastic maximization procedure, called the Gradient Bound Algorithm, to compute near-optimal prices and inventory ordering quantities directly from (random) price-demand samples. We show that a polynomial number of samples suffices to compute the optimal policy to any given error tolerance and certainty level. When the price is not a decision variable, we show that an asymmetric version of the Gradient Bound algorithm can be used to exploit the compact structure of the optimal policy to compute an alternative approximation, and that this approach can be easily adapted to more general classes of problems than inventory problems.

We consider a seller managing a single product in a periodic-review setting. Demands over the periods are stochastic, but independent. The distribution of demand in any period is a function of the price set in that period. Orders for instantaneous replenishment of inventory may be placed at the beginning of each period. Inventory left over at the end of a period is carried over to the next whereas stockouts are fully backlogged. The seller’s problem is to determine an inventory-ordering
and pricing policy to maximize profit, which is revenue earned, minus costs due to ordering, holding, and backordering inventory. Given the concavity of the expected profit function, we show that our Gradient Bound (GB) algorithm for maximizing a stochastic concave function can be used in a recursive fashion to compute an \((\delta, p)\)-Probably Approximately Correct (PAC) policy using a polynomial number of samples. That is, given any error tolerance \(\delta > 0\) and certainty level \(p < 1\), a polynomial number of samples are required to compute a policy \(P\) such that with probability at least \(p\), the expected profit of \(P\) is no more than \(\delta\) below that of the optimal policy. When the price is not a decision variable, we develop and apply an asymmetric version of GB, which finds an approximate maximizer either to the left or to the right of the true maximizer(s), to significantly reduce the number of samples required. We show that the approach works for a much more general class of problems than inventory problems.

There has been substantial research on the joint pricing and inventory control problem. The literature is summarized in two recent survey papers, Yano and Gilbert (2003) and Elmaghraby and Keskinocak (2003).

The papers most closely related to ours are those that consider a seller managing a single product in a finite-horizon, multi-period setting and facing stochastic demands that are independent of prior sales. These works are highly parameter-dependent, modeling the demand as an explicit function of the price. Usually, the demand is linear in the price, with an additive or multiplicative random component whose distribution is known. The focus of these works is on studying the structure of the optimal policy. Federgruen and Heching (1999), Thowsen (1975), and Zabel (1970) determine the optimal inventory and pricing policy in the presence of convex backlogging, holding, and ordering costs. By considering variations on demand, cost structure, lost sales or backlogging, and production lead time, all three papers find that a base-stock list-price (BSLP) policy is optimal for a wide range of settings. A BSLP policy is defined as follows: (i) if the inventory at the start of period \(t\), \(x_t\), is less than some base-stock level \(y_t^*\), produce enough to bring the inventory level up to \(y_t^*\) and charge \(p_t^*\), (ii) if \(x_t > y_t^*\), produce nothing and offer the product at a discounted price of \(p_t^*(x_t)\) where \(p_t^*(x_t)\) is decreasing in \(x_t\). When there is an additional fixed component to the ordering cost, Chen and Simchi-Levi (2004) prove that the profit function is \(k\)-concave and a \((s, S, p)\) policy is optimal when the random term in the price-demand function is additive. Under this policy, whenever the on-hand inventory at the start of period \(t\), \(x_t\), goes below \(s\), the seller replenishes up to level \(S\) and charges price \(p\); otherwise, he orders nothing, and charges price \(p(x_t)\). For more general demand functions, the optimal policy does not necessarily have this form. Using the concept of symmetric \(k\)-concavity and assuming that the random term in the price-demand
function is multiplicative, they show that an \((s, S, A, p)\) policy is optimal. Under this policy, if the inventory at the start of period \(t\), \(x_t\), is less than \(s_t\), or if \(x_t \in A_t\) where \(A_t \subset [s_t, \frac{s_t + S_t}{2}]\), then an order of size \(S_t - x_t\) is made and the seller sets the price equal to \(p(S_t)\). Otherwise, no order is placed and the seller sets a price of \(p(x_t)\).

A significant part of the literature uses a Bayesian framework to address the inventory control problem (without pricing) under the assumption that the distribution of the demand is unknown. Scarf (1960), Iglehart (1964) and Azoury (1985) use a Bayesian framework to estimate the demand parameters and to adaptively update the replenishment quantities as the demand information becomes available.

Sample-based techniques have been applied to supply-chain problems in the past mainly as a computational tool, without results regarding the number of samples required to guarantee, with high confidence, a solution with small relative (or additive) error. Notably, Infinitesimal Perturbation Analysis (IPA) has been very successfully used in multi-stage stochastic supply-chain problems to compute policies that can be characterized in a compact way with a relatively small set of decision parameters (see Glasserman and Ho (1990), Glasserman and Tayur (1995) and Kapuscinski and Tayur (1998) for examples). In this method, the performance measures of interest 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 with respect to the decision parameters. More specifically, independent samples are drawn from the demand distributions, where each sample consists of a path of demands over the horizon. The expected cost is then approximated by averaging the costs arising from the sample paths. A recursion is used to evaluate the partial derivative of each performance measure with respect to each decision parameter. 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. Kunnumkal and Topaloglu (2006) consider inventory-control problems for which base-stock policies are known to be optimal and propose stochastic approximation methods to compute the optimal base-stock levels, using only samples of the demand random variables. In each period, they find approximate gradients of the value function by using approximate base-stock levels for future periods. The gradient information is then incorporated into a search for the optimal base-stock level. They prove that the iterates of their methods converge to the optimal base-stock levels.
A different sample-based approach to multi-stage stochastic supply chain problems is to adaptively control the difference between the expected cost of the policy proposed and the expected optimal cost that would have been incurred if the distribution of the random variables were known. This measure is known as the regret of the algorithm. The total regret from time 0 to the present period $t$ is usually shown to be sublinear in $t$, so that the average regret per period converges over time to 0. The aim here is to show that the average expected cost converges to the optimal expected cost over time. This criterion is weaker than what is required by the PAC approach, in which the costs must be shown to converge in probability. Chang, Fu, Hu, and Marcus (2005) develop an adaptive sampling algorithm for solving a lost-sales, finite-horizon inventory problem using a multi-armed bandit approach. They assume a small finite action space and apply a Multi-Armed Bandit algorithm as subroutine. Huh and Rusmevichientong (2006b) consider stochastic inventory planning with lost sales, where the manager only observes the sales quantities in each period, but not the lost sales. The decision in each period depends only on historical sales data and not on any distributional information about demand. They propose non-parametric adaptive policies that generate ordering decisions over time and give bounds on the regret as a decreasing function of the number of stages. Huh and Rusmevichientong (2006a) consider the capacity allocation problem, where the manager must allocate a fixed capacity among several demand classes that arrive sequentially in the order of increasing fares. The objective is to maximize expected revenue when the demand distributions are unknown and the manager has access only to historical sales. They develop an adaptive algorithm for setting protection levels based on historical sales, showing that the average expected revenue converges to the optimal revenue, and establish the rate of convergence. Huh et al. (2006) study the problem of finding the best base-stock policy in a periodic-review single-location single-product inventory system with lost sales and positive replenishment lead times. They assume that the manager does not know the demand distribution a priori, but must make the replenishment decision in each period based only on the past sales (censored demand) data. They develop a nonparametric adaptive algorithm that generates a sequence of order-up-to levels whose regrets converge to 0.

A number of PAC algorithms exist for multi-stage stochastic programs. However, the complexity, model used, and problem features in these algorithms are different from ours in several important ways. Swamy and Shmoys (2005) show that the Sample Average Approximation method (see more in Section 4) is 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. Levi, Roundy, and Shmoys develop a distribution-independent, polynomial-time
approximation algorithm for the inventory replenishment problem by approximating a subgradient of the value function in each period. Rather than computing a policy whose cost is within $\delta$ of the optimal cost, they obtain one whose cost is within $100\%$ of the optimal cost. Their approach relies on the observability of the demand, the linearity of the costs, and the explicit characterization of the newsvendor solution using a critical fractile of the demand distribution.

We add several important new dimensions to the literature on sample-based techniques for supply-chain management problems. We introduce pricing as a decision variable; this is the first sample-based algorithm for inventory management with a multi-dimensional action space in each time period. Our algorithms encompass a larger class of problems, allowing capacity limitations, minimum-purchase quantities and very general concave profit functions. Finally, our error tolerance is expressed in absolute terms, rather than as a percentage of the optimal cost.

Our algorithms are based on function values. They do not need access to derivatives or subgradients. They do not make use of explicit functional forms or parametrization of the costs, which makes them truly non-parametric algorithms. For pricing problems, this feature is particularly important because of the difficulty of choosing a functional representation for the relationship between the demand distribution and the price.

The paper is organized as follows. In Section 3, we specify our basic model for the joint pricing and inventory control problem. We also state assumptions about the problem that we assume throughout the paper. In Section 4, we develop the GB algorithm for maximizing a concave stochastic function. We apply the GB algorithm to determine a PAC policy for inventory control and pricing in Section 5. Next, in Section 6, we develop asymmetric variations of GB to find approximate maximizers that fall to one side of the true maximizer(s). We show an alternative method that makes use of the asymmetric GB algorithms to solve the inventory problem in Section 7. We conclude the paper in Section 8.

3. The Model

In this section, we specify our mathematical model and notation. We formulate the pricing and inventory-control problem as a dynamic program. We state assumptions about the structure of this problem that we assume to hold throughout the paper.

Consider the problem of managing the price and inventory of a single item over a finite horizon of $1 \ldots T$. At the beginning of each period, a replenishment decision has the effect of increasing the inventory by the replenishment quantity and a pricing decision has the effect of changing the demand for the item in the period. Let $x_t$ be the inventory position at the beginning of period $t$. 
The replenishment decision brings the inventory position to some $y_t \geq x_t$. The pricing decision $r_t$ induces a demand $D_t(r_t, \xi_t)$ to occur in period $t$. Both decisions take effect instantaneously. Initially, there is no constraint on the replenishment quantity. We show how to relax this assumption at the end of Section 5. After the decisions are made, the demand $D_t(r_t, \xi_t)$ occurs and is satisfied by any inventory on hand. Unsatisfied demand is fully backlogged and remaining inventory is carried over to the next period. The inventory position becomes $y_t - D_t(r_t, \xi_t)$ at the end of period $t$. This amount is the same as the inventory position at the beginning of period $t + 1$, $x_{t+1}$.

We assume demands are independent over time and unaffected by past decisions. The demand is an unknown function $D_t(r_t, \xi_t)$ of the prevailing price, $r_t$ and a random variable $\xi_t$, where the sequence $\{\xi_t\}_{t=1}^T$ is independent (but not necessarily identically distributed).

The objective is to find a pricing and inventory-replenishment strategy to maximize the total expected profit, defined to be the revenue minus all costs, over the horizon. We allow a general structure for the profit function. Given $(x_t, y_t, r_t)$, the one-period profit in $t$ is a function $f_t(y_t, r_t, D_t(r_t, \xi_t))$, where $\xi_t$ is a random variable with unknown distribution. This profit function is general enough to capture the traditional revenue obtained from sales in the period, together with the traditional holding, and backlogging costs. Although the traditional linear purchase cost $-c_t(y_t - x_t)$ does not appear in our expression for the one-period profit, it is easy to show that if $c_t$ is non-zero for some $t$, a simple transformation of the problem produces an equivalent problem in which $c_t = 0$. Under this transformation, the new one-period profit is

$$
\tilde{f}_t(y_t, r_t, D_t(r_t, \xi_t)) = -c_t y_t + f_t(y_t, r_t, D_t(r_t, \xi_t)) + c_{t+1}(y_t - D_t(r_t, \xi_t)).
$$

Given this profit structure, our basic inventory and pricing problem has the following formulation in period $t$:

$$
v_t(x_t) = \max_{0 \leq y_t \leq \bar{r}} E\left[f_t(y_t, r_t, D_t(r_t, \xi_t)) + v_{t+1}(y_t - D_t(r_t, \xi_t))\right], \quad t = 1 \ldots T, \quad (1)
$$

$$
v_{T+1}(\cdot) = 0. \quad (2)
$$

where $\bar{r}$ is an upper bound on the price, and $\xi_t$ is a random variable whose distribution is unknown. We assume, without burdening the notation, that in the above equations, the optimization over $r$ takes priority over the optimization over $y$. That is, we optimize first with respect to $r$, then with respect to $y$.

The function $v_t(\cdot)$ is commonly called the value function at stage $t$. The value function gives the maximum expected profit for stages $t \ldots T$, given that the inventory position at the beginning of period $t$ is $x_t$. 
We do not know the analytical form of the one-period profit. Instead, we assume that there is an oracle that allows us to access, for each \((y,r)\), unbiased samples of the time-\(t\) profit \(f_t(y,r,D_t(r,\xi_t))\). This requirement is a very weak one, since we only require black-box knowledge of the profit function, no distributional information about the random variables \(\{\xi_t\}_t\), and only indirect observation of the randomness (i.e., we observe the values \(f_t(y,r,D_t(r,\xi_t))\), but not necessarily \(\xi_t\)).

