Queueing Control Problems for Production/Inventory Systems

by

Michael Howard Veatch

B.A. Mathematics and Physics, Whitman College (1979)
M.S. Oper. Res. and Stat., Rensselaer Polytechnic Inst. (1980)

Submitted to the Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 1992

Author ....
Sloan School of Management
August 1992

Certified by ....
Lawrence M. Wein
Associate Professor of Management Science
Thesis Supervisor

Accepted by ....
Richard C. Larson
Codirector, Operations Research Center

Queueing Control Problems for Production/Inventory Systems

by

Michael Howard Veatch

Submitted to the Sloan School of Management
on August 7, 1992, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Operations Research

Abstract

Two manufacturing facilities are considered that operate in a make-to-stock mode: after production, items are placed in a finished goods inventory that services an exogenous demand. The problem of controlling inventory to minimize holding and backorder or lost sales costs is investigated. To capture the effects of uncertainty and capacity constraints, a Markovian queueing model and dynamic programming are used. The first system consists of two stations in tandem that produces a single product. A general monotonicity theory is formulated, using submodularity, that characterizes the optimal control policies when viewed in the state space. Conditions are found for certain simple controls to be optimal. Optimal controls are computed and compared with kanban and buffer control mechanisms, popular in manufacturing, and with the base stock control mechanism popular in inventory/distribution systems.

The second system consists of one machine that produces several products. Computationally tractable index policies are developed that improve significantly upon the smallest inventory and static priority "$c\mu$" rules found in the literature. One of the indices used is Whittle's "restless bandit" index, which possesses a certain asymptotic optimality. To improve their performance, index policies are combined with other approximations of when to idle. The best idleness policy is obtained from a diffusion approximation, derived from a heavy traffic limit for the backorder case and a similar limit (demand balanced with capacity and large lost sales costs) for the lost sales case.

Thesis Supervisor: Lawrence M. Wein
Title: Associate Professor of Management Science

Acknowlegements

I would like to thank Larry Wein and Dimitris Bertsimas for their continual encouragement and interest in my research. Its timely completion would not have been possible without the loving support and understanding of my wife Cindy. This research was supported by a grant from the National Science Foundation. It is dedicated to the glory of God.

Note to the Reader

All section, equation, definition, and theorem numbers refer to the current chapter unless otherwise noted. All notation is defined within a chapter, and usage varies slightly across chapters.

Contents

Chapter 1

Introduction and Summary

1.1 The Role of Inventory in Manufacturing

The control of inventory is a concern of the first order when managing a manufacturing facility. Inventory raises a myriad of conflicting considerations: the costliness of holding inventory; its remarkable ability to pile up even in facilities that were thought to be balanced; the material handling, process disruption, and quality control problems it can cause; and its essential role in decoupling asynchronous processes from each other, to increase throughput, or from customer demand, to improve fill rates. A number of frameworks have been developed to assist in the efficient management of inventory.

The most popular framework is material requirements planning (MRP), or more generally, manufacturing resource planning (MRP II). In this essentially deterministic model, future requirements are traced back through the system using fixed lead times to identify the time at which each product or component part must be released into the system and begin processing. The inventory in the system due to these lead times is called pipeline stock. Additional inventory, called cycle stock, may be added due batch processing, with batch sizes determined by an economic lot-sizing analysis that trades off set-up costs (or times) and the costs of holding inventories. In queueing notation, this framework is $D/D/\infty$.

A second level of complexity is to introduce randomness but maintain fixed lead times ($G/D/\infty$), or at least lead times that do not depend on the congestion of the system ($G/G/\infty$). There may be requirements uncertainty, i.e., imperfect forecasts, processing time or yield uncertainty, or uncertainty in the supply of inputs. Uncertainty creates a need for additional inventory, termed safety stock. Requirements uncertainty is typically dealt with in inventory systems by holding more finished goods (FG) inventory. Production uncertainty is dealt with by holding work-in-process (WIP) inventory, providing a buffer between the stages of production to reduce disruptions and increase throughput. Once safety stock levels are determined, they can be used to adjust the net requirements or lead times in an MRP system.

The just-in-time philosophy has articulated the disadvantages of holding WIP. Increased lead times hurt timely quality control and responsiveness to customized orders or process changes. FG will also be reduced if lead time can be reduced. Some production may even be changed from a make-to-stock mode, where FG inventory is held, to a make-to-order mode, where production is not initiated until an order is received. On the other hand, time competitiveness is moving some industries toward make-to-stock. In a make-to-stock mode, WIP not only serves as a buffer between production stages, it can also supplement FG to reduce backorders.

1.2 Queueing Models of Production/Inventory Systems

The above frameworks miss a key aspect of safety stock: capacity constraints limit the ability to respond to unexpected requirements. Adding capacity constraints replaces fixed lead times with sojourn times in a queueing system, which depend upon congestion. The queueing framework combines the stochastic nature of the problem, often studied in inventory systems, with the capacity constraints, usually dealt with through deterministic production planning. It provides a more realistic model of the entire production/inventory system. The difficulty of solving queueing control problems has limited the amount of research in this area. Single-product sequential facilities in make-to-order mode, corresponding to a tandem queues with finite buffers, have been studied extensively. Throughput capacity of these systems, operating in an unlimited demand mode, has also been studied. Much less research has been done on the make-to-stock mode (see Chapter 3 for a literature review). A multi-product, single-station, make-to-stock system has been studied using heavy traffic diffusion approximation; see Chapter 4.

Capacity constraints are also addressed by the linear production rule of Graves (1988) without explicitly considering queueing. Processing rates are assumed to be proportional to the backlog of work at a station and stock levels (which smooth production) are set so that the probability of exceeding the capacity constraint is small.

We focus on two queueing models of production/inventory systems. Both assume that demand is Poisson, service times are exponential ($M/M/1$), and linear holding and backorder costs are incurred. There are no set-up costs or times when starting production or switching products. The make-to-stock mode is allowed. The first model, analyzed in Chapter 3, has two machines in tandem. After production, items are placed in a FG inventory that services the exogenous demand. Demand that cannot be met from inventory is backordered. The objective is to control production at each machine to minimize the discounted or long-run average cost of holding WIP, holding FG, and backorders.

The second model, analyzed in Chapter 4, has one machine that can produce a variety of product classes. A FG inventory is kept for each class. The objective is to select a class to produce, or idle, at each time to minimize holding and backorder costs. A lost sales version is also considered, where demands that occur during a stockout are ignored and a cost incurred. Preemption is allowed, so that the machine can switch classes at any time.

1.3 Methodological Contributions

Several of the methods used in this study of production/inventory systems have broader applicability. The first, a monotonicity theory for optimal controls, is developed in Chapter 2. As in Weber and Stidham (1987), submodularity is used to show that the optimal control rates for Markovian queueing networks are monotonic as one moves in certain directions in the state space. Unlike Weber and Stidham's result, we consider directions that are not control directions, arrival routing problems, and some uncontrolled service rates. To establish submodularity, and hence, monotonicity, the cost rate must satisfy a submodularity condition on the set of directions chosen, the boundaries of the state space must satisfy a lattice condition with respect to these directions, and uncontrolled transitions must satisfy a submodularity-preserving condition. Our method could be used to establish monotonicity for many additional problems; two other applications are given in Chapter 2 Section 6. Monotonicity is a useful characterization of optimal controls. For problems such as ours, with no control costs, monotonicity implies that there is a switching curve for each control, dividing the state space into regions where the control is on and off, and the slope of the curve must lie between certain limits. Monotonicity can help in many problems to narrow the scope of policies that must be considered when constructing a heuristic solution.

In Chapter 3 Section 3, a new method is used to verify optimality for a dynamic program on a two-dimensional state space. The usual method of verifying optimality is to compute the value function for a given policy; unfortunately, this requires solving a large linear system and usually cannot be done analytically. Submodularity/monotonicity is one tool to help verify the optimality of a policy. We apply stochastic coupling arguments to bound the difference in the value function at points appearing in the dynamic programming equations. The bound is obtained by computing the relative cost until merging for two coupled processes with different initial states. If the bound determines in which of two states the value function is greater, it can be used to check optimality. Although considerable effort is required to construct each stochastic coupling argument, the technique has broad applicability. For example, it could be used to establish conditions under which the static priority $b\mu$ and $h\mu$ rule of Chapter 4 Section 1 is optimal. In general, stochastic coupling may be a feasible approach whenever the policy being considered has a very simple structure, such as a static priority.

In Chapter 4 Section 3, a successful application of Whittle's (1988) "restless bandit" index is given. In this generalization of the multi-armed bandit problem, arms that are not being played can change state and incur costs. Unlike the Gittens index for the multi-armed bandit, the restless bandit index is not optimal, but it is asymptotically optimal as the number of arms becomes large. We compute the restless bandit index, demonstrate that it is numerically useful, and show that it can be thought of as one of many possible generalized Gittens indices with a variable retirement reward. Viewing the restless bandit index as a generalized Gittens index may provide insights for other restless bandit problems. It opens up the possibility of constructing even better indices for these problems.

In Chapter 4 Section 4, we make a contribution to the heavy traffic scheduling literature by proving that a state-space collapse to a one-dimensional problem can be used even when the one-dimensional problem is a relaxation. The problem of modeling controlled queueing networks with finite arrival buffers within the Brownian framework is also addressed, building on the work of Dai and Harrison (1991). The rejection of arrivals in a queueing network corresponds to lost sales in the production/inventory system.

1.4 Tandem System Results

Considerable insights are gained into the form of the optimal policy for the two-station tandem system in Chapter 3. First, simple policies are shown to be optimal under extreme conditions on the problem parameters. Conditions are found under which no inventory is held, as well as conditions under which all inventory is converted to FG, i.e., the downstream station never idles unless it is starved. Interestingly, the popular base stock policy is shown to never be *exactly* optimal.

Second, numerical results are obtained using dynamic programming. The decision of where to place inventory is shown to depend on the relative holding costs and the rate at which WIP can be converted into FG. Less WIP is held when its holding cost is high, the utilization of the upstream station is low, or the discount rate is high. Optimal policies are also compared to the best base stock, kanban, and fixed buffer policies. It is found that base stock policies are nearly optimal when the upstream station is heavily utilized and the discount rate is small or zero. Kanban policies outperform base stock policies when the downstream station is a bottleneck or discounting is significant. It is encouraging that base stock and kanban policies are within a few percent of optimal for most test cases, since a switching curve policy would be more difficult to implement. Every type of policy is sensitive to the stock levels or buffer sizes, so that obtaining accurate demand and production rate data and setting these levels correctly remains a very important issue.

1.5 Multiclass System Results

Approximate policies for the single-stage, multiclass system are developed and tested in Chapter 4. Policies are represented by switching surfaces in the state space, indicating the preference between classes at each inventory position, and hedging points that indicate when to idle. Switching surfaces are found by computing an index for each class as a function of the inventory for that class; the class with the smallest index is preferred. Hedging points are found using several methods that relate the problem to a single-product subproblem. For the backorder problem, the best heuristic is a service time look-ahead index from Zipkin (1990), combined with a static priority $b\mu$ rule when there are backorders, and a Brownian motion hedging point. This policy is close to optimal for two-product test cases. It is probably more robust over different parameter values than the simple lowest inventory level policy. For the lost sales problem, the best heuristic is the service time look-ahead or restless bandit index combined with the Brownian motion hedging point. This policy is close to optimal for two- and three-product test cases. It is also significantly better than a lowest inventory level policy, suggesting that implementation of index policies should be considered for this type of system.

Chapter 2

Monotone Control of Queueing Networks

2.1 Introduction

The optimal control of arrival and service rates in networks of queues has been the topic of considerable research (see Stidham 1985, 1988 and the references therein). Most studies have dealt with specific network structures or systems with only two queues. Even in these special cases and with memoryless arrival and service processes, one usually must resort to numerical dynamic programming techniques to compute the optimal control. Alternatively, a certain form of control can be analyzed; examples include end-to-end controls for communication systems, and kanban and base stock controls for manufacturing systems.

Another fruitful line of research has been to study the *structure* of optimal policies. In many cases, the optimal policy has been shown to have intuitive monotonicity properties (see, for example, Rosberg, Varaiya, and Walrand 1982, Beutler and Teneketzis 1989, and Hajek 1984). These proofs all use submodularity and are problem-specific. Weber and Stidham (1987) use submodularity to prove *transition monotonicity*: service completion at one station cannot reduce the optimal service rate at another station. Their result applies to a class of network topologies, which are assumed to be Markovian with all service rates controllable, controllable and uncontrollable arrival rates, and discounted or long-run average costs. In a series system with service rate control, service costs and convex holding costs, for example, they show that upon the addition of a customer to a queue, the optimal service rates at that queue and at downstream queues do not decrease and the optimal service rates upstream do not increase.

