SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1061

July 1993

**MULTI-ITEM, MULTI-PERIOD
PRODUCTION PLANNING
WITH UNCERTAIN DEMAND**

by

John A. Muckstadt¹
and Charles R. Sox²

¹Research was supported in part by the National Science Foundation grant DDM-9215328.

²Department of Industrial Engineering, Auburn University.

Abstract

We provide an approximate formulation and solution algorithm for the classical finite-horizon capacitated production planning problem with random demand for multiple products. Using Lagrangian relaxation, we develop a subgradient optimization algorithm to solve this formulation and provide some computational results that indicate this approach works very well when used for rolling-horizon planning. The advantage of our approach is that realistic problem instances can be solved quite quickly while optimal solutions to such instances are computationally intractable.

Key Words: stochastic production planning, finite-horizon planning

1 Introduction

Practical instances of the finite-horizon production and inventory planning problem are difficult to solve. Aggregation is often used to simplify the problem instance and demands are assumed to be known while safety stock is specified by some external calculations. It is well known that aggregate solutions can be capacity infeasible, and that developing a disaggregated solution is not a straightforward task. Furthermore, these problems become more complicated as product variety increases. Both the number of items and the coefficient of variation of the demand process for each time period tend to increase. Thus, production planning schemes that effectively account for demand uncertainty and product variety are of great benefit to companies facing this kind of environment.

Production-inventory systems are often managed with constant base-stock or safety-stock levels. These policies work well in environments with adequate production capacity and low demand variability. However, when there is insufficient capacity to restore all items to their base-stock or safety stock level, production capacity must be allocated among the items, and the question of which items to stock must be answered. In practice, this allocation process is often done in an *ad hoc* manner. Furthermore, seasonal demand patterns require time-dependent production and inventory plans. We propose to solve these problems dynamically by using a rolling-horizon planning algorithm that dynamically specifies inventory levels and production quantities.

In this paper we develop a production planning methodology that accounts for these issues directly and that demonstrates excellent performance in our computational experiments. We formulate the problem as a discrete time, finite-horizon optimization problem. Discrete time models are the most prevalent tools used to study manufacturing and distribution systems because the time structure parallels the normal administrative and planning cycles found in these systems. While there is much interest and value in

real-time planning and scheduling for some environments, many planning activities continue to be performed on a periodic basis for economic and administrative reasons. The discrete time structure readily lends itself to mathematical programming formulations and techniques like dynamic programming and stochastic programming. Moreover, it allows modelling transient and periodic changes in system parameters like demand and cost as well as control parameters like inventory levels and production quantities.

We propose that the finite-horizon algorithm be used on a rolling-horizon basis. The first period's plan is implemented, demands are observed, and the net inventory levels are updated. Then the cost and forecast parameters are incremented, and the algorithm is again applied to the new problem instance. As noted by Baker [1], this approach is quite effective in practice. Our contribution is an algorithm for developing these finite-horizon plans for capacitated systems with multiple products having uncertain demand over the planning horizon.

The rest of the paper is organized as follows. The next section reviews some of the literature on finite-horizon production planning. Section 3 details our model while Section 4 describes our solution algorithm. Then Section 5 describes our experiments and computational results using the planning algorithm, and Section 6 offers some concluding remarks.

2 Literature

The earliest analysis of a dynamic, stochastic, multi-item, production-inventory control problem seems to be by Evans [4] who was able to partially describe the optimal solution for his model. Elmaghraby [3] provides a review of the deterministic version of this problem, the economic lot scheduling problem, ELSP. More recently, we have the work of Graves [9], Williams [16], Gascon [7], and Gascon and Leachman [13, 8]. Williams [16]

studies a deterministic model of an assembly and distribution network. Graves [9] and Gascon [7] study the dynamic stochastic version. Gascon and Leachman [8] analyze the dynamic, deterministic version while Leachman and Gascon [13] develop a more complex heuristic for implementation in a specific instance of the stochastic problem. The works most closely related to ours are Evans [4], Haneveld [10], Peters, Boskma, and Kupper [15], Khang and Fujiwara [12], and Bitran and Yanasse [2].

Using a model essentially the same as ours, Evans [4] formulates the problem as a dynamic program, and, assuming stationarity, he proves the optimality of a type of base-stock policy in which each item has a maximal stock level which is attained whenever there is sufficient production capacity. When there is not enough capacity, he was unable to characterize the optimal policy.

Haneveld [10] and Peters, Boskma, and Kupper [15] formulate the problem as a stochastic program with recourse. Although Peters et al. do not model backordering, Haneveld shows how their model can be generalized to our case as well as others and indicates how more than one production activity resource can be modelled. One solution technique is to discretize the probability distributions and to formulate the problem as a linear program to which one can apply large-scale linear programming algorithms. Another approach is to apply quasi-gradient search algorithms designed to solve stochastic programming problems. The experimentation reported in each of these articles is quite limited. The numerical problem reported in [15] consists of two items, four periods, and three production resources. Our approach improves upon these stochastic programming formulations by efficiently handling a large number of items and by allowing an exact solution without truncation or discretization of the demand distributions. Although recent advances in computer technology have provided greater computational resources, stochastic dynamic programming and stochastic programming are still limited to solving problem instances which contain a relatively small number of products when compared to situations

encountered in practice.

In [2], Bitran and Yanasse describe a heuristic for multi-period production planning with service level constraints. They develop a deterministic approximation that has small relative error for instances with high service levels. They provide numerical results for some small problem instances where either a normal, gamma, or Poisson distribution describes the demand process. They develop a deterministic approximation for both the single stage and multi-stage (sequential) decision problems and derive error bounds. Their single stage problem has linear holding costs and concave production costs while the sequential problem has linear production costs. Khang and Fujiwara [12] borrow heavily from [2]. They formulate a multi-period network flow problem with service level constraints, linear holding costs, capacitated flows, and concave flow costs. They derive a deterministic approximation that can be solved using conventional network flow algorithms and a bound on the relative error. They also provide some computational results when demand follows either a normal, exponential, or Poisson distribution.