We impose minimal structure on the profit function, namely:

**Assumption 1.** For each \(t\) and \(r\), \(D_t(r,\xi_t)\) takes values in a bounded interval of length \(\lambda\) with probability 1. Moreover, for each \(\xi_t\), \(D_t(r,\xi_t)\) is concave in \(r\).

**Assumption 2.** For each \(t\) and \(\xi_t\), \(f_t(y,r,D_t(r,\xi_t))\) is concave in \((y,r)\). Further, \(f_t(y,r,d)\) and \(D_t(r,\xi_t)\) are Lipschitz continuous. That is, there exist constants \(\kappa_d\), \(\kappa_f^d\), \(\kappa_f^r\), and \(\kappa_f^y\) such that

\[
|f_t(y_1,r^1,d^1) - f_t(y_2,r^2,d^2)| \leq \kappa_f^d|y_1 - y_2| + \kappa_f^r|r^1 - r^2| + \kappa_f^d|d^1 - d^2|, \tag{3}
\]

\[
|D_t(r^1,\xi_t) - D_t(r^2,\xi_t)| \leq \kappa_d|r^1 - r^2|. \tag{4}
\]

for all \(y_1, y_2, r^1, r^2, d^1, d^2, \) and \(\xi_t\).

We will use \(\kappa = (\kappa_f^d + \kappa_f^r + \kappa_f^y)\max(1,\kappa_d)\). The last two of these assumptions are satisfied by all of the models of joint inventory and pricing that we have summarized in the Introduction.

**4. Maximizing an Unknown Stochastic Concave Function: the Gradient Bound Algorithm**

In this section we develop the Gradient Bound (GB) Algorithm that will be applied later to the inventory control and pricing problem. The GB algorithm computes a decision \(\hat{x}\) that, with high probability, approximately maximizes a concave function \(f(x) = \mathbb{E}[g(x,\xi)]\) over a compact domain. GB finds an approximate maximizer by examining only black-box samples of the function \(g(x,\xi)\) at any point \(x\), where \(\xi\) is random. We show that it uses a polynomial number of samples in order to find a PAC solution to the problem of maximizing \(f(x)\).

A maximization problem of this type may be formulated as a two-stage stochastic program, usually a linear program. If the random data have a discrete distribution with a finite (and not too large) number of possible realizations called scenarios, with corresponding probabilities, then the two-stage linear program may be formulated as one large linear program. The linear program may have a certain block structure that makes it amenable to decomposition methods such as the L-shaped method developed by Slyke and Wets (1969). See the books by Kall and Wallace (1994) and Birge and Louveaux (1997) for thorough discussions of these approaches.
Two-stage stochastic programs having a large number of scenarios can be solved by using sampling methods, such as the Sample Average Approximation (SAA) method. Following the SAA method, a random sample $\xi_1, \ldots, \xi_N$ of $N$ realizations of $\xi$ is generated, and the expected value function $f(x)$ is approximated by the sample average function $f_N(x) = \sum_{i=1}^{N} g(x, \xi_i)$. Kleywegt et al. (2001) consider the SAA method and show that for two-stage discrete stochastic programs, the optimal value of the SAA problem converges to the optimal value of the original problem with probability 1 as the number of samples grows. They also use large-deviation results to derive PAC bounds. Their result relies on the finiteness of the decision space. They also assume that samples can be observed directly and that the functional form of the objective is known. That is, given a number of samples, it is possible to solve the resulting SAA problem, which is a deterministic problem. Shapiro et al. (2002) focus on the special class of two-stage stochastic models in which the optimal solution to the SAA model is always an optimal solution to the original problem. Under the assumptions that the objective function is piece-wise convex, the decision space is polyhedral, the probability distribution has finite support, and that the original problem has a unique optimal solution, they use large-deviation results to show that this probability converges to 1 exponentially fast as the number of samples grows. Swamy and Shmoys (2005) consider a class of two-stage stochastic linear programs and develop a PAC bound for the number of samples required that is independent of the variance of the probability distribution. Charikar et al. (2005) propose a simpler proof and extended the class of problems to which this result applies. Both approaches take sufficiently many samples so that the SAA function uniformly approximates the true objective over the decision space.

Shapiro and de Mello (1998) consider a two-stage stochastic LP with recourse, where the second-stage objective function is known and the second-stage random variable has a known continuous density. They use the likelihood ratio transformation to compute an approximation that is smooth. They then repeatedly apply a deterministic programming algorithm to the approximation, each time for several steps, then re-approximate. Nesterov and Vial (2000) prove PAC bounds on the number of samples needed to optimize a convex stochastic function by using stochastic gradient optimization. Dyer et al. (2002) propose an algorithm that samples the function value uniformly in a ball about the starting point at each iteration until finding an improvement, so that first and second order information about the function are not needed. They require approximate function evaluation and a feasibility oracle.

Flaxman et al. (2005) take an online approach to stochastic minimization, where an unknown sequence of cost functions is given, one in each period $t$. The sequence may be realizations of
a single unknown random function, or generated adaptively by an adversary. At each period, a feasible point must be chosen, after which the \(t\)-th cost function evaluated at the point (but not the function itself) is revealed. They use an approximation of the gradient that is computed from evaluating the cost at a single (random) point. They show that this biased estimate is sufficient to approximate gradient descent on the sequence of functions and bound the expected regret, or the difference between the cost achieved by following this strategy and any cost achieved with the benefit of hindsight. For references on related papers, see Flaxman et al. (2005).

We will first give an outline of our algorithm. We then develop a subroutine, called the GB subroutine, that is performed at each iteration of the algorithm. We then present the algorithm by building on the subroutine and derive the sample bound that the algorithm achieves.

Formally, let the decision space \(X = [a,b]\) be a finite interval in \(\mathbb{R}\). Note that for a given \(x \in X\), \(g(x,\xi)\) is a random variable. We assume that for each \(x\), there is an interval of length \(M\) containing \(g(x,\xi)\) with probability 1. Our problem is to find \(x^*\) such that

\[
x^* = \arg\max_{x \in X} f(x) = \arg\max_{x \in X} E[g(x,\xi)].
\]

The GB algorithm is based on the well-known Golden Section Search (see Press et al. (1999)). We have adapted the search for stochastic maximization problems, where exact function evaluation is not possible. At each state in the approximation, given a search interval, GB obtains “confidence intervals” around the function values at two well-placed probe points, \(x_1\) and \(x_2\), where \(x_1 < x_2\). It compares these confidence intervals to arrive at one of three possible decisions. Either the search interval can be reduced from the left, or it can be reduced from the right, or the search terminates and any point in the current search interval can be selected to be \(\hat{x}\). In the first case, \(x_1\) becomes the left end point of the new search interval and \(x_2\) becomes one of the probe points in the next iteration. Another probe point is chosen to the right of \(x_1\). In the second case, \(x_2\) becomes the new right end point, \(x_1\) is one of the new probe points, and another probe point is chosen to the left of \(x_1\). The probe points are chosen such that in each iteration except for the last, the search interval can be reduced by a constant ratio of \(\phi\), where \(\phi\) is the golden search ratio: \(\frac{\sqrt{5} - 1}{2}\). For convenience, we define \(\bar{\phi} = 1 - \phi\).

Let \(N(\delta,p) = \frac{M^2}{2\delta^2}\log\left(\frac{2}{p}\right)\). The following GB Subroutine performs a single iteration of the GB algorithm on a concave function \(f(x) = E[g(x,\xi)]\):

**Gradient Bound (GB) Subroutine**

**Input:**
1. Target uncertainty bound $p$;
2. Search interval $[a, b]$;
3. $M$ such that for each $x$, an interval of length $M$ contains $g(x, \xi)$ with probability 1;
4. $\delta > 0$;
5. $N(\delta, p)$ independent samples of $g(x_i, \xi)$ for $i = 1$ or $i = 2$, where $x_1$ is such that $x_1 - a = (b - a)\bar{\phi}$ and $x_2$ is such that $b - x_2 = (b - a)\bar{\phi}$.

**Output:**
1. Interval $[a', b'] \subset [a, b]$ such that $b' - a' = (b - a)\Phi$;
2. $N(\delta, \frac{\pi}{2})$ independent samples of $g(x'_i, \xi)$ for $i = 1$ or $2$, where $x'_1$ is such that $x'_1 - a' = (b' - a')\Phi$ and $x'_2$ is such that $b' - x'_2 = (b' - a')\Phi$;
3. One of the followings two results:
   (a) A guarantee that
   $$P[\arg\max_{[a, b]} f(x) \in [a', b']] \geq 1 - p.$$
   (b) A point $(x_i, m_i)$, $i = 1$ or 2, such that $x_i \in (a', b')$ and with probability at least $1 - p$,
   $$f(x_i) \in [m_i, m_i + \delta] \text{ and } \max_{[a, b]\setminus[a', b']} f(x) \leq m_i + \frac{\Phi \delta}{2\Phi - 1}.$$

**Subroutine:**
1. If samples at $x_1$ are given, draw $N(\delta, \frac{\pi}{2}) - N(\delta, p)$ additional independent samples of $g(x_1, \xi)$ and $N(\delta, \frac{\pi}{2})$ independent samples of $g(x_2, \xi)$. Similarly, if samples at $x_2$ are given then obtain $N(\delta, \frac{\pi}{2})$ independent samples of $g(x_1, \xi)$ for $i = 1, 2$ in the same way. Let $I_i = [m_i, m_i + \delta]$ denote the interval of length $\delta$ centered at the sample mean $y_i$ of the $N(\delta, \frac{\pi}{2})$ samples of $g(x_i, \xi)$.
2. If $m_1 > m_2 + \delta$ then return the interval $[a, x_2]$ and all samples of $g(x_1, \xi)$.
3. Else if $m_2 > m_1 + \delta$ then return the interval $[x_1, b]$ and all samples of $g(x_2, \xi)$.
4. Else if $m_1 \leq m_2 \leq m_1 + \delta$ then return $(x_2, m_2)$, the interval $[x_1, b]$, and all samples of $g(x_2, \xi)$.
5. Else if $m_2 \leq m_1 \leq m_2 + \delta$ then return $(x_1, m_1)$, the interval $[a, x_2]$ and all samples of $g(x_1, \xi)$.

**Lemma 1.** If $m_1 > m_2 + \delta$ then $P[\arg\max_{[a, b]} f(x) \in [a, x_2]] \geq 1 - p$.

By symmetry, we have

**Lemma 2.** If $m_2 > m_1 + \delta$ then $P[\arg\max_{[a, b]} f(x) \in [x_1, b]] \geq 1 - p$.

**Lemma 3.** If $m_2 \leq m_1 \leq m_2 + \delta$ then

$$P[f(x_1) \in [m_1, m_1 + \delta] \text{ and } \max_{[x_2, b]} f(x) \leq m_1 + \frac{\delta \phi}{2\phi - 1}] \geq 1 - p.$$
Again by symmetry,

**Lemma 4.** If \( m_1 \leq m_2 \leq m_1 + \delta \) then

\[
P[f(x_2) \in [m_2, m_2 + \delta] \text{ and } \max_{x \in [a, x_1]} f(x) \leq m_2 + \frac{\delta \phi}{2 \phi - 1}] \geq 1 - p.
\]

The output guaranteed by the GB subroutine is summarized in the following theorem:

**Theorem 1.** The GB Subroutine returns the correct output using \( O\left(\frac{M^2}{\epsilon^2} \log(\frac{1}{p})\right) \) samples.

**Proof.** That the subroutine guarantees the stated output follows directly from Lemmas 1 to 4. It suffices to check the number of samples required by subroutine. The work requires fewer than \( 2N(\delta, \frac{p}{2}) = O\left(\frac{M^2}{\epsilon^2} \log(\frac{1}{p})\right) \) samples. \(\square\)

**Gradient Bound (GB) Algorithm**

**Input:**
1. Concave function \( f(x) = E[g(x, \xi)] \);
2. Constant \( M \) such that for each \( x \), \( g(x, \xi) \) falls in an interval of length \( M \) with probability 1;
3. Target error bound \( \delta \);
4. Target uncertainty bound \( p \);
5. Optimality bound \( \epsilon \) corresponding to \( \delta \). That is, \( \epsilon \) is such that \( |x - x^*| < \epsilon \) implies \( |f(x) - f(x^*)| < \delta \) for all \( x \) and \( x^* \), where \( x^* \) maximizes \( f(x) \);
6. Bounded domain \([a, b]\) of the objective function \( f(\cdot) \), where \( b - a = L \).

**Output:** A pair \((\hat{x}, \hat{m})\) such that with probability at least \( 1 - p \),

\[
|f(\hat{x}) - f(x^*)| \leq 2\delta \text{ and } \hat{m} \leq f(\hat{x}) \leq \hat{m} + \delta.
\]

**Algorithm:**

Initially, \( k = 1, a_1 = a, b_1 = b \). Let \( \delta = \frac{(2\phi - 1)\delta}{\phi} \). Set \( z_1 = a_1 + (b_1 - a_1)\bar{\phi} \). Initially, \((\hat{x}, \hat{m}) = (z_1, -\infty)\).