This paper generalizes the monotonicity results of Weber and Stidham (1987) in several straightforward but useful ways. First, they only consider monotonicity with respect to control directions in the state space. Although other directions can often be addressed by adding fictitious controls, we take the straightforward approach of considering arbitrary directions; the production/inventory model that motivated our research uses this generalization. Second, arrival routing problems, such as Hajek's (1984) two interacting queues, are addressed using controls that choose between two transitions. Finally, they treat uncontrolled transitions that (i) are state-independent, i.e., arrivals or (ii) can be replaced by controlled transitions for which the maximum rate is always optimal (see Weber and Stidham 1987 p.217). We consider other uncontrolled transitions, including the service rates in Hajek (1984) and the downstream server in a series queue.

The following assumptions are made to establish monotonicity:

1. Holding costs satisfy a submodularity condition (related to convexity),

2. The boundaries of the state space satisfy a geometric condition: only one control (or other vector used as a monotone direction) can cross each boundary in each direction,

3. Controls that choose between two directions must have one direction that never leaves the state space, and

4. Service and arrival rates are controllable to zero, state-independent, or satisfy an additional condition (preserving submodularity).

The theory is powerful in the sense that it can be used to easily prove monotonicity for a variety of queueing control problems. We have used it to reproduce some of the two-station monotonicity results of Rosberg, Varaiya, and Walrand (1982) and Hajek (1984).

Another extension of transition monotonicity is the recent paper by Glasserman and Yao (1992). They replace (2), which is a lattice condition on the state space, with a weaker join semi-lattice condition plus some other conditions on the holding costs, use *super-* and submodularity, and consider controls that choose between two directions or idleness. Their results could potentially be applied to routing decisions within a network or to server allocation (scheduling) problems, where our theorem does not apply. However, neither their result nor ours reproduces the monotonicity proven by Hajek for his allocation problem.

The monotonicity theorem is applied to several problems (some of which can be treated using Weber and Stidham 1987). Of particular interest is the following queueing network that arises in a manufacturing setting. Items pass through a sequence of exponential single-server queues, each of which represents a stage of production. Upon exiting the final stage, items are placed in finished goods inventory that is used to meet a Poisson demand. Demand that cannot be met from inventory is backordered and met by the next available finished item. Production and holding costs are incurred at each stage, as well as finished goods backorder costs. The problem is to control the production rates to minimize discounted or average costs. This system is known as a *make-to-stock* queueing system, or in the manufacturing literature as a production/inventory system (see, for example, Buzacott, Price, and Shanthikumar 1991). For a series system, we show that upon the addition of an item between stages, the production rates downstream do not decrease and the production rates upstream do not increase; also, a demand does not decrease production rates.

We also give some simplifying characterizations of the optimal policy under transition monotonicity. It is well-known that, in the absence of service costs, these problems have all-or-nothing (bang-bang) optimal policies. We show that they are characterized by switching functions with certain slope restrictions that bound a region in which all servers work at full capacity. When the system state reaches a boundary, one or more servers turn off. This simplification should ease the task of numerically computing optimal policies. Switching functions are also discussed in Hajek (1984), Stidham (1985,1988), and Chen, Yang, and Yao (1991) for specific two-station networks. Developing numerical techniques that exploited these switching functions to compute optimal policies would be a useful area for further research.

We define a general queueing control problem in Section 2, then present monotonicity theorems in Section 3. Transition-monotone policies are further characterized in Section 4, some generalizations covered in Section 5, and applications to make-to-stock queueing networks and other examples discussed in Section 6. We will use the term increasing (decreasing) in the weaker sense of nondecreasing (nonincreasing) and denote a unit vector whose ith component is one by e_i.

2.2 Problem Description

We are primarily concerned with arrival rate, service rate, and routing control problems for queueing networks. However, the method being used to obtain monotonicity results applies to a broader class of Markov decision processes, so we will follow Weber and Stidham and use a more general notation. To accommodate certain routing problems, we define controls that choose between two transitions. If, instead, the rate of each transition is controlled independently, set $d_i^0 = 0$ in what follows. Consider a continuous time Markov decision process with state $x = (x_1, \ldots, x_m) \in X \subseteq Z^m$. The decision space is $\{\mu = (\mu_1, \ldots, \mu_q) : 0 \le \mu_i \le \overline{\mu}_i\}$ and, given a control μ in state x, the transition $x \rightarrow x + d_i^1$ occurs at rate μ_i while the transition $x \rightarrow x + d_i^0$ occurs at rate $\overline{\mu}_i - \mu_i$, for $i \in \mathcal{C} = \{1, \ldots, q\}$. Let $d_i = d_i^1 - d_i^0$; the control μ_i pushes in the direction d_i. Because the system is memoryless, an optimal control need only depend on the current state x; we denote this control $\mu(x)$. Also, if $x + d_i^1 \notin X$, then $\mu_i(x) = 0$.

Let $d_{q+1}, \ldots, d_{q+p}$ be the uncontrolled transitions with index set $\mathcal{U} = \{q+1, \ldots, q+p\}$. To prevent transitions out of the state space, define the transition function $D_i x = x + d_i$ if $x + d_i \in X$ and $D_i x = x$ otherwise; D_i occurs at the constant rate λ_i.

The cost rate is a separable, nonnegative function of the state,

$$h(x) = \sum_{k=1}^{m} h_k(x_k),$$

with certain restrictions defined later, plus a continuous, separable function of the control,

$$c(\mu) = \sum_{i \in \mathcal{C}} c_i(\mu_i).$$

The functions c_i are assumed to be convex; if they are not, an equivalent optimization problem results when c_i is replaced by its convex lower envelope on $[0, \overline{\mu}_i]$. The objective is to minimize the expected discounted cost. An average cost criterion will be considered in Section 5.

In a queueing network control problem, x_i is the queue length (including customers in service) at queue i; $X = Z_+^m$; $\mathcal{C}$ consists of a customer moving from queue j to queue k ($d_i^1 = e_k - e_j, d_i^0 = 0$), a customer departing the network from queue j ($d_i^1 = -e_j, d_i^0 = 0$), controlled arrivals at queue j ($d_i^1 = e_j, d_i^0 = 0$), and routing arrivals to queue j or k ($d_i^1 = e_k, d_i^0 = e_j$); μ_i is a service, arrival, or routing rate control; $\mathcal{U}$ may contain uncontrolled arrivals or departures; $h_k(x_k)$ is the cost of holding x_k customers at queue k; and $c_i(\mu_i)$ is the cost of service (or arrivals) at rate μ_i.

In make-to-stock queueing networks, one of the state variables, say x_n, measures finished goods inventory and is allowed to be negative, representing backorders. The uncontrolled transitions are demands which decrease the finished goods inventory. The controls μ_i are arrival and service rates with costs c_i; the costs h_k are holding costs at each station $k \neq n$ and h_n is the finished goods holding and backorder cost. A complete description is given in Section 6.

For the infinite-horizon problem with discount rate $\alpha > 0$, the minimum expected cost is

$$V(x) = E_x \int_0^\infty e^{-\alpha t}[h(x(t)) + c(\mu(t))]dt,$$

where E_x denotes expectation given the initial state $x(0) = x$. We will uniformize the process as in Lippman (1975) by defining the potential event rate

$$\Lambda = \sum_{i \in \mathcal{C}} \overline{\mu}_i + \sum_{i \in \mathcal{U}} \lambda_i.$$

Let $V_n(x)$ be the minimum n-stage expected discounted cost for the embedded discrete-time Markov decision process. Then V_n is well-defined and satisfies the dynamic programming equation

$$\begin{aligned}(\alpha + \Lambda)V_{n+1}(x) &= h(x) + \sum_{i \in \mathcal{U}} \lambda_i V_n(D_i x) \\ &\quad + \sum_{i \in \mathcal{C}} \min_{0 \leq \mu_i \leq \overline{\mu}_i} \{c_i(\mu_i) + \mu_i V_n(x + d_i^1) + (\overline{\mu}_i - \mu_i) V_n(x + d_i^0)\}, \quad (2.1)\end{aligned}$$

where we define $V_0(x) = 0$ and $V_n(x) = \infty$, $x \notin X$.

2.3 General Monotonicity Results

Monotonicity results will be obtained by investigating the effect on the optimal control of moving in certain directions in the state space.

Definition. The control $\mu(x)$ is **D**-*monotone* if

$$\mu_j(x) \leq \mu_j(x + d_i)$$

for all controls $j \in \mathcal{C}$ and direction vectors $d_i \in \mathbf{D}$ such that $d_i \neq d_j$.

We will only consider direction sets **D** that contain all of the control directions $d_i, i \in \mathcal{C}$. If $\mathbf{D} = \{d_i : i \in \mathcal{C}\}$ and $d_i^0 = 0$ (i.e., **D** is the set of control transitions), we will say that $\mu(x)$ is *control-monotone* (this is the case considered in Weber and Stidham 1987); if $\mathbf{D} = \{d_i : i \in \mathcal{C} \cup \mathcal{U}\}$ and $d_i^0 = 0$ (i.e., **D** is the set of all transitions), we will say that $\mu(x)$ is *transition-monotone.* For a queueing system with uncontrolled arrivals and controlled servers, control monotonicity means that after a service completion the service rate of other servers increases; transition monotonicity means that the service rate of all servers also increases after an arrival.

Remark. Under transition monotonicity, μ_j is increasing in every other transition. If $\{x(t), t \geq 0\}$ is an irreducible Markov chain under some control $\mu(x)$, then we can write $-d_j = \sum_{i=i_1}^{i_n} d_i$ for some sequence of transitions not including d_i; hence, μ_j is decreasing in d_j. This monotonicity result might be called a law of diminishing returns. Weber and Stidham use this reasoning for a cycle of queues.

Considering the minimization in (2.1), if

$$V_n(x + d_i^1) - V_n(x + d_i^0) \geq V_n(x + d_i^1 + d_j) - V_n(x + d_i^0 + d_j) \tag{3.1}$$

for all $i \in \mathcal{C}$, $d_i \neq d_j \in \mathbf{D}$, and x such that all four points are in X, then the future cost of control μ_i decreases when transition d_j occurs. Because c_i is convex, a $\mathbf{D}$-monotone optimal control must exist at stage $n+1$. For simplicity, change x in (3.1) and write

$$V_n(x + d_i) - V_n(x) \geq V_n(x + d_i + d_j) - V_n(x + d_j). \tag{3.2}$$

Since $V = \lim_{n \to \infty} V_n$, if (3.2) holds with V_n replaced by V then $\mathbf{D}$-monotonicity also holds for the infinite-horizon problem. Condition (3.2) is an example of submodularity of a function on a lattice (Topkis 1978).

Definition. The function $f : X \to R$ is *submodular* w.r.t. $\mathbf{D}$ on X if

$$f(x + d_i) + f(x + d_j) - f(x) - f(x + d_i + d_j) \geq 0 \tag{3.3}$$

for all $d_i \neq d_j \in \mathbf{D}$ and x such that all four points are in X.

The following lattice condition, introduced in Weber and Stidham 1987, will be used to guarantee that (3.2) holds sufficiently close to the boundary of X.

Definition. $\mathbf{D}$ is *compatible* with X if for all $d_i \neq d_j \in \mathbf{D}$, $x + d_i$ and $x + d_j \in X$ implies x and $x + d_i + d_j \in X$.

For most queueing control problems, we can think of X as the discrete equivalent of a polyhedral set and give a simple interpretation of compatibility: only one d_i can cross each bounding hyperplane in each direction. Let $\mathbf{D}(x) = \{d \in \mathbf{D} : x + d \in X\}$ be the set of feasible directions from state x. If $\mathbf{D}$ is compatible with X and f is submodular w.r.t. $\mathbf{D}$ on X, we will say that f is submodular w.r.t. $\mathbf{D}(x)$, since (3.3) holds for all $x \in X$ and $d_i \neq d_j \in \mathbf{D}(x)$.

Weber and Stidham establish submodularity of V by showing that submodularity is preserved under value-iteration. The key step is the following lemma, which is proven in the Appendix (their proof is very similar but does not include choice controls).

Lemma 1 *If* $\mathbf{D}$ *contains all control directions* $d_i, i \in \mathcal{C}$,

(i) $f(x)$ *is submodular w.r.t.* $\mathbf{D}(x)$, *with* $f(x) = \infty$, $x \notin X$,

(ii) $\mathbf{D}$ *is compatible with* X, *and*

(iii) d_i^0 *is always feasible, i.e.,* $x + d_i^0 \in X$ *whenever* $x \in X$,

then for each $d_k, k \in \mathcal{C}$,

$$g(x) = \min_{0 \leq \mu \leq \overline{\mu}_k} \{c_k(\mu) + \mu f(x + d_k^1) + (\overline{\mu}_k - \mu) f(x + d_k^0)\}$$

is also submodular w.r.t. $\mathbf{D}(x)$.

Submodularity must also be preserved by uncontrolled transitions (the first sum in Equation 2.1). One type of transition that preserves submodularity is a state-independent transition, $D_i x = x + d_i$ for all $x \in X$, i.e., Poisson arrivals. Some other transitions also work, depending on the geometry of $\mathbf{D}$ and X; see Section 6 for examples.