The contribution of this paper is the development of a model and solution algorithm for the finite-horizon capacitated production planning problem with random demand for multiple products. Our model includes backorder costs in the objective rather than enforcing service level constraints. Although our approach is an approximate one, as we will explain later, the algorithm described here allows the analysis of problem instances having a very large number of items compared with previous research. We provide theoretical results that allow the use of general discrete or continuous distributions to model item demand.

3 Model

We model a finite horizon, discrete time, production and inventory planning problem with multiple heterogeneous products having general demand distributions. We can model many different limited production resources like multiple production capacities and limited raw materials. Each of these resource limits might have various sources of additional resource like overtime and outsourcing or different component suppliers. Although the theoretical results presented here apply to instances with multiple resource limitations, we will assume a single bottleneck production capacity to simplify the notation. Each item consumes this resource at a linear rate that may vary between items, and the cost of additional capacity is linear with a rate that depends on its source. We include linear holding and backorder costs; but, setup costs and time are not included. Setups can be included at the expense of adding a combinatorial structure to the problem that can be tackled using implicit enumeration or dynamic programming. However, the formulation with setups reduces to the problem described here once the setups are determined, so this model can be viewed as a subproblem for the problem with setups.

The following notation will be used in the mathematical programming formulation of the model. Let

X_{it}	=	the cumulative production of item i through period t ,
X_{i0}	=	the initial inventory of item i ,
h_{it}	=	the holding cost of item i in period t ,
π_{it}	=	backorder or penalty cost of item i in period t ,
a_i	=	the rate of capacity usage for item i ,
$p_{it}(x)$	=	the p.d.f. of cumulative demand for i through t ,
$P_{it}(x)$	=	the c.d.f. of cumulative demand for i through t ,
$\overline{P}_{it}(x)$	=	$1 - P_{it}(x)$,
y_{jt}	=	the amount of capacity type j used in period t ,
b_{jt}	=	the cost of capacity type j in period t , and
u_{jt}	=	the limit of type j capacity available in period t .

Our formulation of the finite horizon, production-inventory planning problem, PIP, can be

stated as

$$\begin{aligned}
& \min \quad \sum_{t=1}^T \sum_{i=1}^N L_{it}(X_{it}) + \sum_{t=1}^T \sum_{j=1}^J b_{jt} y_{jt} \\
& \text{PIP} \quad \text{s.t.} \quad \sum_{i=1}^N a_i (X_{it} - X_{i,t-1}) \leq \sum_{j=1}^J y_{jt}, \quad \forall t, \\
& \quad \quad \quad X_{i,t-1} \leq X_{it}, \quad \forall i \text{ and } t, \\
& \quad \quad \quad y_{jt} \leq u_{jt}, \quad \forall j \text{ and } t, \\
& \quad \quad \quad y_{jt} \geq 0, \quad \forall j \text{ and } t,
\end{aligned}$$

where $L_{it}(x)$ is the expected holding and backorder cost function for product i in period t

$$L_{it}(X_{it}) = h_{it} \int_0^{X_{it}} (X_{it} - x) p_{it}(x) dx + \pi_{it} \int_{X_{it}}^{\infty} (x - X_{it}) p_{it}(x) dx,$$

for continuous demand, and the cost function for discrete demand replaces integration with summation in the preceding expression. The objective is to minimize the expected costs of holding inventory, of backorders, and of extra capacity over the time horizon. The first set of constraints are the production capacity constraints. Note that capacity is a variable with an associated cost. The next set of constraints, which we refer to as consistency constraints, ensures that the cumulative production variables are non-decreasing. The last constraints are bounds on the use of production capacity. The formulation also assumes that $X_{i0} \geq 0$; however, the initial inventory position may be negative due to outstanding backorders. In this situation it is quite reasonable to assume that these backorders will be filled before any additional inventory is produced. Thus, we simply reduce the available capacity in the first period by the appropriate amount needed to cover these outstanding backorders. By convention, we assume that $j = 1$ corresponds to normal capacity and that the capacity costs are ordered: $0 \leq b_{1t} < b_{2t} < \dots < b_{Jt}$, for all t .

Our model is an approximation, as mentioned earlier, because it does not capture the impact of being able to re-plan future production. Conceptually, the model plans production for the entire horizon at the beginning of the planning horizon. Consequently, it

does not allow a new decision to be made for period two, for example, after observing demand in period one. For this reason, we propose using this model in a rolling-horizon manner. The stochastic programming formulations mentioned earlier do not capture this effect either, as is discussed in Peters et al. [15]. Our formulation has the same structure as the stochastic programming formulation with simple recourse described by Haneveld [10], and it produces an objective function and constraint set that are separable across products and time periods using the appropriate relaxations. The assumption of independent, identically distributed random variables is not needed. We only need to compute, or at least approximate, the c.d.f. of the cumulative demand processes.

4 Solution Algorithm

The approach used to solve an instance of this problem is to apply subgradient optimization to a Lagrangian dual of the original problem, PIP. This approach is quite common for solving integer programming problems. Fisher [6] gives an excellent description of this approach and different algorithms for solving the Lagrangian dual and provides a review of some applications. Nemhauser and Wolsey [14] discuss both Lagrangian relaxation and subgradient optimization, and Held, Wolfe, and Crowder [11] provide a good description of subgradient optimization. Everett [5] is the first to have proposed using Lagrangian relaxation in solving constrained resource allocation problems.