Draw \( N(\bar{\delta}, \frac{\bar{\phi}}{2}) \) independent samples of \( g(z_1, \xi) \).

1. If \( b^k - a^k < \epsilon \) then locate an interval of length \( \bar{\delta} \) centered at the mean of the samples at \( x^k \). Let \( z^k \leftarrow x^k \). Let \( m^k \) be the lower end point of that interval. If \( m^k > \hat{m} \) then set \((\hat{x}, \hat{m}) = (x^k, m^k)\). STOP.

2. Otherwise, run the subroutine with input \([a^k, b^k], p^k = \frac{p}{2^k}, \delta\), and the samples of \( g(x_k, \xi) \). Let \((x^k, m^k)\) be the point output, if there is one, of the subroutine. If \( m^k > \hat{m} \) then set \((\hat{x}, \hat{m}) = (x^k, m^k)\).

3. Let \( k \leftarrow k + 1, [a^k, b^k] \leftarrow [a', b'] \). Let \( z^k \) be the point at which samples were returned in the subroutine at the end of Step 2. Go to Step 1.

**Theorem 2.** The GB algorithm returns the correct output by using \( O\left(\frac{M^2}{\epsilon^2} \log(\frac{L}{\epsilon^2}) \log(\frac{1}{p^2})\right) \) samples.
5. Multi-Period Inventory and Pricing

In this section, we develop an algorithm, called the ICP Algorithm, for the inventory control and pricing problem. The ICP algorithm uses the GB algorithm as a subroutine. Recall that the GB algorithm applies to a one-stage optimization problem. To use it in a multi-stage setting, where there is a recourse structure, we will need to work backward from the end of the horizon. At each stage, we will compute an approximation to the value function at stage $t$ using information about the value functions at later stages. We will develop this approach, prove its correctness, and compute its cost in terms of the number of samples required. Finally, we will comment on the applicability of the approach to problems in which there is a constraint on the replenishment quantity (e.g., capacity or minimum order quantity).

Recall the formulation of the problem:

$$v_{t-1}(x_{t-1}) = \max_{y_{t-1} \geq r_{t-1}} \mathbb{E} \left[ f_{t-1}(y_{t-1}, r_{t-1}, D_{t-1}(r_{t-1}, \xi_{t-1})) + v_t(x_t) \right]$$

s.t. $x_t = y_{t-1} - D_{t-1}(r_{t-1}, \xi_{t-1})$,

$$v_{T+1}(\cdot) = 0.$$  \hspace{1cm} (5)

Let

$$G_{t-1}(y, r, \xi_{t-1}) = f_{t-1}(y, r, D_{t-1}(r, \xi_{t-1})) + v_t(y - D_{t-1}(r, \xi_{t-1})), \ t \leq T,$$  \hspace{1cm} (6)

and let

$$V_t(y) = \max_{0 \leq r \leq \bar{r}} \mathbb{E}[G_t(y, r, \xi_t)], \ t = 1, \ldots, T.$$  \hspace{1cm} (7)

Thus, $v_t(x) = \max_{y \geq x} V_t(y)$. Assumption 1 implies the following properties for the functions in each period:

**Theorem 3.** For each $t$ and $\xi_t$, $G_t(\cdot, \cdot, \cdot, \xi_t)$ is concave and $v_t(\cdot)$ is concave and decreasing.

**Lemma 5.** $G_t$ satisfies the following conditions:

1. For each $y^1, y^2, r,$ and $\xi_t$,

$$|G_t(y^1, r, \xi_t) - G_t(y^2, r, \xi_t)| \leq (T - t + 1)\kappa^f_y|y^1 - y^2| \leq (T - t + 1)\kappa|y^1 - y^2|.$$  \hspace{1cm} (8)

2. For each $y$, $r^1$, $r^2$, and $\xi_t$,

$$|G_t(y, r^1, \xi_t) - G_t(y, r^2, \xi_t)| \leq ((T - t)\kappa^f_y + \kappa^f + \kappa^d) |r^1 - r^2| \leq (T - t + 1)\kappa |r^1 - r^2|.$$  \hspace{1cm} (9)
3. For each $x^1$ and $x^2 \in X$, $|v_t(x^1) - v_t(x^2)| \leq (T - t + 1)\kappa |x^1 - x^2|$.  

4. For each $(y, r)$, $G_t(y, r, \xi_t)$ takes values in a bounded interval of length at most

$$(\kappa^d_t + (T - t)\kappa^f_t)\lambda \leq (T - t + 1)\kappa\lambda.$$ 

The proof of Theorem 3 is straightforward and is omitted. See for example, Federgruen and Heching (1999). (The facts that $D_t(r, \xi_t)$ is concave in $r$ and that $v_{t-1}(x)$ is a concave decreasing function imply that $v_{t-1}(y - D_t(r, \xi_t))$ is concave in $(y, r)$.) The proof of Lemma 5 is standard and also omitted.

It follows from Theorem 3 that the following policy is optimal for our model: (i) if the inventory at the start of period $t$, $x_t$, is less than some base-stock level $y^*_t$, produce enough to bring the inventory level up to $y^*_t$ and charge $r^*_t$, (ii) if $x_t > y^*_t$, produce nothing and offer the product at a price of $r^*_t(x_t)$. (This policy is the BSLP policy in Federgruen and Heching (1999), but we do not establish here that $r^*_t(x_t)$ is decreasing in $x_t$.) Hence, to approximate the optimal policy, for each period $t$, we need to approximate $y^*_t$, together with points $\{r^*_t(x_t) \mid x_t \geq y^*_t\}$.

We assume in this section only that the inventory positions before and after ordering, $x_t$ and $y_t$, take integer values within in a bounded interval $X = [x, \bar{x}]$ for all $t$ under all policies of interest. We use $|X|$ to denote the number of integer points within this interval.

The approximation process is as follows. We will apply recursion. Starting with $t = T$, assuming that we can approximate $v_t(\cdot)$ on $X$ with a concave function $v^*_t(\cdot)$, we will replace $v_t(\cdot)$ with $v^*_t(\cdot)$ in the maximization (5) to (7). Specifically, $v^*_{t-1}(x)$ is approximately equal to

$$\max_{y_{t-1} \leq x} \max_{0 \leq r_{t-1} \leq \bar{r}} E[f_{t-1}(y_{t-1}, r_{t-1}, D_{t-1}(r_{t-1}, \xi_{t-1})) + v^*_t(y_{t-1} - D_{t-1}(r_{t-1}, \xi_{t-1}))]. \quad (10)$$

Progressing backward through time, we obtain an approximation $v^*_t(\cdot)$ for the value function $v_t(\cdot)$ in every period, and policy approximations $y^*_t$ for $y^*_t$ and $\{r^*_t(x_t) \mid x_t \geq y^*_t\}$ for $\{r^*_t(x_t) \mid x_t \geq y^*_t\}$.

The work performed by our algorithm at each period involves two parts. In Part I, for $y = \bar{x}, \bar{x} - 1, \ldots, x$, the algorithm computes, by calling the GB algorithm as a subroutine, an interval $I^*_{t-1}(y) = [\tilde{L}^*_{t-1}(y), \tilde{R}^*_{t-1}(y)]$ such that the probability of

$$\max_{0 \leq r \leq \bar{r}} E[f_{t-1}(y, r, D_{t-1}(r, \xi_{t-1})) + v^*_t(y - D_t(r, \xi_{t-1}))] \in I^*_{t-1}(y)$$

is high. The algorithm stops when it finds a downward trend in the intervals as $y$ decreases. During Part I, the algorithm also finds an approximation to the optimal policy. It returns $y^*_{t-1}$ as an approximation to the largest maximizer of the expectation in (10). For each value of $x$ between $\bar{x}$
and \( y_{t-1}^\pi \), it returns \( r_{t-1}^\pi(x) \) as the approximate maximizer output by the GB algorithm. In Part II, using the intervals \( J_{t-1}^\pi(\cdot) \) found in Part I, the algorithm forms a piecewise-linear, concave function \( v_{t-1}^\pi(x) \). We will show in Theorem 4 that this function is a good approximation to (10).

Let

\[
G_{t-1}^\pi(y, r, \xi_{t-1}) = f_{t-1}(y, r, D_{t-1}(r, \xi_{t-1})) + v_{t}^\pi(y - D_{t-1}(r, \xi_{t-1})), \quad t < T,
\]

\[
G_{T}^\pi(y, r, \xi_{T}) = G_{T}(y, r, \xi_{T}),
\]

and let

\[
V_{t}^\pi(y) = \max_{0 \leq r \leq \tilde{r}} E[G_{t}^\pi(y, r, \xi_{t})], \quad t = 1, \ldots, T.
\]

It can be shown that \( G_{t}^\pi \) has nice concavity properties:

**Lemma 6.** For each \( t \) and \( \xi_{t} \), if \( v_{t}^\pi(\cdot) \) is concave then \( G_{t}^\pi(\cdot, \cdot, \xi_{t}) \) is also concave.

Moreover, Lemma 5 implies that

**Lemma 7.** If \( |v_{t}(x) - v_{t}^\pi(x)| \leq \delta \) for all \( x \in X \) then \( G_{t-1}^\pi \) satisfies the following conditions:

1. For each \( y_{1}, y_{2} \in X, \quad r, \quad \xi_{t-1}, \)

\[
|G_{t-1}^\pi(y_{1}, r, \xi_{t-1}) - G_{t-1}^\pi(y_{2}, r, \xi_{t-1})| \leq (\kappa_{y}^{t} + 2\delta + (T - t + 1)\kappa)|y_{1} - y_{2}|
\]

\[
\leq (2\delta + (T - t + 2)\kappa)|y_{1} - y_{2}|.
\]

2. For each \( y \in X, \quad r_{1}, r_{2}, \quad \xi_{t-1}, \)

\[
|G_{t-1}^\pi(y, r_{1}, \xi_{t-1}) - G_{t-1}^\pi(y, r_{2}, \xi_{t-1})| \leq (2\delta + (T - t + 2)\kappa)|r_{1} - r_{2}|.
\]

3. For each \( x_{1}, x_{2} \in X, \)

\[
|v_{t-1}^\pi(x_{1}) - v_{t-1}^\pi(x_{2})| \leq (2\delta + (T - t + 2)\kappa)|x_{1} - x_{2}|.
\]

4. For each \( (y, r) \in X \times R, \)

\( G_{t-1}^\pi(y, r, \xi_{t-1}) \) takes values in a bounded interval of length at most

\( 2\delta + (T - t + 2)\kappa \).

**Inventory Control and Pricing Algorithm (ICP Algorithm)**

(Executed once per stage \( t \), starting with \( t = T \)).

**Input:**

- \( \delta, \ p > 0; \)
- Concave function \( v_{t}^\pi(y) \).

**Output:**
• Piecewise-linear, concave function $v^\pi_{t-1}(\cdot)$ such that with probability at least $1 - p^t$,

$$\left| v^\pi_{t-1}(x) - \max_{0 \leq r \leq \bar{r}} E \left[ f_{t-1}(y, r, D_{t-1}(r, \xi_{t-1})) + v^\pi_t(y - D_t(r, \xi_{t-1})) \right] \right| \leq \delta^t \forall x \in X.$$

• Values $y^\pi_{t-1}$ and $\{r^\pi_{t-1}(x) \mid x \geq y^\pi_{t-1}\}$ such that for all $x \in X$, if $x^\pi = \max(x, y^\pi_{t-1})$ then

$$\left| v^\pi_{t-1}(x^\pi) - E \left[ f_{t-1}(x^\pi, r^\pi_{t-1}(x^\pi), D_{t-1}(r^\pi_{t-1}(x^\pi), x^\pi_{t-1})) \right] + [v^\pi_t(x^\pi - D_t(r^\pi_{t-1}x^\pi, x^\pi_{t-1}))] \right|$$

is bounded by $\delta^t$ with probability at least $1 - p^t$, and for all $x < y^\pi_{t-1}$ and $x \in X$, $v^\pi_{t-1}(x) = v^\pi_{t-1}(y^\pi_{t-1})$.

Algorithm:

Set $k = |X|$, $\bar{k} = k$, $y^k = \bar{x}$, $\delta^{t+1} = 0$.