Definition. The uncontrolled transition $D_i, i \in \mathcal{U}$, *preserves submodularity* w.r.t. $\mathbf{D}(\mathrm{x})$ if $f(D_i(x))$ is submodular whenever $f(x)$ is submodular.

Monotonicity now follows easily from the lemma.

Theorem 1 (Monotonicity) *If* $\mathbf{D}$ *contains all control directions* $d_i, i \in \mathcal{C}$,

(i) $h(x)$ *is submodular w.r.t.* $\mathbf{D}(x)$,

(ii) $\mathbf{D}$ *is compatible with* X,

(iii) d_i^0 *is always feasible, and*

(iv) all uncontrolled transitions preserve submodularity w.r.t. $\mathbf{D}(x)$

then there exists a $\mathbf{D}$*-monotone optimal control.*

Proof. First show that $V_n(x)$ is submodular w.r.t. $\mathbf{D}(x)$ using induction on n. Initially $V_0(x) = 0$ is submodular. In (2.1), $h(x)$ is submodular by (i), the first sum by (iv), and the second sum by the lemma and the inductive hypothesis. Since $V_n \to V$, it follows that V is submodular w.r.t. $\mathbf{D}(x)$ and a $\mathbf{D}$-monotone optimal control exists.
□

Note that the inductive proof establishes monotonicity for finite and infinite-horizon problems.

Weber and Stidham's monotonicity result is equivalent to the following specialization of our theorem.

Corollary (Control Monotonicity) *If* $\mathbf{D} = \{d_i : i \in \mathcal{C}\}, d_i^0 = 0$ *for all* $i \in \mathcal{C}$, *and (i), (ii), and (iv) above, then there exists a control-monotone optimal control.*

The generalization of adding other directions to $\mathbf{D}$ is very useful in some problems because it strengthens the monotonicity result; the strongest way to use this generalization is to consider one additional direction at a time. For systems with uncontrolled transitions, transition monotonicity (if it holds) tends to be a much stronger result than control monotonicity. The next section describes some of the useful implications of transition monotonicity. Choice controls with $d_i^0 \neq 0$ are useful in some arrival routing problems. We have not been able to establish monotonicity without condition (iii); with this condition, the theorem does not apply to problems with server allocation or routing between servers.

Remark: If $\mathbf{D}$-monotonicity holds with e_k or $-e_k$ in $\mathbf{D}$, it can be used to establish monotonicity with respect to the x_k axis. Depending on the geometry of the problem,

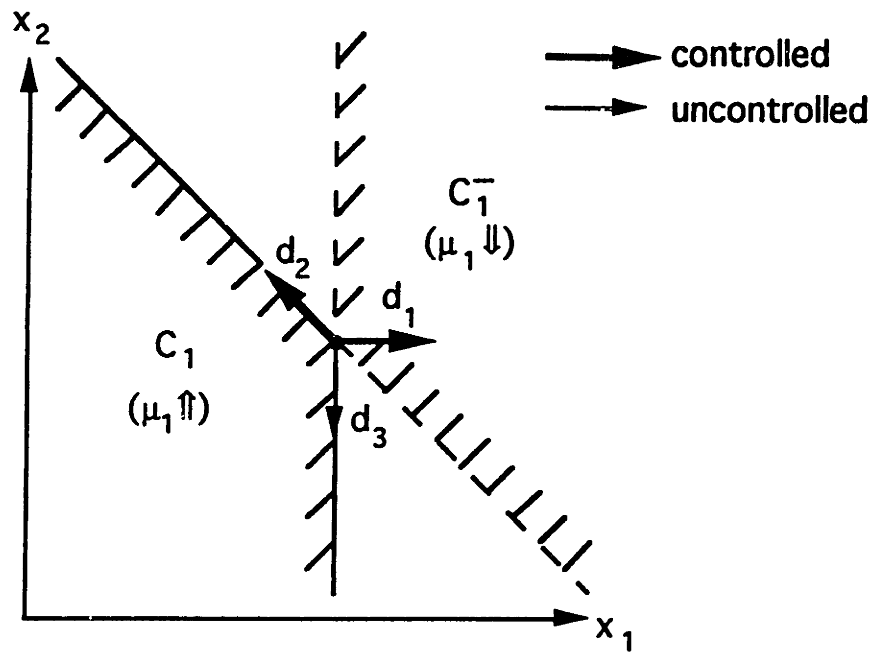

Figure 2-1: Increasing and Decreasing Directions for the Control μ_1

it may be possible to infer state space monotonicity from other direction sets. **D**-monotonicity establishes that the optimal control for a transition d_i increases after any other transition in **D**. It will also increase after any sequence of transitions in **D**; hence, the convex cone C_i of $\mathbf{D} \setminus \{d_i\}$ is a set of directions in which μ_i is increasing and the negative convex cone C_i^- is a set of directions in which μ_i is decreasing. Figure 2-1 shows these cones for the two-stage production/inventory system of Section 6 under transition monotonicity, where $\mathbf{D} = \{d_1, d_2, d_3\}$. If the vector e_k, giving the direction of the x_k-axis, lies within C_i or C_i^-, then monotonicity holds with respect to the x_k coordinate. For example, the control d_1 (and d_2) in Figure 2-1 is monotone with respect to both x_1 and x_2. Control monotonicity corresponds to the convex cones generated by smaller sets **D**. If only control monotonicity is established in Figure 2-1, the convex cone of μ_1 is the single direction d_2, which is not sufficient to establish state space monotonicity.

2.4 Control Regions and Switching Functions

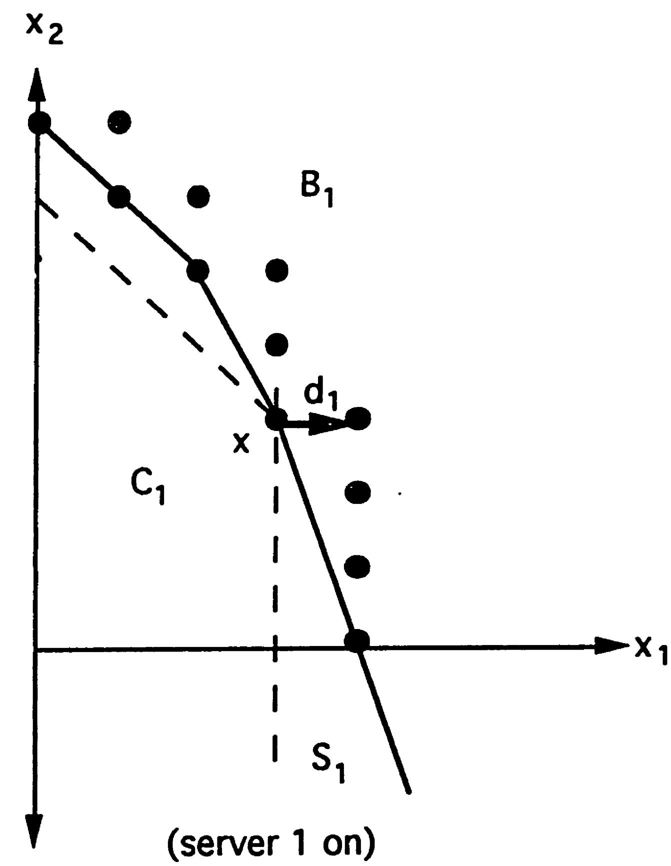

Figure 2-2: The cone C_1 (to the left of the dashed line) is contained in the interior region S_1 (to the left of the solid line); B_1 contains the points outside of S_1.

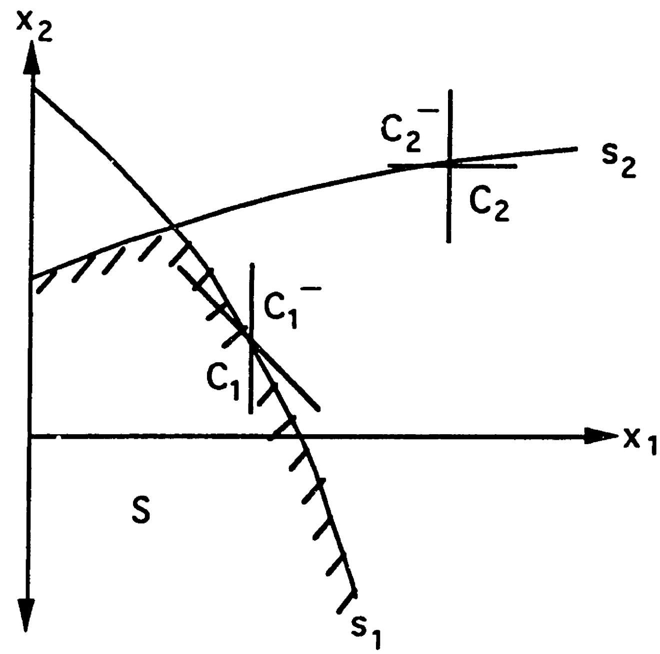

Figure 2-3: Switching Functions and the Interior Region

It is well known that if the service cost c_i in (2.1) is linear, then an all-or-nothing (bang-bang) control is optimal, with $\mu_i(x) = 0$ or $\overline{\mu}_i$. Bang-bang, transition-monotone controls can be characterized very simply as a control region. Let $S_i = \{x \in X : \mu_i(x) = \overline{\mu}_i\}$, $S_i' = \{x + d_i : x \in S_i\}$, and $B_i = S_i' \setminus S_i$ for $i \in \mathcal{C}$. In queueing terminology, S_i is the region where server i is on. As illustrated in Figure 2-2, B_i contains points on the boundary of S_i that are reached by one transition d_i. For any $x \in S_i$, consider the cone C_i for control μ_i with vertex x. Since μ_i is on at x, it must be on in all of C_i (similarly, if μ_i were off at x, it would be off in all of C_i^-); i.e., $C_i \cap X \subseteq S_i$. But all transitions other than d_i lie in C_i, so they cannot cause the system state to leave S_i. By a translation argument, these transitions cannot leave S_i' either. Under such a control, the system never leaves the ***control region*** $S' = \bigcap_{i \in \mathcal{C}} S_i'$. This region consists of an interior $S = \bigcap_{i \in \mathcal{C}} S_i$ in which all controls are on and boundaries in which one or more controls are off. The system can only enter a boundary in which μ_i is off through the transition d_i.

The control region can usually be defined using a *switching function* for each control μ_i. Hajek proves that switching functions exist for a system of two interacting queues; we give sufficient conditions for the existence of switching functions in general **D**-monotone systems. If the direction e_k lies in the cone C_i or C_i^- for control μ_i, then S_i can be defined by a switching function $s_i(x^{(k)})$, where $x^{(k)} = (x_1, \ldots, x_{k-1}, x_{k+1}, \ldots, x_m)$, such that $\mu_i(x) = \overline{\mu}_i$ if and only if $x_k \geq (\leq)\ s_i(x^{(k)})$. In this case, the interior S is simply a subset of X bounded by the switching functions s_i. Moreover, the function s_i must lie between C_i and C_i^- for the control μ_i. Figure 2-3 gives an idealization of these regions (the s_i are actually step functions). In queueing applications, these controls turn a server off when a boundary is reached; such controls have been analyzed on a continuous state space as reflected Brownian motion (Harrison 1985).

2.5 Some Generalizations

Although compatibility is a useful condition because it is easily verified, other conditions can be used in the theorem. If it is possible to extend the cost function h onto all of Z^m in a way that preserves the submodularity of h and makes h much larger at $y \notin X$ than at nearby $x \in X$, then h acts as a barrier and the optimal control for the unbounded problem ($V_0(x) = 0$ for all x) will be feasible for the original problem. Since there are no boundaries, compatibility is not needed to prove the lemma for the unconstrained problem and establish transition monotonicity. Hajek uses a similar extendibility approach to prove monotonicity in a two-station network. However, extendibility of the cost function is at least as strong a condition; it implies compatibility, as the following argument shows.