Given a mathematical program, P , in which the objective and constraints are convex, we can construct a *Lagrangian relaxation* of P , $LR(\theta)$, by incorporating some of the constraints into the objective function using nonnegative Lagrange multipliers, $\theta = (\theta_1, \dots, \theta_m)$. Let Z and $Z_{LR}(\theta)$ be the optimal values of P and $LR(\theta)$, respectively. If the objective is being minimized, then $Z_{LR}(\theta) \leq Z$ for all $\theta \geq 0$. Thus, $Z_{LR}(\theta)$ provides a lower bound on the optimal value of the original problem. The problem of maximizing

$Z_{LR}(\theta)$, LD, is called a *Lagrangian dual* of the original problem. In some cases, $Z_{LD} = Z$; however, this is often not the case for integer programs, and the difference $Z - Z_{LD}$ is called the *duality gap*. In its application to Lagrangian relaxation, *subgradient optimization* is a method of searching for the optimal multipliers, i.e. solving the Lagrangian dual. Because the gradient of the dual objective is difficult to compute or fails to exist, a *subgradient* is used as the search direction instead. As noted by Held et al. [11], a subgradient is not guaranteed to be an improving direction, but it does make an acute angle with any ray from the current solution to an optimal solution. Thus, a sufficiently small step in the direction of a subgradient will be closer to an optimal solution than the current solution. The specific details of our planning algorithm which is based on subgradient optimization will be described later. First, we describe the Lagrangian relaxation used in our algorithm and how it can be easily solved for a given set of multipliers.

There are many possible relaxations to our production planning problem, PIP. We have chosen one in which the T capacity constraints are relaxed with Lagrange multipliers, $\theta_t, t = 1, \dots, T$. This choice yields the following Lagrangian relaxation, $PIP(\theta)$, of the original problem,

$$\begin{aligned}
PIP(\theta) \quad \min \quad & \sum_{i=1}^N \sum_{t=1}^{T-1} [L_{it}(X_{it}) + a_i(\theta_t - \theta_{t+1})X_{it}] + \sum_{i=1}^N [L_{iT}(X_{iT}) + a_i\theta_T X_{iT}] \\
& - \sum_{i=1}^N a_i\theta_1 X_{i0} + \sum_{t=1}^T \sum_{j=1}^J (b_{jt} - \theta_t)y_{jt} \\
\text{s.t.} \quad & X_{i,t-1} \leq X_{it}, \quad \forall i \text{ and } t, \\
& y_{jt} \leq u_{jt}, \quad \forall j \text{ and } t, \\
& y_{jt} \geq 0, \quad \forall j \text{ and } t.
\end{aligned}$$

Note that the production and capacity variables are no longer related through the constraint set. The separability of this Lagrangian relaxation into a capacity subproblem and subproblems for each of the items makes it relatively easy to solve. Indeed, the formulation in terms of cumulative production quantities and the choice of relaxed

constraints create this separability. In fact, this relaxation was chosen for this separability.

We now show how the subproblems can be easily solved.

The subproblem for the capacity variables may be stated as

$$\begin{aligned} \text{CSP}(\theta) \quad & \min \sum_{t=1}^T \sum_{j=1}^J (b_{jt} - \theta_t) y_{jt} \\ & \text{s.t.} \quad y_{jt} \leq u_{jt}, \quad \forall j \text{ and } t, \\ & \quad y_{jt} \geq 0, \quad \forall j \text{ and } t. \end{aligned}$$

Let $y(\theta)$ be the vector of optimal capacity variables for $\text{CSP}(\theta)$. Obviously, if $b_{kt} < \theta_t$, then $y_{kt}(\theta) = u_{kt}$; if $b_{kt} > \theta_t$, then $y_{kt}(\theta) = 0$; and if $b_{kt} = \theta_t$, then $y_{kt}(\theta)$ may take on any value in $[0, u_{kt}]$. However, from the Kuhn-Tucker conditions, we know for $(X(\theta), y(\theta))$ to be an optimal solution of problem PIP, that

$$\sum_{i=1}^N a_i (X_{it}(\theta) - X_{i,t-1}(\theta)) = \sum_{j=1}^J y_{jt}(\theta)$$

when $\theta_t > 0$. Because the capacity types are indexed by increasing cost, b_{kt} , it follows that

$$y_{jt}(\theta) = \begin{cases} 0, & \text{if } \theta_t < b_{jt}, \\ u_{jt}, & \text{if } \theta_t > b_{jt}, \\ \min\{u_{jt}, \max\{0, \sum_{i=1}^N a_i (X_{it}(\theta) - X_{i,t-1}(\theta)) - \sum_{k=1}^{j-1} u_{kt}\}\}, & \text{if } \theta_t = b_{jt}, \end{cases}$$

describes an optimal solution to $\text{CSP}(\theta)$. Therefore, given the Lagrange multipliers and the resulting cumulative production quantities, we can compute the capacity variables.

Next, we must determine the optimal cumulative production quantities for problem $\text{PIP}(\theta)$. Dropping the product index, an item subproblem may be stated as

$$\begin{aligned} \text{SP}(\theta) \quad & \min \quad f(X) = \sum_{t=1}^T L_t(X_t) + \sum_{t=1}^{T-1} a(\theta_t - \theta_{t+1})X_t + a\theta_T X_T \\ & \text{s.t.} \quad X_{t-1} \leq X_t, \quad \forall t, \end{aligned}$$

with $L_t(X_t)$ again representing the expected holding and backorder costs in period t .

Without the consistency constraints, these subproblems would be quite straightforward to

solve. For the moment, we ignore these side constraints to discuss the following result. For convenience, define

$$\gamma_t = \begin{cases} \frac{h_t + a(\theta_t - \theta_{t+1})}{h_t + \pi_t}, & \text{for } t=1, \dots, T-1 \text{ and} \\ \frac{h_T + a\theta_T}{h_T + \pi_T}, & \text{for } t=T. \end{cases}$$

Theorem 1 *If $\gamma_t \in (0, 1)$, then for any unconstrained minimum, X , of $f(X)$*

a) if the demand distribution is continuous, X_t solves

$$\bar{P}_t(X_t) = \gamma_t, \text{ or} \quad (1)$$

b) if the demand distribution is discrete with nonnegative support, X_t is

$$X_t = \min\{X_t \geq 0 : \bar{P}_t(X_t) \leq \gamma_t\}. \quad (2)$$

A detailed proof of this theorem is provided in the appendix. If $\gamma \notin (0, 1)$ then X_t is determined by one of four cases i) $\gamma_t > 1$, ii) $\gamma_t = 1$, iii) $\gamma_t = 0$, and iv) $\gamma_t < 0$. These cases are also detailed in the appendix.