Part I: Finding the intervals $I^\pi_{t-1}(\cdot)$:

1. Apply the GB algorithm to maximize the function $G^\pi_{t-1}(y^k, \xi_{t-1})$ over $r$, using parameters

$$p^t/|X|, \frac{\delta^t}{2}, M = 2\delta^t + (T - t + 2)\kappa\lambda, \epsilon = \frac{\delta^t}{(T - t + 2)\kappa + 2\delta^t},$$

and search domain $[0, \bar{r}]$. Set $(r^\pi_{t-1}(y^k), y^\pi_{t-1}(y^k))$ to be the output of GB. Let $\bar{J}^\pi_{t-1}(y^k) = \frac{\pi}{2} y^\pi_{t-1}(y^k) + \frac{\delta^t}{2}$ and let $\bar{J}^\pi_{t-1}(y^k) = \frac{\bar{J}^\pi_{t-1}(y^k), y^\pi_{t-1}(y^k)}$. Then

$$P[\bar{V}^\pi_{t-1}(y^k) \in \bar{J}^\pi_{t-1}(y^k)] \geq 1 - \frac{p^t}{|X|}.$$

2. If $\bar{J}^\pi_{t-1}(y^k) > \bar{J}^\pi_{t-1}(y^k)$ then $\bar{k} \leftarrow k$.

3. If either $k = 1$ or $\bar{J}^\pi_{t-1}(y^k) \leq \bar{J}^\pi_{t-1}(y^k)$ then for all $k' \geq k$, set $y^*(k') = \arg \max \{J^\pi_{t-1}(y) \mid y \in X, y \geq y^{k'}\}$ and $I^\pi_{t-1}(y^{k'}) = I^\pi_{t-1}(y^*(k'))$. Set $K \leftarrow k$ and $y^\pi_{t-1} = y^k$. For all $y^{k'} \geq y^\pi_{t-1}$, set $r^\pi_{t-1}(y^{k'}) = r^\pi_{t-1}(y^*(k'))$. EXIT.

4. Otherwise, set $y^k \leftarrow y^k - 1, k \leftarrow k - 1$ and go to Step 1.

Part II: Finding concave function $v^\pi_{t-1}(\cdot)$:

1. Find the convex hull $C$ of $\{(y^k, I^\pi_{t-1}(y^k)) | k = 1 \ldots |X|\}$.

2. For each $k$, set $v^\pi_{t-1}(y^k) = \max \{t | (y^k, t) \in C\}$.

Finding the convex hull of the set of points in Part II is an $O(|X|\log|X|)$ procedure using the Graham scan (see Graham (1972)).

The next two lemmas build up partial results for our main theorem, Theorem 4. Lemma 8 proves that given the correct input, our approximation $v^\pi_{t-1}(\cdot)$ to $v_{t-1}(\cdot)$ is a concave function passing through all of the intervals $I^\pi_{t-1}(\cdot)$.

**Lemma 8.** If $v^\pi(\cdot)$ is concave and if $\max_{y \geq x} V^\pi_{t-1}(y) \in I^\pi_{t-1}(x)$ for all $x \in X$, then the algorithm finds a concave function $v^\pi_{t-1}(\cdot)$ such that $v^\pi_{t-1}(x) \in I^\pi_{t-1}(x)$ for all $x \in X$.

Lemma 9 proves that the approximation is sufficiently close with high certainty.
Lemma 9. If \( v_t^\pi(\cdot) \) is concave then the ICP Algorithm ensures that

\[
P \left[ \max \{ V_{t-1}(y) \mid y \in I_{t-1}(x), \forall x \in X \} \geq 1 - p_t \right] \geq 1 - p^t
\]

The following is our main theorem, proving the correctness of the algorithm for inventory and pricing control.

Theorem 4. Given arbitrary values \( \delta, \delta^t, p, p^t > 0 \), and concave function \( v_t^\pi(\cdot) \) for which

\[
P \left[ |v_t(x) - v_t^\pi(x)| < \delta \ \forall x \in X \right] \geq 1 - p,
\]

the ICP Algorithm finds a concave function \( v_{t-1}^\pi(\cdot) \) such that

\[
P \left[ |v_{t-1}(x) - v_{t-1}^\pi(x)| < \delta^t + \delta^t \ \forall x \in X \right] \geq 1 - p - p^t,
\]

using at most

\[
O \left( \frac{|X|\lambda^2 T^2 \kappa^2}{(\delta^t)^2} \log \left( \frac{\bar{r} T \kappa}{\delta^t} \right) \log \left( \frac{\bar{r} |X| T \kappa}{p^t \delta^t} \right) \right) \tag{13}
\]
samples.

Starting in period \( t = T \) and working backward to period 1, for each \( t \), we run one iteration of the ICP Algorithm with input \( \delta^t = \frac{\delta}{T} \) and \( p^t = \frac{p}{T} \). Each time, we use the approximation \( v_t^\pi(\cdot) \) as input to the next iteration \( t - 1 \). Theorem 4 guarantees that at time \( t \), we find \( v_t^\pi(\cdot) \) such that

\[
P \left[ |v_t(x) - v_t^\pi(x)| < \delta \ \forall x \in X \right] \geq 1 - \frac{p(T - t + 1)}{T}.
\]

The complexity of the ICP Algorithm, which is obtained by summing up the sample bound in (13) over all iterations \( t \), is summarized in

Theorem 5. Given \( \delta > 0 \) and \( 0 < p < 1 \) and using \( \delta^t = \frac{\delta}{T} \) and \( p^t = \frac{p}{T} \) at each iteration, the ICP Algorithm computes a \((\delta,p)\)-PAC solution to the dynamic program (1) using

\[
O \left( \frac{|X| T^5 \kappa^2 \lambda^2}{\delta^2} \log \left( \frac{\bar{r} T \kappa}{\delta} \right) \log \left( \frac{\bar{r} T \kappa |X|}{p \delta} \right) \right) \tag{14}
\]
samples.

The ICP Algorithm can be easily adapted to work for problems in which there are constraints on the replenishment quantity in each period. For example, there may be capacity constraints or constraints on the minimum order quantity. In the presence of these or other constraints, in each period \( t \) we can restrict our search to the points in the space \( X \) that correspond to valid order up to levels \( y_t \). We must ensure that all concavity conditions continue to hold in these more
general settings. It can be shown easily that if the demand $D_t$ is a linear function of the price in each period $t$ then under minimum order quantity and capacity constraints, all concavity results indeed continue to hold. Since the value function may not be decreasing in the order-up-to level $y_t$, several aspects of the algorithm, including the termination criterion, will need to change. Some of the proofs will also change. However, the algorithmic approach and the general proofs are still applicable.

6. Finding an Upper or Lower Maximizer: Asymmetric Gradient Bound

In this section, we investigate asymmetric variant of the GB algorithm with an eye toward solving the pure inventory-control problem more efficiently than we have been able to when the decision space includes prices as well. Recall that in Section 5, we approximate the multi-period problem by building an approximation to the entire function $v^*_\pi(\cdot)$ in each period $t$. When the price is fixed, we can also apply the same procedure. This method could involve up to $|X|$ invocations of the GB algorithm per period. In Section 7, we will develop an alternative approximation method, one that makes at most 2 calls to GB per period. The method requires finding an approximate lower maximizer and an approximate upper maximizer to a convex function, which are respectively smaller and larger than the true maximizer(s) with high probability. These asymmetric maximizers are needed to maintain certain convexity properties.

We will show how to modify the GB algorithm to compute approximate lower and upper maximizers. We focus first on lower maximizers in the Lower Gradient Bound (LGB) algorithm. By symmetry, we can extend the ideas to specify an Upper Gradient Bound (UGB) algorithm to find an approximate upper maximizer to a function.

The following LGB subroutine is used to our LGB algorithm in the same way that the GB subroutine was used in the GB algorithm.

**Lower Gradient Bound (LGB) Subroutine:**

**Input:** Same as for the GB Subroutine.

**Output:**

1. An interval $[a', b'] \subset [a, b]$ such that $b' - a' \leq (b - a)\phi$;

2. $N(\delta, \frac{\xi}{2})$ independent samples of $g(x'_i, \xi)$ for $i = 1$ or $2$, where $x'_1$ is such that $x'_1 - a' = (b' - a')\tilde{\phi}$ and $x'_2$ is such that $b' - x'_2 = (b' - a')\tilde{\phi}$;

3. One of the following two results
   - Guarantee that $P[\arg\max_{[a, b]} f(x) \in [a', b']] \geq 1 - p.$
• Point \((x_i, m_i) \in \{a', b']\), \(i = 1, 2\), such that \(x_i \leq x\) for all \(x \in [a, b] \setminus \{a', b']\), and with probability at least \(1 - p\), \[\max_{[a, b] \setminus \{a', b']}(x) \leq m_i + \frac{2\delta}{2\delta - 1}\delta\), and \(m_i \leq f(x_i) \leq m_i + \delta\).

Subroutine:

1. If samples at \(x_1\) are given, draw \(N(\delta, \frac{\delta}{2}) - N(\delta, p)\) additional independent samples of \(g(x_1, \xi)\) and \(N(\delta, \frac{\delta}{2})\) random independent samples of \(g(x_2, \xi)\). Similarly, if samples at \(x_2\) are given then obtain \(N(\delta, \frac{\delta}{2})\) independent samples of \(g(x_i, \xi)\) for \(i = 1, 2\) in the same way. Let \(I_i = \{m_i, m_i + \delta\}\) denote the interval of length \(\delta\) centered at the sample mean \(y_i\) of the \(N(\delta, \frac{\delta}{2})\) samples of \(g(x_i, \xi)\).

2. If \(m_1 > m_2 + \delta\) then return the interval \([a, x_2]\) and the samples of \(g(x_1, \xi)\).

3. Else if \(m_2 > m_1 + \delta\) then return the interval \([x_1, b]\) and the samples of \(g(x_2, \xi)\).

4. Else return the interval \([a, x_2]\) and point \((x_1, m_1)\).

Lemmas 1 and 2 still hold for LGB. Lemmas 3 and 4 are replaced by the following result:

**Lemma 10.** If \(m_1 - \delta \leq m_2 \leq m_1 + \delta\) then \(P\left[\max_{x \in [x_2, b]} f(x) \leq m_1 + \frac{2\delta}{2\delta - 1}\delta\right] \geq 1 - p\).

The proof of the following theorem is analogous to that for Theorem 1.

**Theorem 6.** The LGB subroutine returns the correct output using \(O\left(\frac{1}{\epsilon^2} \log\left(\frac{1}{p}\right)\right)\) samples.

The following LGB algorithm is based on the LGB subroutine.

**Lower Gradient Bound Algorithm:**

**Input:** Same as for the GB Algorithm.

**Output:** A point \((\hat{x}, \hat{m})\) such that \(P\left[|f(\hat{x}) - \max_x f(x)| \leq \delta\right.\) and \(\hat{x} \leq \arg\max_x f(x)\] \(\geq 1 - p\).

**Algorithm:**
Initially, \(k = 1\), \(a^1 = a\), \(b^1 = b\), \(S = \emptyset\), \(\hat{\delta} = \frac{2\epsilon - 1}{4\epsilon - 1}\delta\), \(z^1 = a^1 + (b^1 - a^1)\hat{\phi}\), \((\hat{x}, \hat{l}) = (z^1, -\infty)\), \(p^1 = \frac{\delta}{2}\). Draw \(N(\hat{\delta}, \frac{\hat{\delta}}{2})\) independent samples of \(g(z^1, \xi)\).

Begin loop

1. If \(b^k - a^k < \epsilon\) then let \(x^k = a^k\). Draw \(N(\hat{\delta}, \frac{\hat{\delta}}{2})\) independent samples of \(g(x^k, \xi)\). Locate an interval of length \(\hat{\delta}\) centered at the sample mean. Let \(m^k\) be the lower end point of that interval. Let \(S \leftarrow S \cup \{(x^k, m^k)\}\). Exit loop.

2. Otherwise, run the subroutine with input \([a^k, b^k]\), \(p^k = \frac{\hat{\delta}}{2}\), \(\hat{\delta}\), and the samples of \(g(z^k, \xi)\). If the subroutine returns a point \((x^k, m^k)\) then let \(S \leftarrow S \cup \{(x^k, m^k)\}\).

3. Let \(k \leftarrow k + 1\), \([a^k, b^k] \leftarrow [a', b']\). Let \(z^k\) be the point at which samples were returned by the GB Subroutine at the end of Step 2. Go to Step 1.

End loop.

Elimination round: Consider the list \(S\).

1. Let \(\bar{m} = \max\{m^k \mid (x^k, m^k) \in S\}\). Remove from \(S\) all points \((x^k, m^k)\) such that \(m^k < \bar{m} - \hat{\delta}\).
2. Return the element $(\hat{x}, \hat{m}) \in S$, where $\hat{x} = \min \{x^k \mid (x^k, m^k) \in S\}$.

**Theorem 7.** The LGB algorithm returns the desired output using $O\left(\frac{M^2}{\delta^2} \log\left(\frac{L}{\epsilon}\right) \log\left(\frac{L}{p\epsilon}\right)\right)$ samples.

7. The Inventory Control Problem

In this section, we develop a fast algorithm, called the IC Algorithm, for inventory management problems. These are problems in which the price is fixed throughout the horizon and decisions are made only to replenish inventory. Recall that in Section 5, we approximate the multi-period problem by building an approximation to the entire value function $v^*_\pi(\cdot)$ in each period $t$, a process that involves up to $|X|$ invocations of the GB algorithm per period. This approximation process is necessitated by the complex structure of the optimal policy in which each (of up to $|X|$) inventory positions selected in a period has a corresponding optimal price associated with it. Here, we will develop an alternative approximation method that makes at most 2 calls to GB per period, making use of the fact that the optimal policy is much simpler in this case. We will restate our formulation with some generalizations and comment on structural properties that continue to hold under the generalizations. Next, we will outline our solution approach. Finally, we analyze the IC Algorithm and its cost.