To formalize the concept of nearby values, let

$$\begin{aligned} n_{\text{sup}}(y) &= \max\{h(x):\ x \in X;\ x \pm d = y,\ d \in \mathbf{D}\} \\ n_{\text{inf}}(y) &= \min\{h(x):\ x \in X;\ x \pm d_i = y \text{ or } x \pm d_i \pm d_j = y,\ d_i, d_j \in \mathbf{D},\ i \neq j\}. \end{aligned}$$

If the set used in its definition is empty, then set $n_{\text{sup}}(y)$ or $n_{\text{inf}}(y)$ to zero. Suppose that h is submodular with respect to **D** on X and that for any M there exists an extension $\tilde{h} : Z^m \rightarrow R$ of h that is submodular on Z^m with $\tilde{h}(x) \geq 2n_{\text{sup}}(x) - n_{\text{inf}}(x) +$

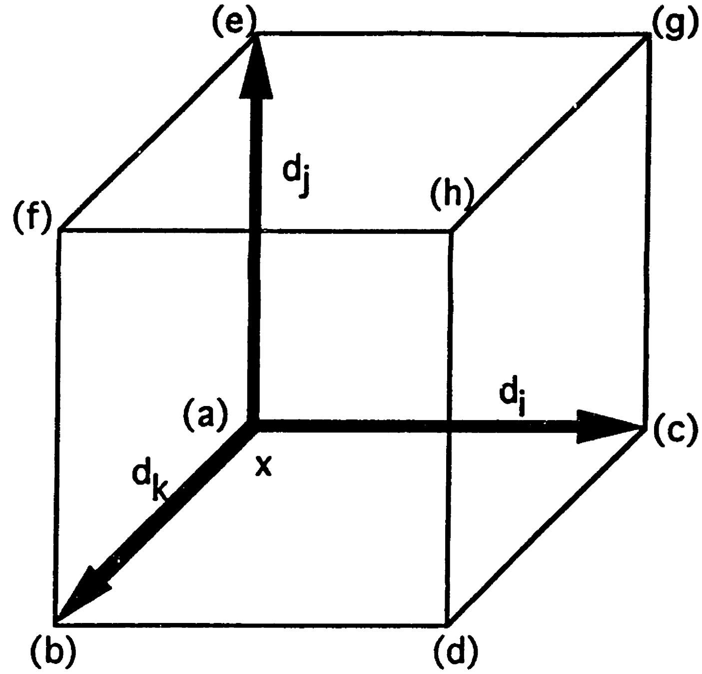

Figure 2-4: Points at which f is Evaluated

M. Then for any controlled transition $d_i \in \mathbf{D}$ and x such that $x \in X$ and $x + d_i \notin X$, we have $\tilde{h}(x + d_i) \geq \tilde{h}(x) + M$, and for sufficiently large M the optimal control for the unbounded problem is $\mu_i(x) = 0$. Hence, this control is feasible (and optimal) for the original problem.

To show compatibility, suppose there exists $d_i, d_j \in \mathbf{D}$, $i \neq j$, and x such that $x + d_i, x + d_j \in X$ and $x \notin X$. Then

$$\begin{aligned}\tilde{h}(x + d_i) + \tilde{h}(x + d_j) - \tilde{h}(x) - \tilde{h}(x + d_i + d_j) \;\leq\; & n_{\text{sup}}(x) + n_{\text{sup}}(x) \\ & -[2n_{\text{sup}}(x) - n_{\text{inf}}(x) + M] - n_{\text{inf}}(x),\end{aligned}$$

which is negative for $M > 0$. But this contradicts submodularity of $\tilde{h}$, so $x \in X$. A similar argument shows that $x + d_i + d_j \in X$.

The compatibility condition can be weakened slightly if there are no service costs. Given four points in X on which f is submodular, as in (3.3), $g(x)$ in the lemma is only submodular on these points if certain combinations of the four points obtained by adding d_k, $k \neq i$ or j, are in X (i.e., points (b), (d), (f), and (g) in Figure 2-4).

Table 2.1: Comparison of Boundary Conditions

Case	Points in X	Compatible	g Submodular
1	b,d,f,h	$\surd$	$\surd$
2	d,f,h		
3	b,f,h		$\surd$
4	b,d,h		$\surd$
5	b,d,f		
6	f,h	$\surd$	$\surd$
7	d,h	$\surd$	$\surd$
8	d,f		
9	b,h		$\surd$
10	b,f		
11	b,d		
12	h	$\surd$	$\surd$
13	f		
14	d		
15	b		$\surd$
16	–	$\surd$	$\surd$

Writing

$$g(x) = \overline{\mu}_k \min\{f(x), f(x + d_k)\},$$

a case-by-case analysis reveals that nine of the 2^4 possibilities guarantee submodularity. In contrast, compatibility allows only five of these possibilities (Table 2.1) and must hold for any four points $x, x + d_i, x + d_j$, and $x + d_i + d_j$, not just those defined as (b), (d) , (f), and (g) are. If $k = i$, a similar analysis shows that compatibility is as weak as possible, allowing three out of 2^2 possible combinations of points in X. However, we have not found any meaningful examples where $g(x)$ is submodular and conditions (i), (iii), and (iv) hold but compatibility does not hold.

We have assumed that the lower control limits are zero. Positive lower control limits $\underline{\mu}_i \leq \mu_i \leq \overline{\mu}_i$ can be modeled by introducing a duplicate uncontrolled transition $d_j = d_i, j \in \mathcal{U}, i \in \mathcal{C}$, and resetting the control limits to $\lambda_k = \underline{\mu}_i$ and $0 \leq \mu_i \leq \overline{\mu}_i - \underline{\mu}_i$. However, the duplicate transition in **D** adds the following conditions to the theorem:

1. By submodularity, $2h(x + d_i) - h(x) - h(x + 2d_i) \geq 0$; i.e., h is convex in the direction d_i.

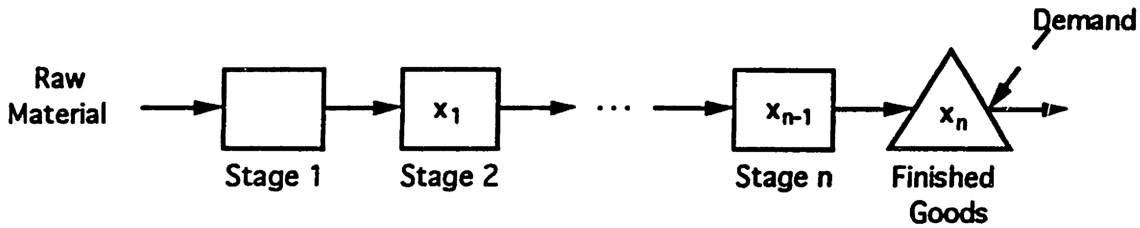

Figure 2-5: A Tandem Make-to-Stock Queueing System

2. By compatibility, if $x + d_i \in X$ then $x, x + 2d_i \in X$; i.e., X has no boundaries in the directions $\pm d_i$.

Monotonicity can be extended to the long-run average cost setting; the proof given by Weber and Stidham applies directly. If the state space X is finite (or h is bounded), well-known methods can be applied to let $\alpha \to 0$. In practice, this is not much of a restriction because it is usually clear how to truncate the state space. Note, however, that monotonicity holds for discounted optimal controls even when the system is unstable and cannot be truncated. For unbounded h we need the following assumptions:

1. There exists a policy with finite average cost.

2. For all $x, y \in X$, there exists a policy under which the undiscounted cost until first passage from x to y is finite.

3. The cost functions h_k are bounded below.

4. For some ordering of the states $x \in X$, $h(x) \to \infty$.

The second assumption is a stronger form of irreducibility and the fourth guarantees that there are only finitely many good states. With these additional assumptions, the theorem holds for average cost optimal policies.

2.6 Applications

Tandem Make-to-Stock Queueing Systems

Consider the tandem queueing system of (Figure 2-5) in which each stage operates as a $\cdot/M/1$ queue. Items are endogenously released into the system, pass through each stage, and then placed in a finished goods inventory that services a Poisson demand. Demand that cannot be met from inventory is backordered and recorded as a negative inventory. Unlike traditional queueing systems, this make-to-stock system allows deficits in the final "queue". One application of this model is a production/inventory system, where each stage represents a machine in a production line; this terminology will be used to discuss make-to-stock systems.

Denote the system state by $x = (x_1, \ldots, x_n)$, where x_i is the number of items at stage $i+1$, $i = 1, \ldots, n-1$ and x_n is the finished goods inventory. Because the supply of raw material is unlimited, there is no queueing and no state variable at stage 1. Stage i has production rate μ_i controlled between 0 and $\overline{\mu}_i$, with transitions $d_1 = e_1$ and $d_i = e_i - e_{i-1}, i = 2, \ldots, n$. Demand is an uncontrolled transition $d_{n+1} = -e_n$ with rate λ. These transitions are illustrated in Figure 2-1 for a two-stage system. The state space is $X = \{x \in Z^n : x_i \geq 0, i = 1, \ldots, n-1\}$. There is a work-in-process holding cost h_k after stage k (items available to stage $k+1$), $k = 1, \ldots, n-1$, a finished goods holding cost h_n, and a finished goods backorder cost b; i.e.,

$$h(x) = \sum_{k=1}^{n-1} h_k x_k + h_n x_n^+ + b x_n^- . \tag{4.1}$$

No holding cost is assessed at stage 1 because it will contain at most one item in production and none queued. The production cost $c(\mu)$ is approximately the same for all stable controls, and so is assumed to be zero. It is interesting to note that this tandem make-to-stock queueing system differs from a traditional tandem queueing system only in the removal of one boundary from X, the convex (as opposed to increasing) cost function, and which transition is uncontrolled. As suggested in Section 1, the corollary (Weber and Stidham's result) can be applied to this system by replacing the uncontrolled demands with a controllable transition with an arbitrarily large reward for the control, so that the maximum rate will always be optimal. A more direct approach is to let $\mathbf{D}$ contain all transitions and check the following conditions.

1. *h is submodular w.r.t.* $\mathbf{D}(x)$. The submodularity condition (3.3) clearly holds

when all four points lie on the same linear segment of h. The exception involves the transitions $e_n - e_{n-1}$ and $-e_n$ at $x_n = 0$; here (3.3) is

$$h(x + e_n - e_{n-1}) + h(x - e_n) - h(x) - h(x - e_{n-1}) \geq 0.$$

Substituting (4.1) into the above gives

$$h(x) + h_n - h_{n-1} + h(x) + b - h(x) - h(x) + h_{n-1} = h_n + b \geq 0.$$

2. **D** *is compatible with* X. The only transitions that cross the boundary $x_k = 0, k < n$, are $e_k - e_{k-1}$ in the positive direction and $e_{k+1} - e_k$ in the negative direction, so the geometric condition for compatibility is satisfied. A case-by-case check of compatibility can also be performed.

There are no choice controls ($d_i^0 = 0$) and Poisson demand preserves submodularity, so we conclude that there exists a transition-monotone optimal control. To determine state space monotonicity, note that

$$\begin{aligned} e_k &= d_1 + \cdots + d_k \\ &= -(d_{k+1} + \cdots + d_n + d_{n+1}). \end{aligned}$$

For $k < j$, the first equality implies that μ_j is increasing in x_k (upstream inventory); for $k \geq j$, the second equality implies that μ_j is decreasing in x_k (downstream inventory). Furthermore, the optimal control is bang-bang with the following switching functions: stage i is on when $x_i \leq s_i(x^{(i)})$ and off otherwise (recall that x_i is the inventory immediately downstream from stage i and that $x^{(i)}$ is the vector of state variables other than x_i). The rates of change of the switching functions are also constrained because the corresponding regions S_i must contain the cones C_i. In the two-stage system of Figure 2-3, s_1 has a slope no greater than -1 (more precisely, the step function s_i is decreasing and has no horizontal segment of length greater than one) and s_2 is increasing. It is interesting to note that the base stock policy that has been proposed for this system (see Buzacott, Price, and Shanthikumar 1991) uses the

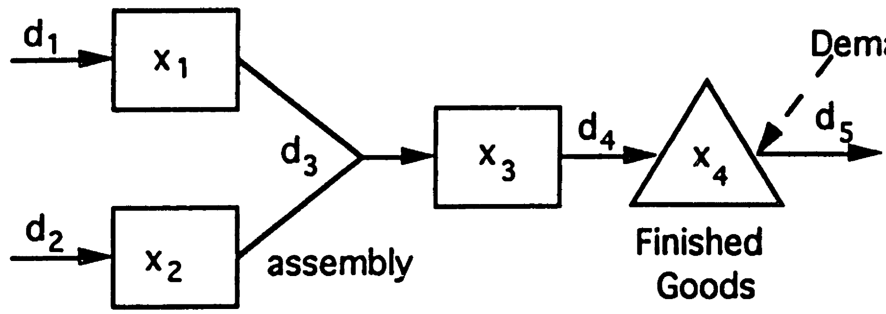

Figure 2-6: An Assembly Make-to-Stock Queueing System

limiting slopes of -1 for s_1 and zero for s_2. Similar characterizations of the optimal control of arrivals to two queues are given in Ghoneim and Stidham 1985.

Assembly and Disassembly Systems

For make-to-stock queueing networks with topologies that include assembly, such as in Figure 2-6, very similar arguments can be used. Each stage except finished goods has a boundary $x_k \geq 0$ which is crossed exactly once in the increasing direction and once in the decreasing direction, so compatibility holds. The cost function $h(x)$ is submodular as before. Transition monotonicity implies that μ_j is increasing in upstream inventory and inventory in other branches, and decreasing in downstream inventory. For the example in Figure 2-6, μ_1 is increasing in x_1 and x_2 and decreasing in x_3 and x_4; μ_3 is increasing in x_1, x_2, and x_3 and decreasing in x_4.

Reversing the above network topology gives a disassembly system, which might arise when a single production process yields multiple products. Similar arguments show that μ_j is increasing in upstream inventory and decreasing in downstream inventory and inventory in other branches. Analogous results hold for traditional queueing networks with assembly or disassembly topologies.