Note how these equations are strikingly similar to the solution of the classic newsboy problem. The formulation of the finite horizon problem in terms of cumulative quantities creates the decomposition into single item, single period problems, so that this formulation and solution are a natural extension of the classic newsboy problem. In fact, the solution of these subproblems can be interpreted as a sequence of newsboy solutions in which the consistency constraints are enforced. As in the newsboy problem, this solution captures the tradeoff between holding and backorder costs. However, it also captures the tradeoff between capacity in the current period and the next through the expression $a_i(\theta_t - \theta_{t+1})$. The Lagrange multipliers may be interpreted as the marginal value of additional capacity in a given period. If $\theta_t < \theta_{t+1}$, holding costs in period t are subsidized by the high value of capacity in the next, and additional inventory is carried over into the next period to release some capacity that may or may not be used for other items. On the

other hand, $\theta_t > \theta_{t+1}$ effectively raises holding costs in period t so that production in period t is penalized thereby making production in future periods seem more attractive. The elegance of Theorem 1 and its interpretation again reinforce this formulation as a natural extension of the newsboy problem to a multi-item, multi-period, capacitated model.

Until now, we have ignored the consistency constraints in $SP(\theta)$. However, the following result neatly handles these constraints.

Theorem 2 *If $f(X)$ is convex and X° is the unconstrained minimum of f and $X_{t-1}^\circ > X_t^\circ$, then $X_{t-1}^* = X_t^*$ for any optimal solution, X^* , of $SP(\theta)$.*

A detailed proof of this result for both continuous and discrete demand distributions is also contained in the appendix. We may interpret this theorem in the following manner. If the unconstrained minimum of $f(X)$ violates any consistency constraint $X_{t-1} \leq X_t$, then that constraint will be tight in any optimal solution of $SP(\theta)$. Observe that when $X_{t-1} = X_t$ there is no production of this product in time period t . The original single-item subproblem may be reformulated with fewer time periods by substituting X_{t-1} for X_t . This reformulation is in terms of periods in which production of this item actually takes place. We call these periods *production periods*. Let t_k denote the k -th production period of which there are K in total, and we may re-write $SP(\theta)$ as

$$\begin{aligned} SP(\theta) \quad \min \quad & f(X) = \sum_{k=1}^K L_{t_k}(X_{t_k}) + \sum_{k=1}^{K-1} a(\theta_{t_k} - \theta_{t_{k+1}})X_{t_k} + a\theta_{t_K}X_{t_K} \\ \text{s.t.} \quad & X_{t_{k-1}} \leq X_{t_k}, \quad \forall k, \end{aligned}$$

where $t_0 = 0$. For continuous demand distributions, the holding and backorder cost functions are

$$L_{t_k}(X_{t_k}) = \sum_{t=t_k}^{t_{k+1}-1} \left[h_t \int_0^{X_{t_k}} (X_{t_k} - x)p_t(x)dx + \pi_t \int_{X_{t_k}}^{\infty} (x - X_{t_k})p_t(x)dx \right],$$

for $k < K$, and for the last production period the summation is from t_K to T . For discrete demand, we may exchange the integration for summation in the obvious manner. One can

see that these holding and backorder cost functions are nonnegative, linear combinations of the original cost functions; so, they are convex and Theorem 2 holds for this reformulation. The equations that define the unconstrained optimal solution for the reformulation are analogous to equations 1 and 2. However, we must still consider the four special cases that are analogous to those discussed following Theorem 1. For now, assume the solution does not fall under one of these cases. Then for $k < K$, $X_{t_k}^o$ is the solution of

$$\sum_{t=t_k}^{t_{k+1}-1} (h_t + \pi_t) \bar{P}_t(X_{t_k}) = h_{t_k, t_{k+1}-1} + a(\theta_{t_k} - \theta_{t_{k+1}}), \text{ or} \quad (3)$$

$$X_{t_k}^o = \min\{X \geq 0 : \sum_{t=t_k}^{t_{k+1}-1} (h_t + \pi_t) \bar{P}_t(X) \leq h_{t_k, t_{k+1}-1} + a(\theta_{t_k} - \theta_{t_{k+1}})\}, \quad (4)$$

where $h_{t_k, t_{k+1}-1} = \sum_{t=t_k}^{t_{k+1}-1} h_t$. When $t_K = T$, $X_{t_K}^o$ is the solution of equation 1 or 2; otherwise, it is the solution of

$$\sum_{t=t_K}^T (h_t + \pi_t) \bar{P}_t(X_{t_K}) = h_{t_K, T} + a\theta_{t_K}, \text{ or} \quad (5)$$

$$X_{t_K}^o = \min\{X \geq 0 : \sum_{t=t_K}^T (h_t + \pi_t) \bar{P}_t(X) \leq h_{t_K, T} + a\theta_{t_K}\}. \quad (6)$$

Thus, an iterative algorithm can be used to solve the item subproblems. First, compute the unconstrained solution of the current formulation. If no consistency constraints are violated, the current candidate solution is optimal. Otherwise, reformulate the problem eliminating the appropriate production periods and iterate.

Having described the solution to the Lagrangian relaxation, we now describe the planning algorithm itself. At each iteration, the algorithm computes the *Lagrangian solution* and its value based on the current multipliers. If this solution is feasible, the algorithm stops with an optimal solution. Otherwise, if its value is greater than the current lower bound on the optimal value, then the lower bound is updated. Next, two feasible production plans are generated using a heuristic that will be described later. If either of these plans has a lower cost than the current best solution, then the best feasible solution

is updated. If the algorithm has completed M iterations or if the cost of the best feasible solution is within $100 \times \delta\%$ of the current lower bound, it stops and returns the best feasible solution and the percent difference from the current lower bound. Otherwise, a subgradient is computed, the step size and multipliers are updated, and the algorithm proceeds to the next iteration. As in Held, Wolfe, and Crowder [11], we use a step size of

$$t_n = \lambda_n \frac{\hat{Z} - Z_{LR}(\theta^n)}{\|w^n\|^2},$$

where λ_n is a multiplicative factor that is reduced every N iterations, $\hat{Z}$ is the cost of the best feasible solution, $Z_{LR}(\theta^n)$ is the cost of the current Lagrangian solution, and w^n is the current subgradient. The subgradient used at each iteration is

$$w_t^n = \sum_{i=1}^N a_i (X_{it}^n - X_{i,t-1}^n) - \sum_{j=1}^J y_{jt}^n.$$

If $\hat{Z}$ is actually the optimal value of the original problem the algorithm can be shown to converge. However, as mentioned by Held et al. [11], the use of an upper bound seems to work well in practice.