We make several changes to the basic formulation (1). First, inventory levels $x_t$ and $y_t$ are now continuous rather than discrete quantities. Thus, we allow $X$ to be an interval on the real line and use $|X|$ to denote the length of that interval. We suppress the price as an argument throughout to reflect the fact that price is no longer a decision variable. In particular, the demand in $t$ is written simply as $D_t(\xi_t)$. We will generalize the period $t$ constraint $y_t \geq x_t$ to a more general constraint, namely $(x, y) \in A_t$, where $A_t$ is any closed convex set in $\mathbb{R}^2$ satisfying the following property:

**Assumption 3.** There exists an $\alpha > 0$ such that for each $t$, $(x, y) \in A_t$, and $x' \in X$, there is $y'$ such that $(x', y') \in A_t$ and $|y - y'| \leq \alpha|x - x'|$.

Assumption 3 is satisfied whenever the upper (concave) and lower (convex) envelopes of $A_t$ have left and right derivatives uniformly bounded by $\alpha$. It is satisfied by the basic (unconstrained) inventory problem with $\alpha = 1$. Finally, it is satisfied by the inventory problem with capacity and minimum order quantity constraints with $\alpha = 1$. We generalize $\kappa$ of Section 3 to $\kappa = (\kappa_{d}^{\prime} + \kappa_{r}^{\prime}) \max(1, \alpha) + \kappa_{l}^{\prime} \max(1, \kappa_{d}^{\prime})$. In addition, we generalize the period $t$ state transition from $x_{t+1} = y_t - D_t(\xi_t)$ to $x_{t+1} = g_t(y_t, \xi_t)$, where $g_t(y_t, \xi_t)$ is any linear function of $y_t$ for each $\xi_t$. This generalization leads us to make the following assumption, in addition to Assumption 1:

**Assumption 4.** For every $t$ and for every fixed $y$, $g_t(y, \xi_t)$ takes values in a bounded interval of length $\lambda$. For every $t$, $\xi_t$, and $y^1$, $y^2$, $|g_t(y^1, \xi_t) - g_t(y^2, \xi_t)| \leq \kappa^{\prime \prime}|y^1 - y^2|$.
Note that this definition of $\lambda$ is slightly different from the definition used in Sections 3 and 5. These generalizations allow us to simultaneously treat multiple classes of problems. For example, we can capture inventory problems involving limited capacity and minimum purchase quantities. As another example, if a random fraction of spoilage, shrinkage, or growth (as in the production of yeast) takes place in period $s$, then we can model the state transition as $x_{s+1} = y_s \theta_s(\xi_s) - D_s(\xi_s)$ for some positive function $\theta_s(\cdot)$.

In light of the changes, our formulation takes on the following form:

$$v_t(x_t) = \max_{(x_t, y_t) \in A_t} E \left[ f_t(y_t, D_t(\xi_t)) + v_{t+1}(g_t(y_t, \xi_t)) \right]$$

$$v_T(x_T) = \max_{(x_T, y_T) \in A_T} E \left[ f_T(y_T, D_T(\xi_T)) \right].$$

Recall that, as we defined earlier,

$$V_t(y) = E \left[ f_t(y_t, D_t(\xi_t)) + v_{t+1}(g_t(y_t, \xi_t)) \right], \quad t = 1, \ldots, T - 1,$$

$$V_T(y) = E \left[ f_T(y_T, D_T(\xi_T)) \right].$$

Define the optimal unconstrained state in each period $t$ to be $y_t^*$, where $y_t^* = \arg\max_y V_t(y)$. The optimal policy has a simple structure when the problem no longer includes pricing decisions. It is based on knowledge of a single quantity $y_t^*$ for each period $t$, and of the constraint set $A_t$. By the concavity of $V_t(\cdot)$ (Theorem 8 below), it is easy to see that given $x_t$, the optimal decision in period $t$ has the form

$$y_t^*(x_t) := \arg\min_{(x_t, y_t) \in A_t} |y_t^* - y|,$$

which can be computed easily given $y_t^*$. We refer to the state-dependent decision $y_t^*(x_t)$ as the decision approaching $y_t^*$. 1.

Our approach is based on the following insight, which was also exploited by Levi et al. (2006). Given that the set $\{y_s^*\}_{t+1}^T$ is known, the policy that chooses decision $y_t$ in $t$ and decisions approaching $y_s^*$ in all periods $s \in (t, T]$ incurs exactly the profit $V_t(y_t)$ in expectation, where the expectation is computed over all random paths $\{\xi_t\}_{t}^T$. Indeed, the profit of this policy is

$$c_t x_t + F_t(y_t, \xi_t, \ldots, \xi_T) = f_t(\tilde{y}_t, D_t(\xi_t))$$

$$+ f_{t+1}(\tilde{y}_{t+1}, D_{t+1}(\xi_{t+1})) + \ldots + f_T(\tilde{y}_T, D_T(\xi_T))$$

where

$$\tilde{x}_s = g_{s-1}((\tilde{y}_{s-1}, \xi_{s-1}), s > t,$$
\[ \tilde{y}_s = \begin{cases} y_s, & s = t; \\ \tilde{y}_s(x_s), & s > t. \end{cases} \]

Hence, a knowledge of the optimal unconstrained states \( \{y^*_s\}_{s \in (t,T]} \) and of the sample paths \( \{\xi_s\}_t \) provides a fast way to estimate the function \( V_t = E[F_t(\cdot, \xi_t, \ldots, \xi_T)] \) without knowing the value function \( v_{t-1}(\cdot) \). Note that we are not observing the samples \( \{\xi_s\}_t \) directly; instead, we observe \( F_t(\cdot, \xi_t, \ldots, \xi_T) \).

In general, we will not be able to compute the sequence \( \{y^*_s\}_{s \in (t,T]} \) exactly. However, if we know upper and lower bounds \( (\underline{y}^*_s, \overline{y}^*_s) \) on each \( y^*_s \), then we can implement a policy that approaches \( y^*_s \) approximately.

This insight suggests the following approach. We will work backward from the end of the horizon. At each stage \( t \), given a random sample \( \xi_t, \ldots, \xi_T \) and a sequence \( \{(\underline{y}^*_s, \overline{y}^*_s)\}_{s \in (t,T]} \) of lower and upper bounds, we can compute a sample \( F^n_t(y_t, \xi_t, \ldots, \xi_T) \) of the profit achieved by the policy that chooses decision \( y_t \) in \( t \) and decisions approaching the bounds \( (\underline{y}^*_s, \overline{y}^*_s) \) in all periods \( s \in (t,T] \). (For a formal definition of \( F^n_t \), see (18) below.) Hence, we will run the LGB and UGB algorithms on the function \( F^n_t(y_t, \xi_t, \ldots, \xi_T) \) to compute bounds \( (\underline{y}^*_t, \overline{y}^*_t) \) on \( \max_y E[F^n_t(y, \xi_t, \ldots, \xi_T)] \). This concludes the work for stage \( t \), at which point we move to \( t-1 \).

In order to establish the validity of this approach, we need to ensure that at each stage, if the pair of bounds \( (\underline{y}^*_s, \overline{y}^*_s) \) are good lower and upper maximizers to \( V^n_t(y_t) = E[F^n_t(y, \xi_t, \ldots, \xi_T)] \) for each \( s > t \), then the same holds for \( s = t \). Further, we will ensure that \( V^n_t(\cdot) = E[F^n_t(\cdot, \xi_t, \ldots, \xi_T)] \) is a good approximation to the true function \( V_t(\cdot) \).

Next, we will make precise the approach described above. Let

\[ [l_t(x), u_t(x)] = A_t(x) = \{y \mid (x,y) \in A_t\}. \]

Because \( A_t \) is convex and closed, \( l_t(x) \) and \( u_t(x) \) are continuous, \( l_t(x) \) is convex, and \( u_t(x) \) is concave. Define the sequence of functions \( \{F^n_s\}_{s = t}^T \) recursively as follows. Let \( F^n_t(\cdot, \xi_T) = f_T(\cdot, D_T(\xi_T)) \). Suppose that \( \{F^n_s\}_{s = t}^T \) is known. Suppose that \( (\underline{y}^*_s, \overline{y}^*_s) \) is a pair of lower and upper maximizers of \( V^n_t \). Then

\[
F^n_{t-1}(y_{t-1}, \xi_{t-1}, \ldots, \xi_T) = f_{t-1}(y_{t-1}, D_{t-1}(\xi_{t-1})) + w^n_t(g_{t-1}(y_{t-1}, \xi_{t-1}), \xi_t, \ldots, \xi_T),
\]

where \( w^n_t(x, \xi_t, \ldots, \xi_T) \) is defined below, according to 4 cases, depending on where \( x \) falls with respect to the following regions:

\[ R^n_t = \{x \in \mathbb{R} \mid l_t(x) > \overline{y}^*_t\}, \]
\[ S_i^\pi = \{ x \in \mathbb{R} | u_i(x) \geq y_i^\pi \}. \]

Note that \( \mathcal{R}_i^\pi \) is the set of states \( x_i \in X \) at which a policy is forced to adopt decisions \( y_i \) that are too large, i.e., larger than \( y_i^\pi \). On the other hand, the complement \( -S_i^\pi \) of \( S_i^\pi \) contains all states \( x_i \) at which a policy is forced to adopt decisions \( y_i \) that are too small, i.e., smaller than \( y_i^\pi \). Note that \( \mathcal{R}_i^\pi \subset S_i^\pi \). Because \( l_i(x) \) is convex and \( u_i(x) \) is concave, \( S_i^\pi \), the complement \( -\mathcal{R}_i^\pi \) of \( \mathcal{R}_i^\pi \), and their (non-empty) intersection \( S_i^\pi \cap \mathcal{R}_i^\pi \), are all closed intervals. We define \( x_i^- \) and \( x_i^+ \) by \( S_i^\pi \setminus \mathcal{R}_i^\pi = [x_i^-, x_i^+] \). Consequently, \( -(S_i^\pi \setminus \mathcal{R}_i^\pi) = \mathcal{R}_i^\pi \cup -S_i^\pi \) is the union of two (possibly empty) intervals, namely \((-\infty, x_i^-)\) and \((x_i^+, \infty)\). We assume that \( x_i^- \leq x_i^+ \) and we allow \( x_i^- = -\infty \) and \( x_i^+ = \infty \).

If \( x_i^+ \) is finite, we define \( y_i^+ = [l_i(x_i^+), u_i(x_i^+)] \cap [y_i^\pi, y_i^\pi] \). Since \( l_i(\cdot) \) and \( u_i(\cdot) \) are continuous, the value \( y_i^+ \) is well-defined. (Note that both \( y_i^+ = y_i^\pi \) and \( y_i^+ = y_i^\pi \) are possible.) If \( x_i^- \) is finite, we define \( y_i^- \) similarly. Note that \( \{y_i^+, y_i^-\} \subset \{y_i^\pi, y_i^\pi\} \). For all \( x \), define

\[
\begin{align*}
&\begin{cases}
    w_i^\pi(x, \xi_1, \ldots, \xi_T) \\
    F_i^\pi(u_i(x), \xi_1, \ldots, \xi_T),
  \end{cases} & \text{if } x \in -S_i^\pi, \\
&\begin{cases}
    F_i^\pi(l_i(x), \xi_1, \ldots, \xi_T),
  \end{cases} & \text{if } x \in \mathcal{R}_i^\pi, \\
&\begin{cases}
    F_i^\pi\left(\frac{x - x_i^-}{x_i^- - x_i^+} y_i^+, \frac{x_i^+ - x}{x_i^+ - x_i^-} y_i^-, \xi_1, \ldots, \xi_T\right),
  \end{cases} & \text{if } x \in S_i^\pi \setminus \mathcal{R}_i^\pi, \ x_i^- < \infty, \ x_i^- > -\infty, \\
&\begin{cases}
    F_i^\pi(y_i^+, \xi_1, \ldots, \xi_T),
  \end{cases} & \text{if } x \in S_i^\pi \setminus \mathcal{R}_i^\pi, \ x_i^+ < \infty, \ x_i^- = -\infty, \\
&\begin{cases}
    F_i^\pi(y_i^-, \xi_1, \ldots, \xi_T),
  \end{cases} & \text{if } x \in S_i^\pi \setminus \mathcal{R}_i^\pi, \ x_i^+ = \infty, \ x_i^- > -\infty, \\
&\begin{cases}
    F_i^\pi(y_i^-, \xi_1, \ldots, \xi_T),
  \end{cases} & \text{otherwise.}
\end{align*}
\]

The decisions \( y_i^\pi(x) \) taken at time \( t \) by the policy \( \pi \), given the initial state \( x \), according to the above cases, are \( u_i(x), l_i(x), \frac{x - x_i^-}{x_i^- - x_i^+} y_i^+, \frac{x_i^+ - x}{x_i^+ - x_i^-} y_i^-, y_i^+, y_i^-, \) and \( y_i^\pi \), respectively. In (19), we have made precise the notion of approaching the bounds \( (y_i^\pi, y_i^\pi) \) with the decision at time \( t \). The quantity \( c(x_i) + F_i^\pi(y_i, \xi_1, \ldots, \xi_T) \) is the profit along a sample path of a policy that starts at \( x_i \) in period \( t \), makes decision \( y_i \) in \( t \), and makes decisions approaching \( (y_i^\pi, y_i^\pi) \), in the sense defined in (19), in all periods \( s > t \).