It should be noted that these results do not apply to systems with more than one transition in or out of a station. For example, if stations 1 and 2 in Figure 2-6 supply the same part to station 3, $d_3 = e_3 - e_1 - e_2$ is replaced by two transitions, $e_3 - e_1$ and $e_3 - e_2$, and compatibility is lost. A distribution network with more than one retail location supplied by a single supplier is also not compatible.

Arrival Routing

The theorem applies to the arrival routing problem shown in Figure 2-7, which is a

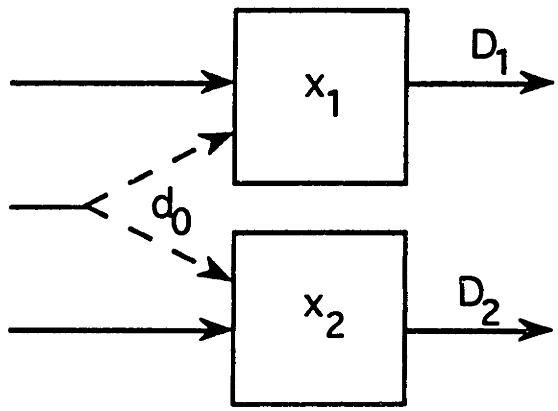

Figure 2-7: A System with Arrival Routing

special case of the problem considered in Hajek (1984). In addition to uncontrolled arrivals at each queue, the choice control $d_0 = e_2 - e_1$ routes arrivals to queue 1 ($d_0^0 = e_1$) or queue 2 ($d_0^1 = e_2$). There are also uncontrolled servers, $D_i x = x - e_i$, $x_i > 0$, $i = 1, 2$. As usual, the state space is $X = R_+^2$ and $h(x)$ is a linear holding cost. Let $\mathbf{D} = \{e_2 - e_1, e_1, -e_2\}$. The only difficult condition to check is (iv). The transitions D_1 and D_2 do not preserve submodularity for general functions $f(x)$; the submodularity conditions for $f(D_i x)$ require that f be convex and increasing in the x_1 and x_2 directions at points on the boundary $x_1 = 0$ or $x_2 = 0$. Hajek (1984) notes that submodularity w.r.t. $\mathbf{D}(x)$ implies convexity; a stochastic coupling argument shows that $V_n(x)$ is increasing in x_k. With these auxiliary conditions, D_1 and D_2 preserve submodularity $V_n(x)$ and $\mathbf{D}$-monotonicity is established; namely, the optimal arrival rate at queue 2 is increasing in x_1 and decreasing in x_2.

Uncontrolled Servers

The previous example illustrates that the theorem applies to some problems with uncontrolled service rates; another example is a series queue. Let x_i be the number of customers at queue i, $i = 1$ (upstream) and 2 (downstream), and $D = \{e_1, e_2 - e_1, -e_2\}$, the set of all transitions. If only the downstream server is uncontrolled, then its transition $D_2 x = x - e_2$, $x_2 > 0$ preserves submodularity and the system is transition-monotone. To see this, again note that submodularity of $f(D_2 x)$ at points on the boundary $x_2 = 0$ requires that f be convex in x_1 and increasing in x_1 and x_2 at these points. If, however, only the upstream server is uncontrolled, its transition

$R_{12}x = x - e_1 + e_2$, $x_1 > 0$ does not preserve submodularity. Submodularity of $f(R_{12}x)$ at points with $x_1 = 0$ requires that f be increasing in the direction $e_1 - e_2$, which is not generally true of $V_n(x)$. Weber and Stidham (1987) construct an example that is neither submodular nor transition-monotone.

2.7 Appendix - Proof of Lemma

We need to show that, for $x \in X$ and $d_i \neq d_j \in \mathbf{D}(x)$,

$$\Delta = g(x + d_i) + g(x + d_j) - g(x) - g(x + d_i + d_j) \geq 0. \qquad (A.1)$$

Let $\mu(x)$ be the control that minimizes $g(x)$. Assume $\mu_k(x + d_i) \geq \mu_k(x + d_j)$.

Case I: $d_k \neq d_j$. If no boundaries are hit, submodularity of f and convexity of c_k give

$$\mu_k(x + d_i + d_j) \geq \mu_k(x + d_i) \geq \mu_k(x + d_j) \geq \mu_k(x),$$

or for brevity, $\mu_{i+j} \geq \mu_i \geq \mu_j \geq \mu$, motivating us to approximate μ_{i+j} by μ_i and μ by μ_j. Substituting into (A.1) and rearranging terms,

$$\begin{aligned}
\Delta \geq\ & c_k(\mu_i) + \mu_i f(x + d_i + d_k^1) + (\overline{\mu} - \mu_i) f(x + d_i + d_k^0) \\
& + c_k(\mu_j) + \mu_j f(x + d_j + d_k^1) + (\overline{\mu} - \mu_j) f(x + d_j + d_k^0) \\
& - c_k(\mu_j) - \mu_j f(x + d_k^1) - (\overline{\mu} - \mu_j) f(x + d_k^0) \\
& - c_k(\mu_i) - \mu_i f(x + d_i + d_j + d_k^1) - (\overline{\mu} - \mu_j) f(x + d_i + d_j + d_k^0) \\
=\ & (\overline{\mu} - \mu_j)[f(x + d_i + d_k^0) + f(x + d_j + d_k^0) - f(x + d_k^0) - f(x + d_i + d_j + d_k^0)] \\
& + (\mu_i - \mu_j)[f(x + d_i + d_j + d_k^0) + f(x + d_i + d_k^1) \\
& \quad - f(x + d_i + d_k^0) - f(x + d_i + d_j + d_k^1)] \\
& + \mu_j [f(x + d_i + d_k^1) + f(x + d_j + d_k^1) - f(x + d_k^1) - f(x + d_i + d_j + d_k^1)].
\end{aligned}$$

By assumption (iii), all four points in the first bracketed expression are in X; hence, it is nonnegative by the submodularity of f w.r.t. d_i and d_j. For the second expression,

$\mu_i > 0$ implies $x+d_i+d_k^1 \in X$. Also, $x+d_i+d_j+d_k^0 \in X$ by assumption (iii). Then, by compatibility of d_j and d_k, all four points are in X and the expression is nonnegative. Similarly, $\mu_j > 0$ implies $x + d_j + d_k^1 \in X$ and, since $\mu_i \geq \mu_j > 0, x + d_i + d_k^1 \in X$. Compatibility of d_i and d_j requires that all four points be in X, and the third expression is also nonnegative by the submodularity of f w.r.t. d_i and d_j.

Case II: $d_k = d_j$. Approximating μ_{i+j} by μ_j and μ by μ_i in (A.1) gives

$$\begin{aligned}\Delta \geq\ & (\overline{\mu} - \mu_i)[f(x + d_i + d_k^0) + f(x + d_j + d_k^0) - f(x + d_k^0) - f(x + d_i + d_j + d_k^0)] \\ & +(\mu_i - \mu_j)[f(x + d_i + d_k^1) + f(x + d_j + d_k^0) - f(x + d_k^1) - f(x + d_i + d_j + d_k^0)] \\ & +\mu_j[f(x + d_i + d_k^1) + f(x + d_j + d_k^1) - f(x + d_k^1) - f(x + d_i + d_j + d_k^1)].\end{aligned}$$

The first and third terms are nonnegative as before; the second term is zero, since $d_k = d_j$.□

Chapter 3

Two-Station Tandem Production/Inventory System

Considerable attention has been given in recent years to viewing manufacturing facilities as production/inventory systems. This framework recognizes the importance not only of inventory control but also of queueing due to capacity constraints and uncertainty. In this paper, we consider a production/inventory system that consists of several stages in series that produce a single product. Each stage of the facility contains a single workstation that is modeled as a queue with controllable service rate. Once completed, items are counted as finished goods until they are consumed by an exogenous demand. As in inventory/distribution systems, production is driven by demand, but unlike standard inventory models there are capacity constraints and queueing in the production process. Demand that cannot be met from inventory is backordered and met by the next available finished item. Holding costs are incurred at each stage, as well as finished goods holding and backordering costs.

We consider the problem of finding optimal controls of the production rates for a long-run average or discounted cost criterion. In contrast, most studies of production line control problems assume that a relatively simple mechanism, such as buffers or kanbans, is used to control the system. Its performance is evaluated or the best policy using that mechanism is found. Restricting attention to a certain mechanism may be practical, given that optimizing a cost function is difficult and "optimal" policies may be impractical to implement; however, it is also desirable to know how the mechanisms compare to each other and to the optimal policy.

The demand environment of our production/inventory system is make-to-stock with complete backordering; other environments have been more widely studied. A make-to-order environment, where production of an item cannot begin until a demand is received, may be dictated by customization requirements on orders or chosen for economic reasons when customers will tolerate the waiting time. This environment corresponds to a tandem queue; its optimal control has been studied by Rosberg, Varaiya and Walrand (1982) and Weber and Stidham (1987), among others. A buffer control mechanism results in a tandem queue with blocking; lost sales can be incorporated in this model by including a finite first buffer. Approximate performance evaluation and optimal buffer placement have been studied; see Perros (1984), Smith and Daskalaki (1988), Hillier, Boling and So (1986), and the references therein. All of the environments described thus far are exogenous demand, or "pull" systems. Unlimited demand "push" systems can be modeled as closed queueing networks; buffer placement for these systems has been studied by Conway et al. (1988). Kanban policies, pioneered by Toyota (Sugimori et al. 1977), are studied by Mitra and Mitrani (1990) and Muckstadt and Tayur (1991) in an unlimited demand setting.

Several previous studies of multi-stage, single-product production/inventory systems have obtained approximate results for evaluating a specific control mechanism. Mitra and Mitrani (1991) and Cheng and Yao (1991) both study kanban policies; they also establish sample path and stochastic dominance of kanban mechanisms over traditional buffer mechanisms. The dominance is essentially due to the movable buffers within a kanban cell, in contrast with the traditional fixed buffers. Base stock policies, motivated by distribution/inventory systems, are evaluated by Lee and Zipkin (1990) and Buzacott, Price and Shanthikumar (1991) using stage decomposition approximations. Base stock is not really a new concept in manufacturing since, as the second paper points out, MRP systems essentially use a base stock mechanism with a demand forecast included in the target stock levels. Constant work-in-process (CONWIP) can be viewed as a special case of this policy.

Our approach is to find optimal policies, using analytical and numerical methods, and compare them with some of the simpler control mechanisms being used in manufacturing. To accomplish this program, a simple two-station problem is considered. It is assumed that demand is Poisson, service times are exponential, and there are no set-up costs. It is hoped that the insights gained from this idealized system, with careful attention to its limitations, will be applicable to more realistic systems. It is encouraging to note that van Ryzin, Lou and Gershwin (1991) and Lou and van Ryzin (1989) obtained similar numerical results for a similar system in which the only source of uncertainty is unreliable machines, indicating that there is at least some robustness to our findings.

Two types of results are obtained in this paper. First, very simple policies are shown to be optimal under certain extreme conditions on the problem parameters. Chapter 2 established that the optimal policies generally consist of a switching curve for each station, dividing the state space into an idle and busy region. We use stochastic coupling arguments to show that, for certain parameter values, these switching curves become essentially static priority rules. Conditions are found under which no inventory is held, as well as conditions under which all inventory is converted to finished goods (FG), i.e., the downstream station never idles unless it is starved. These results are comparable to those of Bielecki and Kumar (1988) for a single-stage production/inventory system. Interestingly, the popular base stock policy is shown to never be *exactly* optimal.

Second, numerical results are obtained using dynamic programming. These results further illustrate the tradeoffs of whether or not to hold inventory and whether to hold work-in-process (WIP) or FG. Holding WIP may seem to fly in the face of the just-in-time goal of eliminating WIP; in fact, our model provides a cost basis for deciding whether or not to hold WIP and FG. In a production/inventory system, WIP can perform two functions: not only does it serve as a buffer between asynchronous stations to increase throughput capacity (as in make-to-order systems), it can also supplement finished goods (FG) inventory to reduce backorders. The decision of where to place inventory depends on the relative holding costs and the rate at which

WIP can be converted into FG. Less WIP is held when its holding cost is high, the utilization of the upstream station is low, or the discount rate is high.

Optimal policies are also compared to the best base stock, kanban, and fixed buffer policies. It is found that base stock policies are nearly optimal when the upstream station is heavily utilized and the discount rate is small or zero. Kanban policies outperform base stock policies when the downstream station is a bottleneck or discounting is significant. Fixed buffer policies are consistently the worst, though the degradation is not always significant. It is encouraging that base stock and kanban policies are within a few percent of optimal for most test cases, since a switching curve policy would be more difficult to implement. Every type of policy is sensitive to the stock levels or buffer sizes, so that obtaining accurate demand and production rate data and setting these levels correctly remains a very important issue.