To generate a feasible production plan, we compute production quantities based on a total capacity usage schedule derived from the current Lagrangian solution. One feasible plan is computed using the sum of the capacity variables, $\sum_{j=1}^J y_{jt}$, as the total capacity usage in period t . For the second plan, the total capacity usage in period t is the minimum of the total available capacity, $\sum_{j=1}^J u_{jt}$, and the total capacity required by the production quantities in the Lagrangian solution, $\sum_{i=1}^N a_i (X_{it} - X_{i,t-1})$. To compute the production quantities based on a total capacity usage, C_t , we substitute η_t for $\theta_t - \theta_{t+1}$ in equations 1 and 2 and for $t=1, \dots, T$ search for a value of η_t that gives

$$C_t - \epsilon < \sum_{i=1}^N a_i (X_{it} - X_{i,t-1}) \leq C_t.$$

Note that when this algorithm reaches period t , the production plan for periods $1, \dots, t-1$ is already fixed. When there is only one capacity type, $J = 1$, this heuristic often finds an

optimal solution. Certainly, other heuristics might be used but this one seems to be sufficient for the problem instances that have been tested.

5 Computational Results

In order to evaluate our algorithm, we compared its performance with that of a stochastic dynamic programming solution for a small problem instance. The problem instance has two items each with a holding cost of \$0.50 per period and a backorder penalty cost of \$7.00 per period. Each item has a discrete probability distribution that is detailed in Table 1. There is only a single source of production capacity, and it is expressed in units of finished goods because both items require the same amount of production capacity. In order to reduce the end of horizon effect that causes a finite horizon solution to carry less inventory at the end of the time horizon, the unit purchase and production costs are zero.

We simulated a rolling horizon implementation of each algorithm at different utilizations with different planning horizon lengths. Using the net inventory levels at the beginning of a period, a production plan is computed for the given horizon length, T . The solution for the first period is executed, and then the demands are observed. The simulation then moves to the next time period, a new plan is computed using the same horizon length, T . Again the first period's solution is executed, and demands are observed. Forty independent runs were made for each problem instance. Each run was initialized for 200 periods, and statistics were then collected for 1000 periods. Common random number streams were used across all problem instances and both algorithms. The planning horizon length was 4, 8, 12, or 16. The production capacity was 7, 8, or 9 with long-run average utilization of 90%, 78%, or 70%, respectively.

There are several parameters in the planning algorithm that can be adjusted to improve its performance on a particular problem instance. However, we used the same

parameter settings for each of the problem instances and simulation experiments reported here. The initial value of λ was 1.0, and it was reduced by a factor of 0.95 every $N = 5$ iterations. The algorithm stopped after $M = 500$ iterations or if the cost of the best feasible solution was within 1% of the current lower bound. A binary search procedure was used to find an appropriate value of η_t when computing feasible production plans.

The planning algorithm demonstrated excellent performance in these computational tests. For each problem instance, Table 2 shows the average cost per period for the planning heuristic (H) and the dynamic programming (DP) solutions and the percent difference between them. For the longest planning horizon, $T=16$, the heuristic's average cost is less than 3% above the dynamic program's cost. Although the standard deviations reported in Table 3 may be significant, common random number streams were used for each problem instance so that comparisons are, in fact, meaningful. When generating the production plans, most of the solutions had a duality gap less than the 1% tolerance. However, with 90% utilization and $T=8, 12$, and 16 this tolerance was violated in some cases, but the difference between the cost of a feasible solution and its lower bound was at most 2.4%. Therefore, the planning algorithm performs extremely well from two perspectives. First, it generates excellent solutions to the formulation PIP, and secondly, the average cost of its rolling horizon implementation comes within a few percent of the cost of the implementation of the optimal finite-horizon solution.

We may draw some further conclusions by looking at Tables 2 and 4 more closely. First, we see that as the average capacity utilization increases, the difference between the heuristic and DP performance increases. It appears that at a higher average utilization the additional information available for future production decisions has a more significant effect on the first period's decisions. Second, as the average capacity utilization increases, a longer planning horizon is required to generate good production plans. With higher utilization, more inventory is required to avoid long periods of persistent backorders

	0	1	2	3	4	5	6	7	8	9	Mean
1	0.05	0.06	0.08	0.10	0.10	0.10	0.10	0.10	0.11	0.20	5.40
2	0.82	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.90

Table 1: Demand Distribution for Each Item

Util.	T											
	4			8			12			16		
	H	DP		H	DP		H	DP		H	DP	
90% (7)	22.11	20.86	6.0%	19.37	18.26	6.1%	17.00	16.37	3.8%	16.04	15.59	2.9%
78% (8)	7.41	7.34	1.0%	7.40	7.17	3.2%	7.40	7.17	3.2%	7.38	7.17	2.9%
70% (9)	5.81	5.80	0.2%	5.82	5.80	0.3%	5.82	5.80	0.3%	5.82	5.80	0.3%

Table 2: Mean Cost per Period

because of the queueing effects across time. Plans generated using a planning horizon that is too short will not build sufficient inventory because they assume no need for inventory beyond the end of the horizon. Thus, we see that our results also provide some interesting insights into the rolling horizon use of finite-horizon planning models.