Define \( \beta = \alpha \kappa^\beta \), where \( \alpha \) is given by Assumption 3. Let

\[
\sigma_1(\beta, t) = \begin{cases} 
1, & \text{if } \beta < 1, \\
t, & \text{if } \beta = 1, \\
\beta^t, & \text{if } \beta > 1
\end{cases}
\]

and

\[
\sigma_2(\beta, t) = \begin{cases} 
1, & \text{if } \beta < 1, \\
t^2, & \text{if } \beta = 1, \\
\beta^t, & \text{if } \beta > 1.
\end{cases}
\]

Define \( \kappa_{t-1,y}^F \) to be the Lipschitz constant for \( F_{t-1}^\pi \). That is, \( \kappa_{t-1,y}^F \) satisfies

\[
|E[F_{t-1}^\pi(y_1^1, \xi_{t-1}, \ldots, \xi_T)] - E[F_{t-1}^\pi(y_2^1, \xi_{t-1}, \ldots, \xi_T)]| \leq \kappa_{t-1,y}^F |y_1^1 - y_2^1|
\]
for all \( y^1 \) and \( y^2 \). Similarly, define \( \kappa_{i-1,x}^w \) to be the Lipschitz constant for \( w_i^x \), so that

\[
|E[w_i^x(x^1, \xi_t, \ldots, \xi_T)] - E[w_i^x(x^2, \xi_t, \ldots, \xi_T)]| \leq \kappa_{i,x}^w |x^1 - x^2|
\]

for all \( x^1 \) and \( x^2 \). We can prove that

**Lemma 11.** For all \( t \), the following properties hold:

1. \( \kappa_{t-1,y}^F = \kappa_y^f \sum_{k=0}^{T-t+1} \beta^k = \kappa O(\sigma_1(\beta, T - t + 1)) \).
2. \( \kappa_{t,x}^w = \alpha \kappa_y^f \sum_{k=0}^{T-t} \beta^k = \kappa O(\sigma_1(\beta, T - t + 1)) \).
3. For each \( y \), with probability 1, \( F_i^y(\xi_t, \xi_T) \) takes values in a bounded interval of length \( \Lambda_{t-1}^F = (T - t + 2) \kappa_y^f + \lambda \sum_{s=t}^{T} \kappa_{s,x}^w = \lambda \kappa O(\sigma_2(\beta, T - t + 2)) \).

Let \( v_i^x(x) = E[w_i^x(x, \xi_t, \ldots, \xi_T)] \). Note that this definition of \( v_i^x \) differs from the one used in Section 5. The following theorem shows that our approach maintains concavity for all functions \( V_i^x(\cdot) \) and \( v_i^x(\cdot) \).

**Theorem 8.** If \( V_i^x(\cdot) = E[F_i^x(x, \xi_t, \ldots, \xi_T)] \) is concave and \( \underline{y}_i^x \leq y_i^x = \arg \max_y V_i^x(y) \leq \overline{y}_i^x \) then \( v_i^x(\cdot) = E[w_i^x(\cdot, \xi_t, \ldots, \xi_T)] \) and \( V_i^x(\cdot) \) are well-defined, concave functions.

The following lemma shows that \( v_i^x(x) \), the expected profit incurred by approaching bounds \( \{(y_s, \overline{y}_s)\} \) on all \( s \), \( s \geq t \), differs from \( \max_{(x,y) \in A_t} V_i^x(y) \) by at most the error induced by approximating \( \arg \max_{(x,y) \in A_t} V_i^x(y) \) with \( \overline{y}_i^x \) or \( \underline{y}_i^x \).

**Lemma 12.** If \( V_i^x(\cdot) \) is concave and \( \underline{y}_i^x \leq y_i^x \leq \overline{y}_i^x \) then for every \( x \),

\[
\max_{(x,y) \in A_t} V_i^x(y) - v_i^x(x) \leq \max_{(x,y) \in A_t} V_i^x(y) - \min\{V_i^x(\underline{y}_i^x), V_i^x(\overline{y}_i^x)\}.
\]

**Proof** We showed in the proof of Theorem 8 that \( v_i^x(x) \) is identical to the concave function \( \max_{(x,y) \in A_t} V_i^x(y) \) everywhere on \( R_i^x \cup (-S_i^x) \). It suffices to check the lemma for \( x \in S_i^x \setminus R_i^x \). By definition, \( v_i^x(x) \) is equal to a convex combination of \( V_i^x(y_i^x) \) and \( V_i^x(\overline{y}_i^x) \) here.

We now show that as we progress backward from stage \( t = T \) to stage \( t = 1 \), we incur at most an additive loss in the accuracy and certainty with which we can estimate \( V_i(\cdot) \).

**Lemma 13.**

\[
P(V_i^{x+1}(\cdot) \text{ is concave and } V_{i+1}(y) - V_{i+1}^x(y) \leq \delta_2 \text{ for all } y) \geq 1 - p_2 \tag{20}
\]

and

\[
P(\max_y V_i^{x+1}(y) - \min\{V_i^{x+1}(\underline{y}_{i+1}), V_i^{x+1}(\overline{y}_{i+1})\} \leq \delta_1 \text{ and } \underline{y}_{i+1} \leq \overline{y}_{i+1} \leq \overline{y}_{i+1}) \geq 1 - p_1 \tag{21}
\]

together imply that

\[
P(V_i^x(\cdot) \text{ is concave and } V_i(y) - V_i^x(y) \leq \delta_1 + \delta_2 \text{ for all } y) \geq 1 - p_1 - p_2.
\]
Algorithm for Inventory Control using Asymmetric Gradient Bound (IC Algorithm)

**Input:**
- $\delta > 0$, $\epsilon', p \in [0,1]$;
- Bounds $\{(y^\pi_t, \bar{y}^\pi_t)\}_{t+1}^T$ specifying $V^\pi_t(\cdot)$ and $v^\pi_t(\cdot)$ via (18) and (19), where

$$V^\pi_t(\cdot) = E[F^\pi_t(y, \xi_t, \ldots, \xi_T)],$$

and an interval $[a^t, b^t]$ with the following properties:
1. $|y_1 - y_2| < \epsilon^t$ implies $|V_t(y_1) - V_t(y_2)| < \frac{\delta}{T}$ for all $y_1, y_2$;
2. With probability at least $1 - \frac{(T-t)}{pT}$,

$$V^\pi_t(\cdot) \text{ is concave and } V_t(y) - V^\pi_t(y) \leq \frac{(T-t)\delta}{T} \text{ for all } y;$$
3. With probability 1, $V^\pi_t(\cdot)$ has a maximizer $y^\pi_t$ on the bounded interval $[a^t, b^t]$, with $a^t - b^t \leq |X|$.

**Output:**
- $(y^\pi_t, \bar{y}^\pi_t)$ such that with probability at least $1 - \frac{(T-t+1)}{pT}$,

$$V^\pi_{t-1}(\cdot) \text{ is concave and } V_{t-1}(y) - V^\pi_{t-1}(y) \leq \frac{(T-t+1)\delta}{T} \text{ for all } y.$$

**Algorithm:**
- Run UGB and LGB on function $F^\pi_t(\cdot, \xi_t, \ldots, \xi_T)$ with inputs $p^t = \frac{p}{2T}$, $M_t = \Lambda^F_{t,y}$, $\epsilon^t = \frac{\delta}{T\kappa^F_{t,y}}$, $\delta^t = \frac{\epsilon}{T}$, $[a^t, b^t]$. (Formulas for $\kappa^F_{t,y}$ and $\Lambda^F_{t,y}$ are given in Lemma 7.)

**Theorem 9.** The IC Algorithm returns the correct output by taking

$$O\left(T^3\sigma_2(\beta, T)^2\kappa^2\lambda^2 \log(T|X|\kappa\sigma_1(\beta, T)\delta) \log\left(T|X|\kappa\sigma_1(\beta, T)\delta\right)p\delta\right)$$

samples at iteration $t$.

**Proof.** UGB and LGB return $y^\pi_t$ and $\bar{y}^\pi_t$ respectively, where

$$P\left(y^\pi_t \leq y^\pi_t \leq \bar{y}^\pi_t \text{ and max}_y V^\pi_t(y) - \min\{V^\pi_t(y^\pi_t), V^\pi_t(\bar{y}^\pi_t)\} \leq \frac{\delta}{T}\right) \geq 1 - \frac{p}{2T}.$$ 

Note that the IC Algorithm takes samples of $V^\pi_{t+1}(\cdot)$, and that each sample of $V^\pi_t(\cdot)$ requires $T-t+1$ samples of $\xi_t$. The total number of samples required is

$$O\left(T^3\sigma_2(\beta, T)^2\kappa^2\lambda^2 \log(T|X|\kappa\sigma_1(\beta, T)\delta) \log\left(T|X|\kappa\sigma_1(\beta, T)\delta\right)p\delta\right).$$

We apply Lemma 5 and Theorem 7.

By Lemma 12, the conclusion of the Theorem follows. □
Finally, we check that initial conditions apply, so that the IC Algorithm can be applied at \( t = T \), and count the number of samples required for computation at all stages.

**Theorem 10.** Let \( \delta > 0 \), \( p \in (0,1) \). Assume that for each \( t \) there is a known interval \([a_t, b_t] \) of length at most \( |X| \), where \( y_t^* \) lies. Then a policy \( \pi \) can be computed, such that

\[
P(v_t(x) - v_t^*(x) < \delta \text{ for } t = 1, \ldots, T) > 1 - \phi
\]

using

\[
O \left( \frac{T^4 \sigma_2(\beta, T)^2 \kappa^2 \lambda^2}{\delta^2} \log \left( \frac{|X| \sigma_1(\beta, T) \kappa}{\delta} \right) \log \left( \frac{T|X| \sigma_1(\beta, T) \kappa}{p \delta} \right) \right)
\]

samples.

**Proof.** We can conclude from Lemma 11 that for \(|V_t^*(y_1) - V_t^*(y_2)| \leq \frac{\delta}{T} \), we need \(|y_1 - y_2| \leq \epsilon_t = \frac{\delta}{|V_{t+1}^*(\cdot)|} \). Note that for \( t = T \), \( V_T^* (\cdot) = V_T (\cdot) \) meets the conditions for input into the IC Algorithm. (In (18), \( w_{T+1}^*(\cdot, \cdot) \equiv 0 \).) Hence, Theorem 9 applies for \( t = T \). Assume that the input conditions of the theorem apply at \( t + 1 \). Then the theorem and the algorithm guarantee that the conditions are satisfied for \( t \). When the dynamic program has been approximately solved for \( t = 1 \), we have obtained that with probability at least \( 1 - \frac{(T-t+1)}{pT} \),

\[
V_{t-1}^*(\cdot) \text{ is concave and } V_{t-1}(y) - V_{t-1}(y) \leq \frac{(T-t+1)\delta}{T} \text{ for all } y
\]

for all \( t = 1, \ldots, T \), which implies the desired conclusion. The sample bound for \( t = 1 \), summed over all \( t \), is as claimed. \( \square \)

Note that when \( \beta = 1 \), as in the inventory problem with capacity and minimum order quantity constraints, the sample bound in Theorem 10 above reduces to

\[
O \left( \frac{T^5 \kappa^2 \lambda^2}{\delta^2} \log \left( \frac{|X| T \kappa}{\delta} \right) \log \left( \frac{|X| T \kappa}{p \delta} \right) \right).
\]

**8. Conclusion**

We have established a non-parametric approach to the joint inventory and pricing problem in the presence of stochastic demands whose relationship to the price is unknown. We apply a new stochastic maximization algorithm, the Gradient Bound Algorithm, to compute near-optimal prices and inventory ordering quantities directly from (random) price-demand samples. We show that a polynomial number

\[
O \left( \frac{|X| T^5 \kappa^2 \lambda^2}{\delta^2} \log \left( \frac{\bar{r}T \kappa}{\delta} \right) \log \left( \frac{\bar{r}T \kappa |X|}{p} \right) \right)
\]
of samples suffices to compute a \((\delta, p)\)-PAC policy using the ICP Algorithm. When capacity and minimum-purchase-quantity constraints are present, our approximation approach applies under the stronger assumption that the demand is a linear function of the price.