The patterns that appear in the numerical study can only be extrapolated to more complex systems tentatively and qualitatively. It is reasonable to expect that the desirability of holding WIP would be similar for systems with more stations, probably with most WIP being held downstream. However, the amount of WIP held is always modulated by the service time variability, which is sometimes less in real systems than our exponential assumption. It also should be noted that we apply a cost to the average WIP; most studies of buffer allocation apply a cost or constraint to the maximum WIP, e.g., total buffer capacity (see McClain and Moodie 1991). As is well known, WIP can have other adverse effects than just the inventory holding cost (see, for example, Schonberger 1982).

Another result that may be of use in future research is a transformation of make-to-stock systems into make-to-order systems. This equivalence allows some of the methods developed for traditional tandem queues to be applied to make-to-stock systems. For example, approximate evaluation of stationary distributions for tandem queues with blocking can be used to quickly identify suboptimal policies for our system.

The remainder of the paper is organized as follows. The problem is formulated mathematically in Section 1 and several control mechanisms are defined in Section

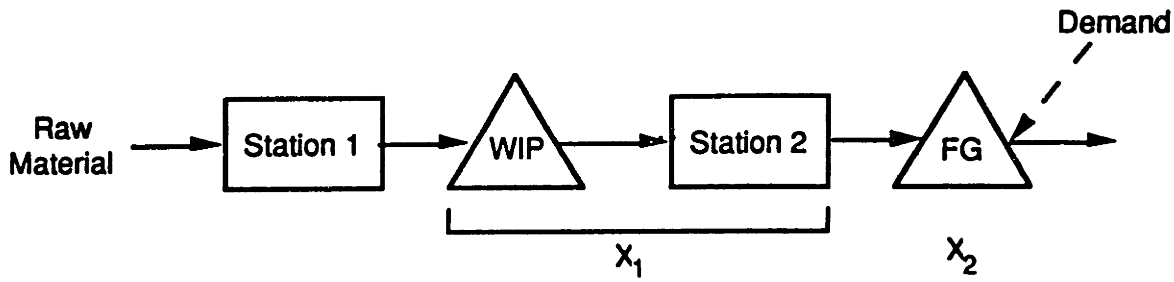

Figure 3-1: A Two-Stage Production/Inventory System

2. The optimality of simple controls under special parameter values is proven in Section 3, with some of the proofs deferred to the Appendix. A connection with traditional make-to-order queues is made in Section 4 and dynamic programming numerical results are presented in Section 5.

3.1 Problem Description

Consider the two-stage tandem production system of Fig. 3-1. Jobs are released into the system, processed at stage 1, then held in a work-in-process (WIP) buffer. When released into stage 2, they are processed there and then placed in a finished goods (FG) inventory that services an exogenous demand. Demand that cannot be met from inventory is backordered and recorded as a negative inventory. Denote the system state at time t by $X(t) = (X_1(t), X_2(t))$, where X_1 is the number of jobs available for stage 2 processing (including any item being processed at stage 2) and X_2 is the FG inventory. Because the supply of raw material is unlimited, there is no queueing and no state variable at stage 1.

Stage i consists of a single machine that operates as a $\cdot/M/1$ queue with production rate μ_i controlled between 0 and $\overline{\mu}_i$. Associated with these controls are the transitions $x \rightarrow x + d_i$, where $d_1 = e_1$ for μ_1 and $d_2 = e_2 - e_1$ for μ_2, as shown in Fig. 3-2. Here e_i is the unit vector along the ith axis. Demands occur according to a Poisson process with rate λ and cause the transition $d_0 = -e_2$. Stability of the system requires that $\lambda < \overline{\mu}_i$ for $i = 1, 2$. An *admissible* control policy π is a function $\mu(X, t)$ that is nonanticipating, i.e., depends only on $\{X(s); s \leq t\}$, and obeys the control limits

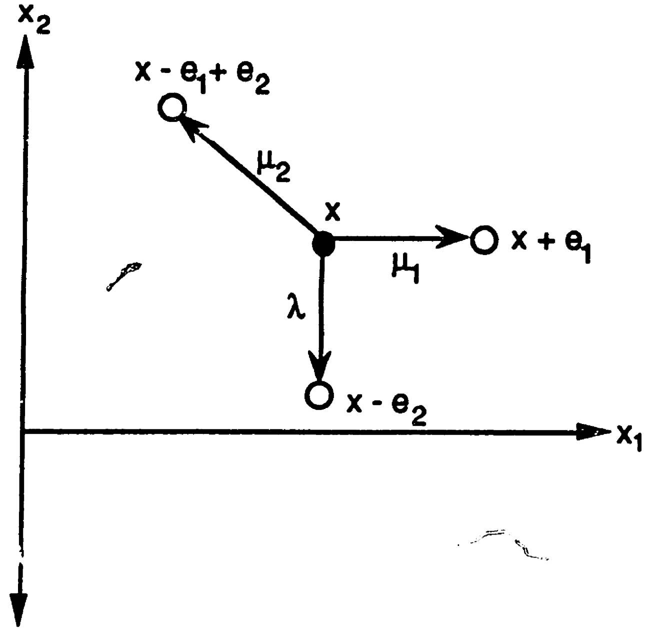

Figure 3-2: State Space Diagram

$0 \le \mu_i \le \overline{\mu}_i$ and $\mu_i(X,t) = 0$ if $X(t^-) + d_i \notin \mathbf{X} = \{x \in Z^2 : x_1 \ge 0\}$. Let Π denote the class of admissible policies. Because the system is memoryless, a Markov policy depending only on the current state x will be optimal; we denote this policy $\mu(x) = (\mu_1(x), \mu_2(x))$ for $x \in \mathbf{X}$.

The objective is to minimize WIP holding cost (incurred at a rate of one per job per unit time), FG holding cost $h > 1$, and FG backorder cost b, all discounted at a rate $\alpha > 0$ over an infinite time horizon. For policy π, the expected cost is

$$V^\pi(x) = E_x \int_0^\infty e^{-\alpha t} c(X(t))dt, \tag{3.1}$$

where $c(x) = x_1 + hx_2^+ + bx_2^-$. Here E_x denotes expectation given the initial state $X(0) = x$ and policy π. The optimal policy $\mu(x)$ achieves the minimum

$$V(x) = \min_{\pi \in \Pi} V^\pi(x) \tag{3.2}$$

simultaneously for all x. We will uniformize the process as in Lippman (1975) by defining the potential event rate $\Lambda = \overline{\mu}_1 + \overline{\mu}_2 + \lambda$. The n-stage cost function satisfies the dynamic programming equations

$$V_{n+1}(x) = \mathbf{T}V_n(x) \tag{3.3}$$

$$\mathbf{T}V(x) = \frac{1}{\Lambda + \alpha}[c(x) + \lambda V(x - e_2) + \overline{\mu}_1 \min\{V(x), V(x + e_1)\}$$
$$+ \overline{\mu}_2 \min\{V(x), V(x - e_1 + e_2)\}], \tag{3.4}$$

where we define $V_0(x) = 0$ and $V_n(x) = \infty$, $x \notin \mathbf{X}$. The infinite-horizon cost function satisfies

$$V(x) = \mathbf{T}V(x). \tag{3.5}$$

The form in which we have written (4) emphasizes that the optimal policy is *bang-bang*, i.e., $\mu_i(x) = 0$ or $\overline{\mu}_i$. Such a policy is specified by its idle and busy sets $\mathcal{I}_i = \{x \in \mathbf{X} : \mu_i(x) = 0\}$ and $\mathcal{B}_i = \mathbf{X} \setminus \mathcal{I}_i$. The existence of a Markov policy that achieves the minimum in (2) and the convergence of the n-stage policy and cost function to the infinite-horizon optimal policy and cost follow from the fact that only finitely many controls are considered at each state; see Bertsekas (1976).

An undiscounted, long-run average cost criterion will also be considered. In this case the average cost per stage, g, and the relative cost of starting in state x, $V(x)$, satisfy

$$V(x) + g = \mathbf{T}V(x), \tag{3.6}$$

where we arbitrarily set $V(0,0) = 0$. Existence and convergence results can be obtained for (6) by letting $\alpha \to 0$ in (4) and exploiting the fact that there are only a finite number of "good" states; see Weber and Stidham (1987).

It is shown in Veatch and Wein (1991) that optimal policies have the following monotonicity property: there exist *switching functions* $s_i(x_1)$ such that $\mu_i(x) = 0$ if and only if $x_2 > s_i(x_1)$. Furthermore, these functions have derivatives (or more precisely, differences, since they are defined on Z^+) $s_1'(x_1) \leq -1$ and $s_2'(x_1) \geq 0$, as illustrated in Fig. 3-3.

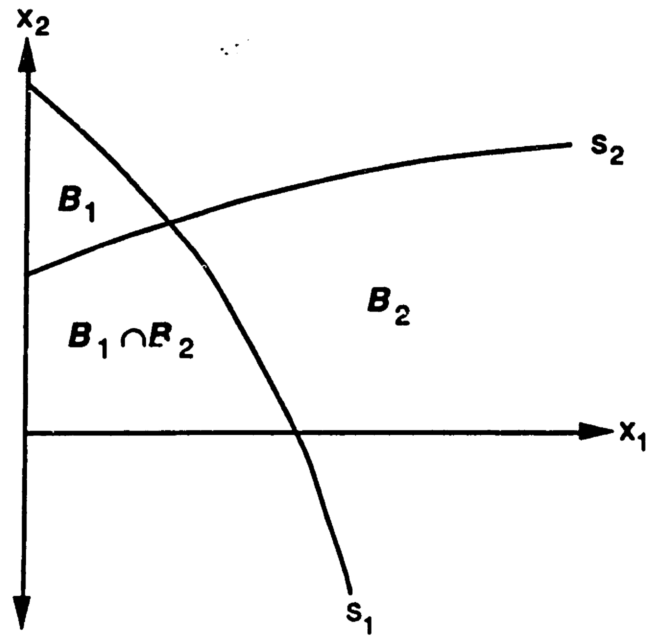

Figure 3-3: Properties of Optimal Policies

3.2 Control Mechanisms

This section describes several control policies that have been studied and used in production lines. In order to describe the mechanism by which they are usually implemented, we begin with a more detailed model of material and information flow through the system. Associated with each station is a physical or organizational *cell*. As shown in Fig. 3-4, orders (hereafter called demands) are placed on cell 1 according to a policy that depends only on the state of cell 2. Either a release occurs

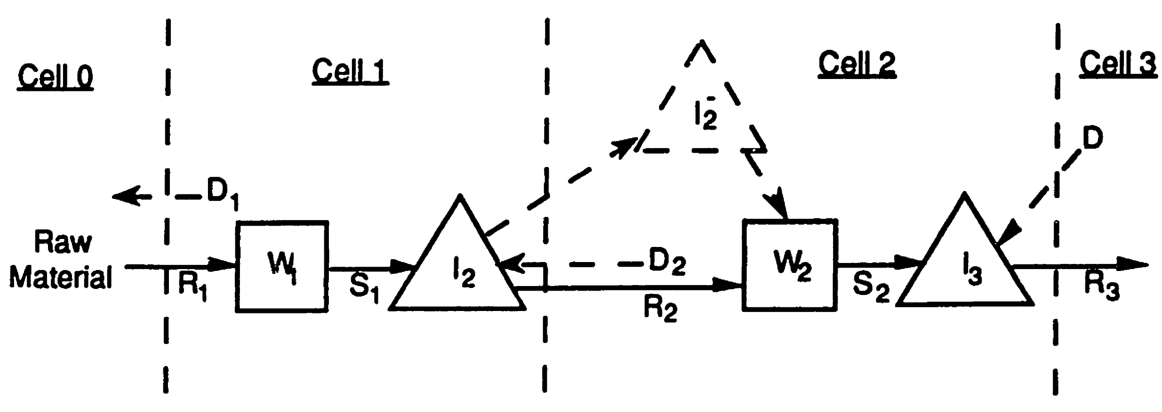

Figure 3-4: Material and Information Flow

immediately or the demand is backordered until there is inventory at the cell 1 output buffer. Define

$D_i(t)$ = demand placed by cell i on cell $i-1$ in $(0,t]$
$D(t)$ = exogenous demand in $(0,t]$
$R_i(t)$ = work released into cell i in $(0,t]$
$S_i(t)$ = service completions at station i in $(0,t]$
$W_i(t)$ = work at station i at time t
$I_i(t)$ = inventory position, cell i output buffer at time t.