Finally, we conducted some computational tests to demonstrate the algorithm's speed on a realistic problem instance. The instance is composed of twenty items that are

Util.	T							
	4		8		12		16	
	H	DP	H	DP	H	DP	H	DP
90% (7)	7.88	7.85	7.12	7.15	6.69	6.49	6.37	6.17
78% (8)	1.08	1.08	1.08	0.98	1.08	0.98	1.06	0.98
70% (9)	0.34	0.32	0.33	0.32	0.36	0.32	0.35	0.32

Table 3: Standard Deviation of Cost per Period

Util.	T							
	4		8		12		16	
	H	DP	H	DP	H	DP	H	DP
90% (7)	6.53	6.83	8.06	8.35	9.95	9.96	10.87	10.79
78% (8)	8.84	8.91	8.92	9.82	8.98	9.82	9.20	9.82
70% (9)	8.92	8.94	8.88	8.94	8.92	8.94	8.92	8.94

Table 4: Average Total Inventory at Period's End

produced in a single workcenter by a manufacturer of aerospace components. We fit generalized negative binomial distributions to the mean and standard deviation of weekly demand over a year. The weekly holding cost is estimated to be \$0.13 and the backorder cost is estimated to be \$2.47 per week. Long-run average capacity utilization was assumed to be 80%, and a planning horizon of 16 weeks was used. Assuming a uniform distribution between zero and 1.5 times the average weekly demand for each item, we generated 40 random vectors of initial inventory levels. The algorithm parameters were the same as those reported above except that the initial value of λ was 2.0, and it was decreased by a factor of 0.5 every 10 iterations. The average number of iterations of the subgradient algorithm was 27.9 with standard deviation 3.75. The algorithm required an average of 36.5 seconds of CPU time with a standard deviation of 2.78 on a Sun Sparc Station IPX. Although further refinements and computational tests are needed, these results show that the algorithm has great potential for solving practical problem instances.

6 Conclusions

We have described an approximate formulation and solution algorithm for the finite-horizon capacitated production planning problem with random demand for multiple

products. Using Lagrangian relaxation, we developed a subgradient optimization algorithm to solve this formulation and provided some computational results that indicate this approach works very well when used for rolling-horizon planning. The advantage of our approach is that realistic problem instances can be solved quite quickly while optimal solutions to these problem instances are computationally intractable.

There are many avenues for further extensions of this work. More computational work should be done to determine the effect of a greater number of products and production resources on the convergence of the solution algorithm. Also, good lower bounds for realistic problem instances are needed to evaluate performance. Knowledge of firm customer orders over some part of the planning horizon can be modeled and an appropriate solution algorithm developed. Extensions that incorporate setup times and cost are also needed. This modeling and solution approach can also be extended to multi-echelon distribution and assembly structures. Each of these extensions drives toward a general methodology for finite-horizon, multi-item, production planning that explicitly considers randomness and its effects on production planning.

7 Appendix

Proof of Theorem 1

Proof: Clearly, the objective function of $SP(\theta)$, $f(X)$, is separable in the decision variables and may be written as

$$f(X) = \sum_{t=1}^T f_t(X_t),$$

where

$$\begin{aligned} f_t(X_t) &= L_t(X_t) + a(\theta_t - \theta_{t+1})X_t, \text{ for } t=1, \dots, T-1 \text{ and} \\ f_T(X_T) &= L_T(X_T) + a\theta_T X_T. \end{aligned}$$

First, we present the proof of part a.

a) Note that $\frac{\partial f}{\partial X_t}|_X = \frac{\partial f_t}{\partial X_t}|_{X_t}$, and that

$$\begin{aligned} L_t(X_t) &= h_t \int_0^{X_t} (X_t - x)p_t(x)dx + \pi_t \int_{X_t}^{\infty} (x - X_t)p_t(x)dx \\ &= h_t(X_t - \mu_t) - h_t \int_{X_t}^{\infty} (X_t - x)p_t(x)dx \\ &\quad + \pi_t \int_{X_t}^{\infty} (x - X_t)p_t(x)dx \\ &= h_t(X_t - \mu_t) + (h_t + \pi_t) \int_{X_t}^{\infty} (x - X_t)p_t(x)dx, \end{aligned}$$

where μ_t is the expected demand for this product in period t . Note that the terms in $f(X)$ corresponding to the Lagrangian capacity costs are linear. Thus, we merely need to show the convexity of the L_t 's to prove $f(X)$ is convex. We see that the first derivative of L_t is

$$\begin{aligned} \frac{\partial L_t}{\partial X_t} &= h_t + (h_t + \pi_t) \frac{\partial}{\partial X_t} \int_{X_t}^{\infty} (x - X_t)p_t(x)dx \\ &= h_t - (h_t + \pi_t) \int_{X_t}^{\infty} p_t(x)dx \\ &= h_t - (h_t + \pi_t) \bar{P}_t(X_t). \end{aligned}$$

Thus, the second derivative is

$$\frac{\partial^2 L_t}{\partial X_t^2} = (h_t + \pi_t)p_t(X_t).$$

Therefore, $f_t(X)$ is strictly convex if $p_t(0) = 0$ and $p_t(X_t) > 0$ for $X_t > 0$. Therefore, the X which solves $\nabla f(X) = 0$ is the global optimum of $f(X)$. For the last time period, the solution of

$$\frac{\partial f_T}{\partial X_T} = 0$$

gives us

$$h_T - (h_T + \pi_T)\bar{P}_T(X_T) + a\theta_T = 0,$$

and

$$\bar{P}_T(X_T) = \frac{h_T + a\theta_T}{h_T + \pi_T}.$$

For $t=1, \dots, T-1$, the solution of

$$\frac{\partial f_t}{\partial X_t} = 0$$

gives us

$$h_t - (h_t + \pi_t)\bar{P}_t(X_t) + a(\theta_t - \theta_{t+1}) = 0,$$

and

$$\bar{P}_t(X_t) = \frac{h_t + a(\theta_t - \theta_{t+1})}{h_t + \pi_t}.$$

b) When demand is described by a discrete distribution, the holding and backorder cost function can be written as:

$$\begin{aligned} L_t(X_t) &= h_t \sum_{x=0}^{X_t} (X_t - x)p_t(x) + \pi_t \sum_{x=X_t}^{\infty} (x - X_t)p_t(x) \\ &= h_t(X_t - \mu_t) - h_t \sum_{x=X_t}^{\infty} (X_t - x)p_t(x) + \pi_t \sum_{x=X_t}^{\infty} (x - X_t)p_t(x) \\ &= h_t(X_t - \mu_t) + (h_t + \pi_t) \sum_{x=X_t}^{\infty} (x - X_t)p_t(x), \end{aligned}$$

where μ_t is the expected demand for this product in period t . Let

$$\Delta h(x) = h(x+1) - h(x).$$

Now,

$$\begin{aligned}
\Delta L_t(X_t) &= h_t + (h_t + \pi_t) \left[\sum_{x=X_t+1}^{\infty} (x - X_t - 1)p_t(x) - \sum_{x=X_t}^{\infty} (x - X_t)p_t(x) \right] \\
&= h_t + (h_t + \pi_t) \left[- \sum_{x=X_t+1}^{\infty} p_t(x) \right] \\
&= h_t - (h_t + \pi_t)\bar{P}_t(X_t);
\end{aligned}$$

so,

$$\Delta^2 L_t(X_t) = (h_t + \pi_t)p_t(X_t + 1).$$

Therefore, $f_t(X_t)$ is strictly convex if $p_t(X_t) > 0$ for $X_t \in \{1, 2, \dots\}$, and $\min_{X_t \geq 0} f_t(X_t)$ is attained by $X_t^0 = \min\{X_t \geq 0 : \Delta f_t(X_t) \geq 0\}$. However, $\Delta f_t(X_t) \geq 0$ is equivalent to

$$h_t - (h_t + \pi_t)\bar{P}_t(X_t) + a(\theta_t - \theta_{t+1}) \geq 0,$$

and solving for $\bar{P}_t(X_t)$ we see that X_t^0 is defined by 2.

□

Discussion of Special Cases Related to Theorem 1

Without loss of generality, assume the demand process has a continuous distribution with support on $[0, M]$ and that $t < T$. Recall that

$$\begin{aligned}
\frac{\partial f}{\partial X_t} &= h_t + a(\theta_t - \theta_{t+1}) - (h_t + \pi_t)\bar{P}_t(X_t), \text{ and} \\
\gamma_t &= \frac{h_t + a(\theta_t - \theta_{t+1})}{h_t + \pi_t}.
\end{aligned}$$

We now describe the unconstrained solution for X_t for each of the four special cases.

1. $\gamma_t > 1 \Rightarrow h_t + a(\theta_t - \theta_{t+1}) > h_t + \pi_t > (h_t + \pi_t)\bar{P}_t(X_t), \forall X_t$. Therefore, $\frac{\partial f}{\partial X_t} > 0, \forall X_t$, and any unconstrained optimal solution has $X_t = -\infty$.
2. $\gamma_t = 1 \Rightarrow h_t + a(\theta_t - \theta_{t+1}) = h_t + \pi_t = (h_t + \pi_t)\bar{P}_t(X_t), \forall X_t \leq 0$. Therefore, $\frac{\partial f}{\partial X_t} = 0, \forall X_t \leq 0$, and any unconstrained optimal solution has $X_t \leq 0$. However, we assume that all backorders have been filled which implies that $X_t = 0$ in an optimal solution.

3. $\gamma_t = 0 \Rightarrow h_t + a(\theta_t - \theta_{t+1}) = 0$. Therefore, $\frac{\partial f}{\partial X_t} = 0, \forall X_t \geq M$, and any unconstrained optimal solution has $X_t \geq M$. In order to generate good feasible solutions, any remaining available production capacity is allocated to items in this case.
4. $\gamma_t < 0 \Rightarrow h_t + a(\theta_t - \theta_{t+1}) < 0$. Therefore, $\frac{\partial f}{\partial X_t} < 0, \forall X_t$, and any unconstrained optimal solution has $X_t = +\infty$, and the optimal value is $f(X^*) = -\infty$.

Proof of Theorem 2

Proof: For both continuous and discrete demand, we prove the theorem from first principles using the Kuhn-Tucker conditions. For continuous demand, note that the Kuhn-Tucker conditions for optimality for $SP(\theta)$ are

$$\frac{\partial f}{\partial X_T}|_{X^*} - \theta_T = 0, \quad (7)$$

$$\frac{\partial f}{\partial X_t}|_{X^*} - \theta_t + \theta_{t+1} = 0, \text{ for } t=1, \dots, T-1, \quad (8)$$

$$\theta_t(X_{t-1}^* - X_t^*) = 0, \text{ for } t=1, \dots, T, \text{ and} \quad (9)$$

$$\theta_t \geq 0, \forall t. \quad (10)$$

Since $f(X)$ is separable, $\frac{\partial f}{\partial X_t} = \frac{\partial f_t}{\partial X_t}$. Suppose $X_{t-1}^* < X_t^*$. Then by equation 9, $\theta_t = 0$ which implies that $\frac{\partial f_t}{\partial X_t}|_{X_t^*} = 0$ when $t=T$, and $\frac{\partial f_t}{\partial X_t}|_{X_t^*} = -\theta_{t+1}$ when $t=1, \dots, T-1$. Thus, condition 10 implies $\frac{\partial f_t}{\partial X_t}|_{X_t^*} \leq 0$ for any t . Because $f_t(X)$ is convex and $\frac{\partial f_t}{\partial X_t}|_{X_t^0} = 0$ and $\frac{\partial f_t}{\partial X_t}|_{X_t^*} \leq 0$, we have that $X_t^* \leq X_t^0$.