When the price is not a decision variable and the number of inventory states \(|X|\) is much larger than the horizon length \(T\), we exploit the simple structure of the optimal policy to derive an alternative algorithm, the IC Algorithm, whose sample bound does not depend on \(|X|\). The algorithm uses an asymmetric version of the Gradient Bound algorithm and requires

\[
O \left( \frac{T^8 \kappa^2 \lambda^2}{\delta^2} \log \left( \frac{|X| T \kappa}{\delta} \right) \log \left( \frac{|X| T \kappa}{p \delta} \right) \right)
\]

samples in total. Levi et al. (2006) most recently proved a PAC bound of

\[
O \left( \frac{T^6}{\delta^2} \left( \frac{b+h}{\min(b, h)} \right)^2 \log \left( \frac{T}{p} \right) \right)
\]

for the pure inventory problem only, where \(b\) and \(h\) denote the unit backorder and holding cost. (The number of samples that they require per period is

\[
O \left( \frac{T^5}{\delta^2} \left( \frac{b+h}{\min(b, h)} \right)^2 \log \left( \frac{T}{p} \right) \right)
\]

Note that their work uses samples of subgradients of the value function, applies to a smaller class of models, and the error bound they obtain is a fraction \(\delta\) the optimal cost, rather than an absolute quantity. However, some similarities are worthy of notice. The term \(\delta^2\) appear in the denominator of both bounds because of a common use of the Hoeffding inequality. In the model of Levi et al. (2006), for fixed \(\min(b, h)\), as \(b+h\) grows larger, so does the ratio \(\frac{b+h}{\min(b, h)}\). This growth is matched by the growth in the Lipschitz constant \(\kappa\) in our problem.

The term \(T^8\) in our bound for the IC algorithm accounts for several sources of complexity. We perform \(T\) approximation rounds, one for each period, in order to derive an approximate policy for that period; this work contributes one factor of \(T\) to the final bound. For each period, we draw samples of demand paths, each path consisting of \(O(T)\) demand samples; hence, we have another contribution of \(T\). The approximation error in each period is \(\delta^t = \frac{\delta}{T}\); since \(\delta^t\) appears as a square in the Hoeffding bound, the accuracy with which we must approximate the optimal policy at each period contributes an additional factor of \(T^2\). Lastly, the variation of \(F^t(y, \xi_t, \ldots, \xi_T)\), which is measured by the length of the interval within which \(F^t(y, \xi_t, \ldots, \xi_T)\) falls for each fixed \(y\), is \(O(T^2)\) (Recall that \(F^t_\pi\) is the function whose expectation we must maximize at each period); this term is squared in the Hoeffding bound, leading to a final contribution of \(T^4\) to the final bound. This last contribution is an important place where we differ from Levi et al. (2006). Since Levi et al. (2006)
work with gradients of the value functions, whose variations are \( O(T) \), rather than with the value functions themselves, whose variations are \( O(T^2) \), they are able to obtain a bound of \( T^6 \) rather than our \( T^8 \).

Alternatively, for the pure inventory problem we can apply the ICP algorithm if the inventory quantities are known to be discrete. In Part I, Step 1 of the ICP Algorithm, instead of applying the GB Algorithm to maximize \( G_{\pi}(t) = \max_{y_k,v_t} \sum_{t=0}^{T-1} (y_k,r,\xi_t) \) over the price \( r \) at a cost of 
\[
O\left( \frac{T^4 \kappa^2 \lambda^2}{\delta^2} \log\left( \frac{r T \kappa}{\delta} \right) \right)
\]
samples, we can simply draw
\[
N(\delta^t, p^t) = N\left( \frac{\delta}{2T}, \frac{p}{T|X|} \right) = O\left( \frac{T^4 \kappa^2 \lambda^2}{\delta^2} \log\left( \frac{T|X|}{p} \right) \right)
\]
samples of \( G_{\pi}(t) \), keeping the price \( r \) fixed. The total number of samples required by this approach is
\[
O\left( \frac{|X| T^5 \kappa^2 \lambda^2}{\delta^2} \log\left( \frac{T|X|}{p} \right) \right).
\]

Appendix. Technical Proofs

Proof of Lemma 1. By the Hoeffding inequality (see Hoeffding (1963)),
\[
P[|f(x_i) - y_i| \geq \frac{\delta}{2}] \leq \exp\left( -\frac{2N(\delta, \frac{\delta}{2}) \delta^2}{M^2} \right) \leq \frac{p}{2}.
\]
Hence, \( f(x) \) passes through the two intervals \( I_i \) with probability at least \( 1 - p \). If \( f(x) \) goes through both intervals \( I_i \), then \( f \) must be decreasing on \([x_2, b] \). □

Proof of Lemma 3. By the Hoeffding inequality,
\[
P[|f(x_i) - y_i| \geq \frac{\delta}{2}] \leq \exp\left( -\frac{2N(\delta, \frac{\delta}{2}) \delta^2}{M^2} \right) \leq \frac{p}{2}.
\]
Hence, \( f(x) \) passes through the two intervals \( I_i \) with probability at least \( 1 - p \). If \( f(x) \) goes through both intervals \( I_i \) then the slope of \( f \) on \([x_2, b]\) is no more than \( \frac{\delta}{2\phi-1} \). Hence, the maximum value achievable by \( f \) on \([x_2, b]\) is no more than \( m_1 + \frac{\delta \phi}{2\phi-1} \). □

Proof of Theorem 2. The algorithm terminates in \( K = O(\log(\frac{1}{\epsilon})) \) iterations because the search interval is reduced by a constant factor of \( \phi \) at each iteration. With probability at least \( 1 - \sum_k p^k \leq 1 - p \), in every iteration \( k < K \),
\[
\arg \max_{[a^k, b^k]} f(x) \in [a^{k+1}, b^{k+1}]
\]
or
\[
\max_{[a^k, b^k] \setminus [a^{k+1}, b^{k+1}]} f(x) \leq m^k + \frac{\bar{\delta} \phi}{2\phi - 1} = m^k + \delta, \text{ and } m^k \leq f(x_k) \leq m^k + \bar{\delta} \leq m^k + \delta,
\]
where \((x^k, m^k)\) is the output of iteration \(k\). Further,
\[
m^K \leq f(x^K) \leq m^K + \delta
\]
and
\[
\hat{m} \leq f(\hat{x}) \leq \hat{m} + \delta.
\]
We assume that the above statements hold. This assumption is valid with probability at least \(1 - p\). We will finish proving that the output is correct by showing that \(f(x^*) \leq f(\hat{x}) + 2\delta\). If \(x^* \in [a^K, b^K]\) then since \(b^K - a^K \leq \epsilon\),
\[
f(x^*) = \max_{[a^K, b^K]} f(x) \leq m^K + 2\delta \leq \hat{m} + 2\delta \leq f(\hat{x}) + 2\delta.
\]
We now assume that \([a^K, b^K]\) does not contain a maximizer of \(f(\cdot)\) and we define \(x^*\) to be the last maximizer of \(f(\cdot)\) that the algorithm eliminated. If \(x^*\) was excluded from the active search interval at the end of iteration \(k\), then the GB Subroutine ended in either Step 4 or Step 5, and
\[
f(x^*) > \max_{[a^{k+1}, b^{k+1}]} f(x).
\]
Hence,
\[
f(x^*) \leq m^{k+1} + \delta \leq m^k + \hat{m} + \delta \leq f(\hat{x}) + 2\delta.
\]
Now, the number of samples required at each iteration \(k\) is
\[
O\left(\frac{M^2}{\delta^2} \log\left(\frac{1}{p_k}\right)\right) = O\left(\frac{M^2}{\delta^2} \log\left(\frac{2^k}{p}\right)\right)
\]
by Theorem 1. Summing from \(k = 1\) to \(k = K\), we see that the total number of samples required at all iterations is
\[
O\left(\frac{M^2}{\delta^2} \log\left(\frac{L}{\epsilon}\right) \log\left(\frac{L}{pe}\right)\right).
\]
□

Proof of Lemma 7. Note that the first statement holds if \(y^1 = y^2\). If \(y^1 \neq y^2\) then \(|y^1 - y^2| \geq 1.\)
\[
|G^\pi_{t-1}(y^1, r, \xi_{t-1}) - G^\pi_{t-1}(y^2, r, \xi_{t-1})|
\leq |f_t(y^1, r, D_{t-1}(r, \xi_{t-1})) - f_t(y^2, r, D_{t-1}(r, \xi_{t-1}))|
+ |v^\pi_t(y^1 - D_{t-1}(r, \xi_{t-1})) - v^\pi_t(y^2 - D_{t-1}(r, \xi_{t-1}))| \quad \text{by (11)}
\leq \kappa^G_{y^1} |y^1 - y^2| + |v_t(y^1 - D_{t-1}(r, \xi_{t-1})) - v_t(y^2 - D_{t-1}(r, \xi_{t-1}))| + 2\delta
\]
by Assumption 2 and by hypothesis
\[
\leq \kappa'_y|y^1 - y^2| + (T - t + 1)\kappa|y^1 - y^2| + 2\delta \quad \text{by Lemma 5}
\]
\[
\leq (\kappa'_y + 2\delta + (T - t + 1)\kappa)|y^1 - y^2|.
\]

The proof of the second and third statements are similar and follow from the corresponding results in Lemma 5. Finally, we prove the last statement. By the last statement of Lemma 5, we have that for each fixed \((y, r)\), with probability 1, \(G_{t-1}(y, r, \xi_{t-1})\) takes values in an interval of length at most \((T - t + 2)\kappa\lambda\). Comparing (8) and (11) and recalling that \(|v_t(x) - v_t^*(x)| \leq \delta\) for all \(x \in X\), we see that with probability 1, \(G_{t-1}(y, r, \xi_{t-1})\) takes values in an interval of length at most \((T - t + 2)\kappa\lambda + 2\delta\).

**Proof of Lemma 8.** We know by construction that \(v_{t-1}^\pi(\cdot)\) is concave and that \(v_{t-1}^\pi(y^k) \geq \bar{i}_{t-1}^\pi(y^k)\) for \(k = 1 \ldots |X|\). It suffices to check that \(v_{t-1}^\pi(y^k) \leq \bar{i}_{t-1}^\pi(y^k)\) for all integers \(k = 1 \ldots |X|\). Let

\[
h(x) = \max_{y \geq x} V_{t-1}^\pi(y).
\]

Then \(h(\cdot)\) is concave on \(X\) by the concavity and monotonicity of \(v_t^\pi(\cdot)\), and by the convexity of \(D_t\). Further, we are given that \(h(y^k) \in I_{t-1}^\pi(y^k)\) for all \(k = 1 \ldots |X|\). Let \(E\) be the convex hull of \(\{(x, t)|x \leq x \leq \bar{x}, \ t \leq h(x)\}\). Then \(E\) is a convex set containing \(C\), which as we recall, is the convex hull of \(\{(y^k, \bar{i}_{t-1}^\pi(y^k))|k = 1 \ldots |X|\}\). Hence, we have that

\[
\bar{i}_{t-1}^\pi(y^k) \geq \max\{t|(y^k, t) \in E\} \geq \max\{t|(y^k, t) \in E\} = v_{t-1}^\pi(y^k)
\]

for all \(k\). □

**Proof of Lemma 9.** Recall that \(K\) is the value that \(h\) had when Step 3 of Part I was executed. For \(k \geq K\), the Gradient Bound algorithm returns \(J_{t-1}^\pi(y^k)\) of length \(\frac{p^t}{2}\) such that

\[
P[V_{t-1}^\pi(y^k) ] \in J_{t-1}^\pi(y^k)] \geq 1 - \frac{p^t}{|X|}.
\]

In other words, with probability at least \(1 - (|X| - K + 1)\frac{p^t}{|X|} \geq 1 - p^t\), the concave function \(V_{t-1}^\pi(y^k)\) passes through \(J_{t-1}^\pi(y^k)\) for all \(k \geq K\). Assume this is the case.