The controls can be described in terms of demands as follows. Station i is idle when $W_i = 0$ and busy when $W_i > 0$; this defines $S_i(t)$. The dynamic equations are

$$R_1(t) = D_1(t) \tag{3.7}$$

$$R_i(t) = \min\{D_i(t), I_i(0) + S_{i-1}(t)\},\ i = 2,3 \tag{3.8}$$

$$I_i(t) = I_i(0) + S_{i-1}(t) - D_i(t) \tag{3.9}$$

$$W_i(t) = R_i(t) - S_i(t),\ i = 1,2. \tag{3.10}$$

In the notation of section 1, WIP is $x_1 = I_2^+ + W_2$, FG is $x_2 = I_3$, and the dynamic equations are

$$x_1(t) = x_1(0) + S_1(t) - S_2(t) \tag{3.11}$$

$$x_2(t) = x_2(0) + S_2(t) - D(t). \tag{3.12}$$

Base Stock

Under a base stock, one-for-one ordering policy each cell places an order upstream as soon as it receives an order. Hence, demands propagate through the system immediately and $D_i(t) = D(t)$. The application of this policy to production/inventory systems is discussed in Buzacott, Price and Shanthikumar (1991). Let c_1 and c_2 be the base stock levels for WIP and FG, respectively. The policy is characterized by the busy sets $\mathcal{B}_1 = \{x : x_1 + x_2 < c_1 + c_2\}$ and $\mathcal{B}_2 = \{x : x_1 > 0,\ x_2 < c_2\}$.

Kanban

A kanban policy has been applied to the make-to-stock environment by Mitra and Mitrani (1991). In terms of our model, the number of kanbans or cards in cell i is $c_i = I_i^- + W_i + I_{i+1}^+$; I_i^- represents the bulletin board and I_{i+1}^+ the output hopper in cell i. Demands occur when a job is released to the next cell, freeing a card: $D_i(t) = R_{i+1}(t)$. The policy is $\mathcal{B}_1 = \{x : x_1 + x_2^+ < c_1 + c_2\}$ and $\mathcal{B}_2 = \{x : x_1 > 0, x_2 < c_2\}$.

Fixed Buffer

Under a finite buffer policy (see, e.g., Conway et al. 1988) the system operates as two tandem, finite capacity $\cdot/M/1/c_i$ queues. The buffer size is c_1 between stations and c_2 for FG; the policy is $\mathcal{B}_1 = \{x : x_1 < c_1\}$ and $\mathcal{B}_2 = \{x : x_1 > 0, x_2 < c_2\}$. This policy is also known as *fixed buffer* to distinguish it from a kanban policy where buffers are dynamically shifted as cards move in a cell, or *local control* because control of a station depends only on the number of jobs immediately downstream; i.e., station 1 is independent of the FG inventory.

CONWIP

The constant work-in-process (CONWIP) policy can be viewed as a kanban system with a single kanban cell (Muckstadt and Tayur 1991). For make-to-order or unlimited demand systems, CONWIP keeps the number of unfinished jobs in the system constant; for a make-to-stock system, the analogous policy is to keep WIP plus FG inventory constant. This policy is a special case of base stock with $c_1 = 0$.

3.3 Optimal Controls

It seems impossible to find a general solution to this control problem. For given parameter values λ, $\overline{\mu}_1$, $\overline{\mu}_2$, h and b, an optimal policy can be found numerically using dynamic programming; this is done in Section 5. One can also analyze a proposed policy π to determine conditions on the parameters under which it is optimal. The method used here to prove optimality is to establish bounds on $V^\pi(x + d_i) - V^\pi(x)$

using stochastic coupling arguments (see Hajek 1984 for another example of this technique). In special cases, these bounds can be used to verify that a policy is optimal.

For a given Markov policy π, let $\Delta_i(x) = V^\pi(x+e_i) - V^\pi(x)$ and $\Delta_{12}(x) = V^\pi(x-e_1+e_2) - V^\pi(x)$. The optimality condition (5) can be written

$$\Delta_1(x) \leq 0 \text{ iff } x \in \mathcal{B}_1 \tag{3.13}$$

$$\Delta_{12}(x) \leq 0,\ x \in \mathcal{B}_2 \tag{3.14}$$

$$\Delta_{12}(x) \geq 0,\ x \in \mathcal{I}_2 \text{ and } x_1 > 0. \tag{3.15}$$

Note that (14) and (15) only apply to points with $x_1 > 0$. We will write μ_i for $\overline{\mu}_i$ and μ_0 for λ when convenient. The subsections below make use of the following results. A cost of $c(x) = 1$ applied indefinitely yields a discounted cost of $1/\alpha$. Generalizing (5) to arbitrary policies gives

$$V^\pi(x) = \frac{1}{\Lambda(x)+\alpha}\left[c(x) + \sum_i \mu_i V^\pi(x+d_i)\right], \tag{3.16}$$

where $\Lambda(x)$ is the transition rate out of state x and the sum is taken over transitions i that are active in state x.

Let $(X(t), Y(t))$ be a coupled Markov process with state space

$$\mathbf{C} = \{(x,y) : x, y \in \mathbf{X} \text{ and } y = x, x+e_1, x+e_2, \text{ or } x-e_1+e_2\}, \tag{3.17}$$

where $X(t)$ and $Y(t)$ each have the same marginal distribution as the process of Section 1 under policy π with initial states $X(0)$ and $Y(0)$, and they share the same Poisson point processes of potential transitions. For example, in state $(x, x+e_1)$ the process transitions at rate μ_1 to $(x+e_1, x+e_1)$ if $x \in \mathcal{B}_1$ and $x+e_1 \in \mathcal{I}_1$, at rate μ_2 to $(x, x+e_2)$ if $x \in \mathcal{I}_2$ and $x+e_1 \in \mathcal{B}_2$, at rate μ_1 to $(x+e_1, x+2e_1)$ if $x, x+e_1 \in \mathcal{B}_1$, at rate μ_2 to $(x-e_1+e_2, x+e_2)$ if $x, x+e_1 \in \mathcal{B}_2$, and at rate λ to $(x-e_2, x+e_1-e_2)$. Only the first two transitions change the relative position $Y(t) - X(t)$. We say that the process merges at the first time at which $Y(t) = X(t)$; then $Y(s) = X(s)$ for all

$s \geq t$. We consider only policies that have the monotonicity properties of Fig. 3-3 (all optimal policies have these properties), thus limiting $Y(t) - X(t)$ to the values in $\mathbf{C}$. For example, if $x \in \mathcal{B}_1$ then $x - e_1 + e_2 \in \mathcal{B}_1$, and transitioning from $(x, x - e_1 + e_2)$ to $(x + e_1, x - e_1 + e_2)$ is impossible.

The usefulness of the coupled process lies in the fact that $\Delta_i(x)$ is the total cost resulting from a cost rate $c(Y(t)) - c(X(t))$ at time t, where $X(0) = x$ and $Y(0) = x + e_i$, and similarly for $\Delta_{12}(x)$, except that $Y(0) = x - e_1 + e_2$. The possible values of the cost rate are 0, 1, h, $h-1$, $-b$, and $-b-1$, corresponding to the values of $y - x$ for $(x, y) \in \mathbf{C}$.

3.3.1 No Inventory

Perhaps the simplest policy is to never hold inventory, releasing a job into station 1 only when there are backorders and station 2 is starved: $\mathcal{B}_1 = \{x : x_1 = 0,\ x_2 < 0\}$. Assume that station 2 is busy in states $(1,-1),(1,-2),(1,-3)\ldots$ so that the set of recurrent states is $\{(0,0),(0,-1),(0,-2),\ldots;(1,-1),(1,-2),(1,-3),\ldots\}$. For completeness, assume that the optimal control is used for station 2 in other, transient states. The task of checking optimality is much easier if only the recurrent states are checked. Although this condition is not sufficient for general Markov chains, it is for the chain defined here, as the following lemma shows.

Lemma 2 *For the no-inventory policy, (13) is implied by*

$$\Delta_1(0, x_2) \leq 0,\ x_2 < 0 \tag{3.18}$$

$$\Delta_1(1, x_2) \geq 0,\ x_2 < 0 \tag{3.19}$$

$$\Delta_1(0, 0) \geq 0. \tag{3.20}$$

Proof. We must show that $\Delta_1(x) \geq 0$ when $x_1 > 1$ or $x_2 > 0$ (all cost functions are for the no-inventory policy). First, establish the condition for the states $\{(0, x_2) : x_2 > 0\}$ using induction on x_2. To evaluate $\Delta_1(0, x_2)$, we must consider the states $(1, x_2)$. Since station 2 uses the optimal control in these states, monotonicity holds: station 2 is busy up to some x_2 and idle beyond. For states $(1, x_2) \in \mathcal{B}_2$, we will establish that $\Delta_1(0,x_2)$ is increasing. Initially $\Delta_1(0,0) \geq 0 \geq \Delta_1(0,-1)$. Assume that $\Delta_1(0,x_2) \geq \Delta_1(0,x_2-1)$ for some $x_2 > 0$. Since $(0,x_2+1) \in \mathcal{I}_1 \cap \mathcal{I}_2$ and $(1,x_2+1) \in \mathcal{I}_1 \cap \mathcal{B}_2$, there are two transitions for the coupled process corresponding to $\Delta_1(0,x_2+1)$, and (16) yields

$$\Delta_1(0,x_2+1) = \frac{1}{\lambda+\mu_2+\alpha}\left[1+\lambda\Delta_1(0,x_2)+\mu_2\Delta_2(0,x_2+1)\right]. \tag{3.21}$$

Similarly, using the transitions from $(x, x+e_2)$ gives

$$\Delta_2(0,x_2+1) = \frac{1}{\lambda+\alpha}\left[h+\lambda\Delta_2(0,x_2)\right]. \tag{3.22}$$

Since the cost rate for the coupled process is never more than h, $\Delta_2(0,x_2) \leq h/\alpha$; eliminating h in (22) gives $\Delta_2(0,x_2+1) \geq \Delta_2(0,x_2)$. Using this fact and the inductive hypothesis in (21),

$$\begin{aligned}\Delta_1(0,x_2+1) &\geq \frac{1}{\lambda+\mu_2+\alpha}\left[1+\lambda\Delta_1(0,x_2-1)+\mu_2\Delta_2(0,x_2)\right] && (3.23)\\ &= \Delta_1(0,x_2). && (3.24)\end{aligned}$$

Therefore $\Delta_1(0,x_2) \geq \Delta_1(0,0) \geq 0$ for $(1,x_2) \in \mathcal{B}_2$. For $(1,x_2) \in \mathcal{I}_2$, the last term in (21) is omitted and a simple induction argument shows that $\Delta_1(0,x_2+1)$ remains nonnegative.

Now consider $\Delta_1(x)$ for states with $x_1 > 0$. The corresponding coupled process has a cost rate of one until the first time t such that $X(t) = y$ for some y such that $y_1 = 1$ and $y_2 < 0$, or $y_1 = 0$ and $y_2 > 0$. Thereafter the cost rate is the same as for $\Delta_1(y)$. But $\Delta_1(y) \geq 0$ for such y; hence, $\Delta_1(x) \geq 0$. □

Using rather crude stochastic coupling bounds, the following parameter ranges are obtained from (14), (15), and (18-20).

Theorem 2 *The following conditions are sufficient for the no-inventory policy to be optimal:*

$$b\mu_2 \geq \mu_1 + \alpha, \tag{3.25}$$

$$1+\mu_2\min\{\phi,0\}+\frac{\mu_2}{\mu_1+\mu_2+\alpha}\left[1-\frac{\mu_2(b+1)}{\alpha}\right]\leq 0, \tag{3.26}$$

$$\text{and}\ \ 1+\frac{\lambda}{\mu_1+\mu_2+\alpha}\left[1-\frac{\mu_2(b+1)}{\alpha}\right]+\frac{\mu_2}{\lambda+\alpha}\left[h-\frac{\lambda(b+1)}{\alpha}\right]\geq 0, \tag{3.27}$$

$$\text{where}\ \ \phi=\left(\frac{\mu_2}{\mu_2+\alpha}\right)\frac{h}{\lambda+\alpha}+\left(\frac{\mu_1+\mu_2}{\mu_1+\mu_2+\alpha}\right)\left(\frac{\mu_2}{\mu_1+\mu_2}-\frac{\lambda}{\lambda+\alpha}\right)\frac{b+1}{\alpha}. \tag{3.28}$$

The proof of Theorem 1, which is a lengthy application of the coupled process, is given in the Appendix. Conditions (25-27) only hold for very large α, on the order of μ_2; one example is $h=2$, $b=4$, $\mu_1/\alpha=2$, $\mu_2/\alpha=1$, and $\lambda/\alpha=1/2$. This result is reasonable because a policy of not holding inventory is very shortsighted. It can be optimal only if the time horizon $1/\alpha$ is sufficiently short.

3.3.2 No FG Inventory

Now consider a policy that consists of the optimal control of station 1 and operating station 2 only when there are backorders: $\mathcal{B}_2=\{x:x_1>0,x_2<0\}$. A lemma again allows us to omit the transient states $\{x:x_2>0\}$ from the optimality check. Here the assumption that $h>1$ is critical; otherwise all WIP would be converted to FG.