Similarly, for $t=2, \dots, T$ condition 8 implies $\frac{\partial f_{t-1}}{\partial X_{t-1}}|_{X_{t-1}^*} = \theta_{t-1} \geq 0$, and for $t=1$, $X_{t-1} = X_0$, a constant. Thus, $\frac{\partial f_{t-1}}{\partial X_{t-1}}|_{X_{t-1}^*} \geq 0$ for any t for which $X_{t-1}^* < X_t^*$. Because $f_{t-1}(X)$ is convex and $\frac{\partial f_{t-1}}{\partial X_{t-1}}|_{X_{t-1}^0} = 0$ and $\frac{\partial f_{t-1}}{\partial X_{t-1}}|_{X_{t-1}^*} \geq 0$, we have that $X_{t-1}^0 \leq X_{t-1}^*$. Therefore, $X_t^* \leq X_t^0 < X_{t-1}^0 \leq X_{t-1}^*$ implying that $X_t^* < X_{t-1}^*$, which contradicts the hypothesis.

For a discrete demand distribution, note the optimality conditions for the mixed

integer nonlinear program $SP(\theta)$:

$$X_T^* = \min\{X_T \geq 0 : \Delta f_T(X_T) - \theta_T \geq 0\} \quad (11)$$

$$X_t^* = \min\{X_t \geq 0 : \Delta f_t(X_t) - \theta_t + \theta_{t+1} \geq 0\}, \text{ for } t=1, \dots, T-1 \quad (12)$$

$$\theta_t(X_{t-1}^* - X_t^*) = 0, \text{ for } t=1, \dots, T \quad (13)$$

$$\theta_t \geq 0, \forall t. \quad (14)$$

Suppose $X_{t-1}^* < X_t^*$. Then by equation 13, $\theta_t = 0$. So, 11 and 12 imply that

$$X_T^* = \min\{X_T \geq 0 : \Delta f_T(X_T^*) \geq 0\} \text{ when } t=T, \text{ and } X_t^* = \min\{X_t \geq 0 : \Delta f_t(X_t^*) \geq -\theta_{t+1}\}$$

when $t=1, \dots, T-1$. Because $f_t(X)$ is convex and $X_t^0 = \min\{X_t \geq 0 : \Delta f_t(X_t) \geq 0\}$ and

$\theta_{t+1} \geq 0$, we have that $X_t^* \leq X_t^0$. Similarly, for $t=2, \dots, T$, condition 12 implies

$$X_{t-1}^* = \min\{X_{t-1} \geq 0 : \Delta f_{t-1}(X_{t-1}) \geq \theta_{t-1}\}, \text{ and for } t=1, X_{t-1}^* = X_0 = X_0^0, \text{ a constant.}$$

Because $f_{t-1}(X)$ is convex and $X_{t-1}^0 = \min\{X_{t-1} : \Delta f_{t-1}(X_{t-1}) \geq 0\}$ and $\theta_{t-1} \geq 0$, we have

that $X_{t-1}^0 \leq X_{t-1}^*$. Therefore, $X_t^* \leq X_t^0 < X_{t-1}^0 \leq X_{t-1}^*$ which implies that $X_t^* < X_{t-1}^*$,

which contradicts the hypothesis. $\square$

References

- [1] Baker, Kenneth R., "An Experimental Study of the Effectiveness of Rolling Schedules in Production Planning," *Decision Sciences*, 8, 19–27 (1977).
- [2] Bitran, Gabriel R., and Yanasse, Horacio H., "Deterministic Approximations to Stochastic Production Problems," *Operations Research*, 32, 999–1018 (1985).
- [3] Elmaghraby, S.E., "The Economic Lot Scheduling Problem (ELSP): Review and Extensions," *Management Science*, 24, 587 – 598 (1978).
- [4] Evans, Richard V., "Inventory Control of a Multi-Product System with a Limited Production Resource," *NRLQ*, 14, 2, 173–184 (1967).
- [5] Everett, Hugh, "Generalized Lagrange Multiplier Method for Solving Problems of Optimum Allocation of Resources," *Operations Research*, 11, 399–417 (1966).
- [6] Fisher, Marshall L., "Lagrangian Relaxation Method for Solving Integer Programming Problems," *Management Science*, 27, 1–18 (1981).
- [7] Gascon, André, "The Lookahead Heuristic for Multi-Item Single Machine Production Scheduling with Dynamic, Stochastic Demands," *INFOR*, 26, 114 – 126 (1987).
- [8] Gascon, André, and Leachman, Robert C., "A Dynamic Programming Solution to the Dynamic, Multi-Item, Single-Machine Scheduling Problem," *Operations Research*, 36, 50 – 56 (1988).
- [9] Graves, Stephen C., "The Multi-Product Production Cycling Problem," *AIIE Transactions*, 12, 233 – 240 (1980).

- [10] Haneveld, Willem K. Klein, "A Stochastic Programming Approach to Multiperiod Production Planning," Research Memorandum 276, University of Groningen, Institute of Economic Research (1988).
- [11] Held, Michael, Wolfe, Philip, and Crowder, Harlan P., "Validation of Subgradient Optimization," *Mathematical Programming*, 6, 62–88 (1974).
- [12] Khang, Do Ba, and Fujiwara, Okitsugu, "Multi-Period Stochastic Network Flow Problems with Service Level Requirements: Approximation and Error Estimates," *IIE Transactions*, 25, 104–110 (1993).
- [13] Leachman, Robert C., and Gascon, André, "A Heuristic Scheduling Policy for Multi-Item, Single-Machine Production Systems with Time-Varying, Stochastic Demands," *Management Science*, 34, 377 – 390 (1988).
- [14] Nemhauser, G. L., and Wolsey, L. A., *Integer and Combinatorial Optimization*, Wiley, New York (1988).
- [15] Peters, Robert J., Boskma, Klaas, and Kupper, Hendrik A.E., "Stochastic Programming in Production Planning: A Case with Non-Simple Recourse," *Statistica Neerlandica*, 31, 113 – 126 (1977).
- [16] Williams, Jack F., "A Hybrid Algorithm for Simultaneous Scheduling of Production and Distribution in Multi-Echelon Structures," *Management Science*, 29, 77 – 92 (1983).