We need to show that \(\max_{y \geq y^k} V_{t-1}^\pi(y^k) \leq I_{t-1}^\pi(y^k)\) for all \(k\). Because \(I_{t-1}^\pi(y^k) = J_{t-1}^\pi(y^*)\) and \(y^*(k) \geq y^k\), we have

\[
\bar{i}_{t-1}^\pi(y^*(k)) \leq V_{t-1}^\pi(y^*(k)) \leq \max_{y \geq y^k} V_{t-1}^\pi(y) \leq \max_{y \geq y^k} \bar{i}_{t-1}^\pi(y) = \bar{i}_{t-1}^\pi(y^*(k)).
\]

□
Proof of Theorem 4. By Lemma 9 and Theorem 2, the ICP Algorithm returns intervals $I^\pi_{i-1}(\cdot)$ of length $\frac{d^i}{2}$ such that $P\left[\max_{y \geq x} V^\pi_{i-1}(y) \in I^\pi_{i-1}(x) \forall x \in X\right] \geq 1 - p^i$ using

$$O\left(\frac{|X|(2\delta^i + \lambda T \kappa)^2}{(\delta^i)^2} \log\left(\frac{\bar{r}T \kappa}{\delta^i}\right) \log\left(\frac{|X|T \kappa}{p^i \delta^i}\right)\right) = O\left(\frac{|X|\lambda^2 T^2 \kappa^2}{(\delta^i)^2} \log\left(\frac{\bar{r}T \kappa}{\delta^i}\right) \log\left(\frac{|X|T \kappa}{p^i \delta^i}\right)\right)$$

samples. The algorithm also computes a concave function $v^\pi_{i-1}(\cdot)$ such that $v^\pi_{i-1}(x) \in I^\pi_{i-1}(x)$ for all $x \in X$. Hence,

$$P\left[\max_{y \geq x} V^\pi_{i-1}(y) - v^\pi_{i-1}(x) \leq \delta^i \forall x \in X\right] \geq 1 - p^i.$$ 

Since we are given that $P[|v_t(x) - v^\pi_t(x)| < \delta \forall x \in X] \geq 1 - p$,

it follows that

$$P\left[|v_{t-1}(x) - v^\pi_{t-1}(x)| \leq \delta^i \forall x \in X\right] \geq 1 - p^i - p.$$

□

Proof of Lemma 10. By the Hoeffding inequality,

$$P[|f(x_i) - y_i| \geq \frac{\delta}{2}] \leq \exp\left(-\frac{2N(\delta, \frac{\delta}{2})^2}{M^2}\right) \leq \frac{p}{2}.$$ 

Hence, $f(x)$ passes through the two intervals $I_i$ with probability at least $1 - p$. If $f(x)$ goes through both intervals $I_i$, then the slope of $f$ on $[x_2, b]$ is no more than $\frac{2\delta}{(b-a)(2\phi-1)}$. Hence, the maximum value achievable by $f$ on $[x_2, b]$ is no more than $m_1 + \frac{2\phi}{2\phi-1} \delta$. □

Proof of Theorem 7. The algorithm terminates in $K = O(\log(\frac{1}{\delta}))$ iterations because the search interval is reduced by a constant factor of $\phi$ at each iteration. With probability at least $1 - \sum_k p^k \geq 1 - p$, we may assume that at every iteration $k < K$,

$$\arg \max_{[a^k, b^k]} f(x) \in [a^{k+1}, b^{k+1}]$$

or

$$\max_{[a^k, b^k], [a^{k+1}, b^{k+1}]} f(x) \leq m^k + \frac{2\phi}{2\phi-1} \delta,$$

where $(x^k, m^k)$ is the output at iteration $k$. Further,

$$m^K \leq f(x^K) \leq m^K + \tilde{\delta},$$

and

$$\hat{m} \leq f(\hat{x}) \leq \hat{m} + \tilde{\delta}.$$ 

Assume the above statements hold.
Let \( x^* = \arg \max_{[a,b]} f(x) \). Note that the algorithm returns \( \hat{m} \), and that all \( m^k \) satisfy \( m^k \leq \hat{m} + \delta \), including those eliminated in the Elimination Round. If \( x^* \in [a^K, b^K] \) then \( a^K = x^K \leq x^* \), and

\[
f(x^*) \leq m^K + \bar{\delta} \leq \hat{m} + 2\bar{\delta} < f(\hat{x}) + \delta.
\]

Otherwise, \( x^* \) was excluded from the active search interval at the end of iteration \( \bar{k} < K \), and \( x^* \leq x^\bar{k} \). Furthermore,

\[
f(x^*) \leq m^\bar{k} + \frac{2\phi}{2\phi - 1} \bar{\delta} \leq \hat{m} + \frac{2\phi}{2\phi - 1} \bar{\delta} + \bar{\delta} = \hat{m} + \delta \leq f(\hat{x}) + \delta.
\]

Now, the number of samples required at each iteration \( k \) is

\[
O\left( \frac{M^2}{(\delta)^2 \log \left( \frac{1}{p^k} \right)} \right) = O\left( \frac{M^2}{\delta^2 \log \left( \frac{2^k}{p} \right)} \right)
\]

by Theorem 1. Hence, the total number of samples required at all iterations is

\[
O \left( \frac{M^2}{\delta^2} \log \left( \frac{L}{\epsilon} \right) \log \left( \frac{L}{p\epsilon} \right) \right).
\]

Note that the work done in the elimination round is only \( O(K) = O(\log \left( \frac{L}{\epsilon} \right)) \).

Proof of Lemma 11. First, we claim that for all \( x^1 \) and \( x^2 \),

\[
|y^\pi_t(x^1) - y^\pi_t(x^2)| \leq \alpha |x^1 - x^2|.
\]

To prove the claim, examine (19). It is simple to show the claim in all cases except for the third line of (19). When the third line holds, note that

\[
\left| \frac{\partial y^\pi_t(x)}{\partial x} \right| = \left| \frac{y^+_t - y^-_t}{x^+_t - x^-_t} \right| \leq \frac{\bar{y}^\pi_t - \bar{y}^\pi_t}{x^+_t - x^-_t} \leq \alpha.
\]

The last inequality follows from the definitions of \( \alpha \), \( x^+_t \) and \( x^-_t \). Integrating over \( x \), we establish the claim.

Next, we prove the first statement of the Lemma. Examining (18) and (19) and applying the claim, we see that

\[
\kappa^F_{t-1, y} = \kappa^f_y + \alpha n^3 \kappa^F_{t, y}.
\]  \hfill (22)

Moreover, \( \kappa^F_t = \kappa^f_y \). Note that when we iterate (22), terms of the type \( \beta, \beta^2, \ldots \) appear. We conclude that

\[
\kappa^F_{t-1, y} = \kappa^f_y \sum_{k=0}^{T-t} \beta^k \sum_{k=0}^{T-t} \beta^k,
\]

which simplifies to the expression in the first statement of the Lemma.
Using the first statement of the lemma, the claim, and \( (19) \), we obtain the second statement of the lemma as follows:

\[
\kappa_{t,x}^w = \alpha \kappa_{t,y}^F = \alpha \kappa_y^f \sum_{k=0}^{T-t} \beta^k = \kappa O(\sigma_1(\beta, T - t + 1)).
\]

Lastly, we prove the final statement of the lemma. Recall that by Assumption 1, \( \Lambda_t^F = \lambda \kappa_d^f \). From \( (18) \), Assumption 4, and the second statement of the lemma,

\[
\Lambda_{t-1}^F = \lambda \kappa_0^d + \lambda \kappa_{t,x}^w + \Lambda_t^F = (T - t + 2) \kappa_d^f + \lambda \sum_{s=t}^T \kappa_{s,x}^w = \lambda \kappa O(\sigma_2(\beta, T - t + 2)),
\]

which is as we claim in the third statement of the lemma. \( \square \)

**Proof of Theorem 8.** First, note that \( v_t^\pi(\cdot) \) and \( V_t^\pi(\cdot) = E[F_t^\pi(\cdot, \xi_t, \ldots, \xi_T)] \) are well defined by Lemma 11. We now show that \( v_t^\pi(\cdot) \) is concave. Note that by definition, \( w_t^\pi(x, \xi_t, \ldots, \xi_T) \) is a continuous function of \( x \).

Note that the conditions on the right hand side of \((19)\) all depend on \( x \) rather than on \( (\xi_t, \ldots, \xi_T) \). Therefore, by taking expectations in \((19)\), we obtain an expression for \( v_t^\pi \) in terms of \( V_t^\pi \). Since \( V_t^\pi(\cdot) \) is concave, the function \( \max_{(x, y) \in A_t} V_t^\pi(y) \) is a concave function of \( x \).

We claim that \( v_t^\pi(x) \) is identical to the concave function \( \max_{(x, y) \in A_t} V_t^\pi(y) \) everywhere on \( R_t^\pi \cup \neg S_t^\pi = (\infty, x_t^-) \cup (x_t^+, \infty) \). To establish the claim, notice that \( R_t^\pi \) and \( \neg S_t^\pi \) are open and disjoint. Consequently, each of the intervals \( (\infty, x_t^-) \) and \( (x_t^+, \infty) \) is a subset of either \( R_t^\pi \) or \( \neg S_t^\pi \). For \( x \in R_t^\pi \), if \( (x, y) \in A_t \), we have

\[
y > \bar{y}_t^\pi \geq y_t^\pi,
\]

so that

\[
v_t^\pi(x) = V_t^\pi(l_t(x)) = V_t^\pi(\min_{(x, y) \in A_t} y) = \max_{(x, y) \in A_t} V_t^\pi(y).
\]

Similarly, \( v_t^\pi(x) \) is equal to \( \max_{(x, y) \in A_t} V_t^\pi(y) \) on \( \neg S_t^\pi \). This proves the claim.

Note that by \((19)\), \( v_t^\pi(x_t^-) = V_t^\pi(y_t^-) \) if \( x_t^- \) is finite, and that \( x_t^+ \) and \( y_t^+ \) are similarly related. We now prove that \( v_t^\pi(\cdot) \) is concave everywhere by considering 4 cases:

1. If \( x_t^- \) and \( x_t^+ \) are both finite then we have shown that \( v_t^\pi(x) = E[w_t^\pi(x, \xi_t, \ldots, \xi_T)] \) is concave and equal to \( \max_{(x, y) \in A_t} V_t^\pi(y) \) on \( (\infty, x_t^-) \cup [x_t^+, \infty) \). On \( [x_t^-, x_t^+] \), \( v_t^\pi(x) \) is concave because by \((19)\), it is the composition of the concave function \( V_t^\pi(\cdot) = E[F_t^\pi(\cdot, \xi_t, \ldots, \xi_T)] \) and a linear function. Because \( w_t^\pi \) is Lipschitz-continuous, so is \( v_t^\pi \). It suffices to show that \( v_t^\pi \) is locally concave at \( x_t^- \) and \( x_t^+ \).
At $x_t^-$, the left-hand derivative of $v_t^\pi$ is equal to the left-hand derivative of $\max_{(x,y) \in A_t} V_t^\pi(y)$. But the right-hand derivative of $v_t^\pi$ at $x_t^-$ is less than or equal to the right-hand derivative of $\max_{(x,y) \in A_t} V_t^\pi(y)$, since

$$\max_{(x,y) \in A_t} V_t^\pi(y) \geq V_t^\pi(y_t^\pi(x)) = v_t^\pi(x)$$

for all $x \in [x_t^-, x_t^+]$. The right-hand derivative of $\max_{(x,y) \in A_t} V_t^\pi(y)$ is in turns less than or equal to the left-hand derivative of $\max_{(x,y) \in A_t} V_t^\pi(y)$ by the concavity of $\max_{(x,y) \in A_t} V_t^\pi(y)$. Hence, at $x_t^-$, the left-hand derivative of $v_t^\pi$ is greater than or equal to the right-hand derivative of $v_t^\pi$. We conclude that $v_t^\pi$ is locally concave at $x_t^-$. Similarly, $v_t^\pi$ is locally concave at $x_t^+$.

2. If $x_t^-$ is finite and $x_t^+ = \infty$, then by (19), $v_t^\pi(x) = \max_{(x,y) \in A_t} V_t^\pi(y)$ on $(-\infty, x_t^-]$ and $v_t^\pi(x) = \max_{(x,y) \in A_t} V_t^\pi(y)$ on $[x_t^-, \infty)$. It suffices to show that no maximizer of $\max_{(x,y) \in A_t} V_t^\pi(y)$ is in $(-\infty, x_t^-)$, i.e., that if $x < x_t^-$ then $(x, y_t^\pi) \notin A_t$. This follows directly from the definitions of $x_t^-$, $R_t^\pi$, and $S_t^\pi$.

3. If $x_t^- = -\infty$ and $x_t^+ = \infty$, then the proof of the second case applies by symmetry.

4. If $x_t^- = -\infty$ and $x_t^+ = \infty$, then $v_t^\pi(x, \xi_1, \ldots, \xi_T)$ is constant in $x$, and $v_t^\pi(x)$ is concave. By the linearity of $g_t(.)$, $v_t[g_{t-1}(y_{t-1}, \xi_{t-1}), \xi_t, \ldots, \xi_T]$ is concave in $y_{t-1}$, so taking expectations in (18), we see that $V_{t-1}^\pi(.)$ is concave. \hfill \square

**Proof of Lemma 13.** Assume that the events in parentheses in (20) and (21) are true. (This happens with probability $1 - p_1 - p_2$.) Theorem 8 implies that $V_t^\pi(.)$ is concave. Since

$$V_t(y) = E[f_t(y, D_t(\xi_t)) + v_{t+1}(g_t(y, \xi_t))], \quad \text{and}$$

$$V_t^\pi(y) = E[f_t(y, D_t(\xi_t)) + v_{t+1}^\pi(g_t(y, \xi_t))],$$

it suffices to show that

$$v_{t+1}^\pi(x_{t+1}) \geq v_{t+1}(x_{t+1}) - \delta_1 - \delta_2. \quad (23)$$

But because

$$V_{t+1}(y) - V_{t+1}^\pi(y) \leq \delta_2 \text{ for all } y,$$

$$\max_{(x_{t+1}, y_{t+1}) \in A_{t+1}} V_{t+1}^\pi(y_{t+1}) \geq \max_{(x_{t+1}, y_{t+1}) \in A_{t+1}} V_{t+1}(y_{t+1}) - \delta_2 = v_{t+1}(x_{t+1}) - \delta_2,$$

and by Lemma 12,

$$\max_{(x_{t+1}, y_{t+1}) \in A_{t+1}} V_{t+1}^\pi(y_{t+1}) \leq v_t^\pi(x_{t+1}) + \delta_1.$$

Hence, (23) follows. \hfill \square
References


Huh, W. T., P. Rusmevichientong. 2006b. A nonparametric approach to stochastic inventory planning with lost sales and censored demand. Multi-Echelon Conference, Atlanta, GA.