Lemma 3 *For the no-FG inventory policy, $\Delta_{12}(x_1,0)\geq 0$ implies (15).*

Proof. Again using the coupled process, $\Delta_{12}(x)$ for $x_2>0$ corresponds to the cost rate $h-1$ until $X_2(t)=0$, then the same costs as $\Delta_{12}(y_1,0)$ for some y_1. Both of these costs are nonnegative. □

Theorem 3 *The following conditions are sufficient for the no-FG inventory policy to be optimal:*

$$\mu_2\geq\lambda\left(\frac{\lambda+\alpha}{\alpha}\right)\left(\frac{b+1}{h-1}\right)-\lambda-\alpha, \tag{3.29}$$

$$h-1+\frac{\lambda}{\lambda+\mu_2+\alpha}(-b-1+\lambda V_5+\mu_2V_3)\geq 0, \tag{3.30}$$

$$\text{and}\ \ h-1+\lambda V_2\geq 0, \tag{3.31}$$

$$\text{where}\ \ V_3 \;=\; \theta_1[\lambda+\mu_2+\alpha+\lambda(1+\lambda V_5)] \tag{3.32}$$

$$V_5 = -(b+1)/\alpha \tag{3.33}$$

$$\theta_1 = \frac{1}{(\lambda+\alpha)(\lambda+\mu_2+\alpha)-\lambda\mu_2} \tag{3.34}$$

$$V_2 = \theta_2\left[-b-1+\lambda V_5+\frac{\mu_2}{\lambda+\alpha}\left(1+\frac{\lambda(1+\lambda V_5)}{\lambda+\mu_2+\alpha}\right)\right] \tag{3.35}$$

$$\theta_2 = \frac{(\lambda+\alpha)(\lambda+\mu_2+\alpha)}{(\lambda+\alpha)(\lambda+\mu_2+\alpha)^2-\lambda\mu_2^2}. \tag{3.36}$$

In particular, it is optimal for sufficiently large μ_2.

The proof in the Appendix suggests that other conditions for optimality could be obtained if desired.

3.3.3 Non-Idling at Station 2

In some sense the opposite of the previous policy is to never idle at station 2: $\mathcal{B}_2 = \{x : x_1 > 0\}$ and optimal control of station 1. Reversing the cost structure so that it is more costly to hold WIP than FG makes this policy optimal.

Theorem 4 *If* $h < 1$ *then non-idling at station 2 is optimal.*

Proof. Suppose a policy π includes idling at station 2 in some state x with $x_1 > 0$. For this initial state x, construct a policy π' that is identical to π except that the start time of the next job at station 2 is moved up to zero. Their cost rates differ by $h-1$ or $-b-1$, both nonpositive, for the time interval between the completion of the next job at station 2 under π' and under π. Therefore $V^{\pi'}(x) < V^{\pi}(x)$ and only policies that are non-idling at station 2 are optimal. □

Theorem 3 can be strengthened to the case $h = 1$ using a more elaborate proof. A situation where $h \leq 1$ might occur when the benefits of just-in-time manufacturing are incorporated as additional WIP holding costs.

When $h > 1$ the decision of whether to operate station 2 depends on the likelihood of incurring FG holding costs as a result. If the optimal control for station 1 prevents FG inventory from being held, then non-idling is optimal at station 2.

Theorem 5 *If* $\mathcal{B}_1 \cap \{x : x_2 \geq 0\} = \emptyset$ *then it is optimal not to idle at station 2 in all states that are recurrent under some station 2 control.*

Proof. Let $\overline{\mathcal{B}}_1 = \{x : x \in \mathcal{B}_1 \text{ or } x - d_1 \in \mathcal{B}_1\}$. Recall from Section 1 that $\mathcal{B}_1$, and also $\overline{\mathcal{B}}_1$, consists of all states below a switching curve $s_1(x_1)$ with slope $s_1'(x_1) \leq -1$. Thus, regardless of the station 2 control the system cannot leave $\overline{\mathcal{B}}_1$. Other states cannot be recurrent because the transition d_1 cannot occur outside of $\overline{\mathcal{B}}_1$. Now, for x recurrent with $x_1 > 0$, we must have $x \in \overline{\mathcal{B}}_1$ and $x_2 < 0$. Hence, from the initial state x, $X(t) \in \overline{\mathcal{B}}_1$ for all t and the coupled process cost rate for $\Delta_{12}(x)$ is either $-b-1$ or $-b$. This implies that $\Delta_{12}(x) \leq 0$ and non-idling is optimal at station 2. □

The above proof also generalizes the proof of (14) for the no-inventory policy of Theorem 1.

3.3.4 Base Stock Policies are Never Optimal

Despite their popularity in inventory systems, base stock policies are never optimal for this problem because they can accumulate large amounts of WIP that, due to the capacity constraint, will remain in the system for long periods of time.

Theorem 6 *The base stock policy of Section 2 is not optimal.*

Proof. Consider a base stock policy with $\mathcal{B}_1 = \{x : x_1 + x_2 < c_1 + c_2\}$ and $\mathcal{B}_2 = \{x : x_1 > 0,\ x_2 < c_2\}$. For the coupled process associated with $\Delta_1(x)$, let T_1 be the time of first departure from a cost rate of one, T_m be the merge time, $p(x) = \Pr\{T_1 < T_m\}$, V_1 be the discounted cost until T_1, and V_2 be the discounted cost from T_1 to T_m given that $T_1 < T_m$. Then

$$\Delta_1(x) = V_1 + p(x)V_2. \tag{3.37}$$

We will show that, for $x = (x_1, c_1 + c_2 - x_1 - 1)$ and $x_1 \to \infty$, $p(x) \to 0$. Since $V_1 > 0$ and $V_2 \leq h/\alpha$, this implies that $\Delta_1(x) \geq 0$ for some $x = (x_1, c_1 + c_2 - x_1 - 1)$ and the base stock policy is not optimal.

From the initial state x, the event $\{T_1 < T_m\}$ requires $X_1(t) = 0$ for some $t < T_m$; hence, no merge can occur during the first x_1 potential transitions of X. Let N_0, N_1, and N_2 be the number of the first x_1 transitions that are of type λ, μ_1, and μ_2, respectively, and T be the time of the x_1-th transition. Then

$$p(x) \leq \Pr\{T_m > T\} \tag{3.38}$$

$$\leq \Pr\{N_1 - N_0 < 1\}, \tag{3.39}$$

since $N_1 - N_0 \geq 1$ implies that the first occurence of $X_1(t) + X_2(t) = c_1 + c_2$ was at some time $t \leq T$ and the process merged at that time. But (N_0, N_1, N_2) have a multinomial distribution with x_1 trials and probabilities $(\lambda/\Lambda, \mu_1/\Lambda, \mu_2/\Lambda)$, so that $\mu_1 > \lambda$ implies that $\Pr\{N_1 - N_0 < 1\} \to 0$ as $x_1 \to \infty$. □

3.4 An Equivalent Make-to-Order System

The production/inventory system of Section 1 is not a tandem queue in the usual sense because backorders ($x_2 < 0$) are allowed; instead, it has been viewed as a queueing network with assembly where demands enter a queue and are joined with FG (Mitra and Mitrani 1991). However, if total inventory is bounded, we can transform this system into an equivalent make-to-order system. Consider only policies π and initial states $X(0) = (c_1, c_2)$ for which $X_1(t) + X_2(t) \leq c_1 + c_2$, or equivalently, $S_1(t) \leq D(t)$, with probability one. Define

$$Z_1(t) = c_1 + c_2 - X_1(t) - X_2(t) \tag{3.40}$$

$$Z_2(t) = X_1(t). \tag{3.41}$$

Then Z is a make-to-order system (a tandem queue) with infinite first buffer; rates λ, μ_1, and μ_2; and cost function $c^Z(z) = c^X(z_2, c_1 + c_2 - z_1 - z_2) = z_2 + h(c_1 + c_2 - z_1 - z_2)^+ + b(c_1 + c_2 - z_1 - z_2)^-$. All statistics of X can be recovered from Z by solving (40) and (41) for X. The distinguishing feature of a make-to-stock system is seen to be its concave (as opposed to linear) cost function, not its dynamics. Such a transformation can be made for n-stage systems as well.

Because of this equivalence, any method that obtains the state probability distribution for tandem queues can be used to evaluate a make-to-stock system under the corresponding control policy. The fixed buffer mechanism provides a good example. A fixed buffer of size c_1 (with infinite first buffer) for Z gives the policy $\mathcal{B}_1 = \{x : x_1 + x_2 < c_1 + c_2,\ x_1 < c_1\}$ and $\mathcal{B}_2 = \{x : x_1 > 0\}$ for X, where c_2 is arbitrary. This control mechanism, motivated by the linear cost function of make-to-order systems, will not always be appropriate for the concave cost function. However, when it is reasonable, the following method could be used to obtain a nearly optimal policy of this form. For a given c_1, generate a steady-state distribution using one of the approximations noted in Smith and Daskalaki (1988). Then, given c_2, compute the appropriate cost measure. Use an optimization scheme to find the best c_1 and c_2.

3.5 Dynamic Programming Computational Results

Dynamic programming value iteration was used on a truncated state space to compute the optimal policy for several cases. For undiscounted problems, the average cost per unit time g/Λ is reported; in the discounted case, the cost $V(0,0)$ is reported. Up to 2000 iterations were required to achieve four digit accuracy. Larger and larger state spaces were tested until the results were insensitive to increasing the state space. The largest state space required was 21 by 43. To avoid solving large linear systems, value iteration was also used to evaluate candidate policies. A coordinate search algorithm was employed to find the best parameters (c_1, c_2) for a given type of policy. The algorithm assumes convexity of V^{π}; to check this assumption different initial values of (c_1, c_2) were tried and gave the same results.

Three undiscounted cases are presented in Table 3.1. Case 1 is a balanced system with a utilization of 5/6. In case 2 station 1 is faster, while in case 3 station 2 is faster. As is known for a variety of manufacturing systems, it is better to have the faster machine downstream so that the bottleneck is upstream (case 3). Among the suboptimal policies, base stock performs very well for cases 1 and 3. When the utilization of the upstream machine is low, as in case 2, stockpiling WIP when there are many backorders is unnecessary and the base stock policy does not perform as

Table 3.1: Gain per Unit Time Under Various Policies ($\lambda = 1$, $h = 2$, $b = 4$, $\alpha = 0$)

Policy	Case 1 $\mu_1 = \mu_2 = 1.2$	Case 2 $\mu_1 = 2$, $\mu_2 = 1.2$	Case 3 $\mu_1 = 1.2$, $\mu_2 = 2$
Optimal	21.50	14.88	11.48
Best Base Stock	21.57	15.9	11.56
Best Kanban	22.1	15.3	11.63
Best Fixed Buffer	23.7	16.4	11.83
Revised Base Stock	21.54	15.2	11.56

Table 3.2: Suboptimality and Stock Levels for Various Policies ($\lambda = 1$)

		Base Stock		Kanban		Fixed Buffer	
Case	$(\mu_1, \mu_2, h, b, \alpha)$	Subopt	(c_1, c_2)	Subopt	(c_1, c_2)	Subopt	(c_1, c_2)
1	1.2,1.2,2,4,0.0	0.3	4,8	3.0	6,8	10	12,7
2	2.0,1.2,2,4,0.0	7	1,6	3	1,6	10	5,6
4	1.2,1.2,2,4,0.1	1.2	1,3	0.8	2,3	2.2	5,3
5	2.0,2.0,2,4,0.0	0.9	1,2	5.5	1,2	17	3,1
6	2.0,1.2,1,1,0.0	24	1,3	6	1,4	15	4,4

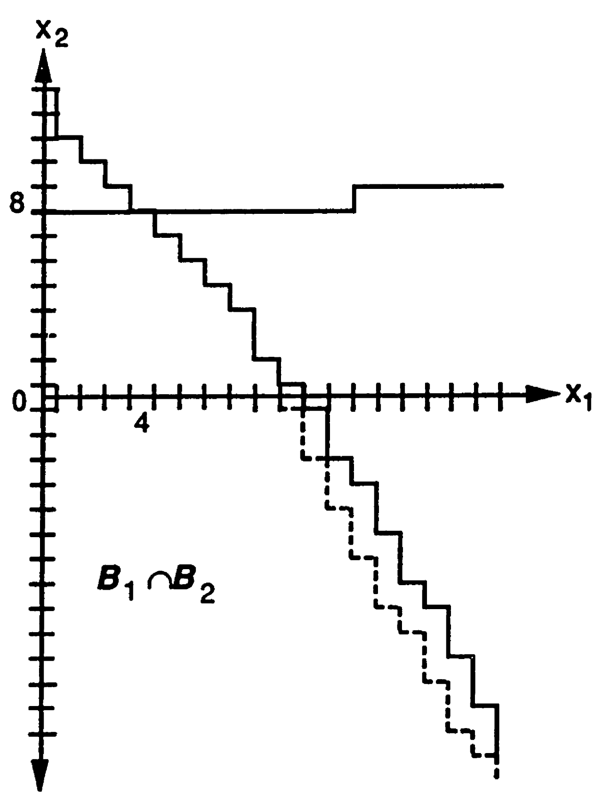

Figure 3-5: Optimal Policy for Case 1 (dashed line is revised base stock)

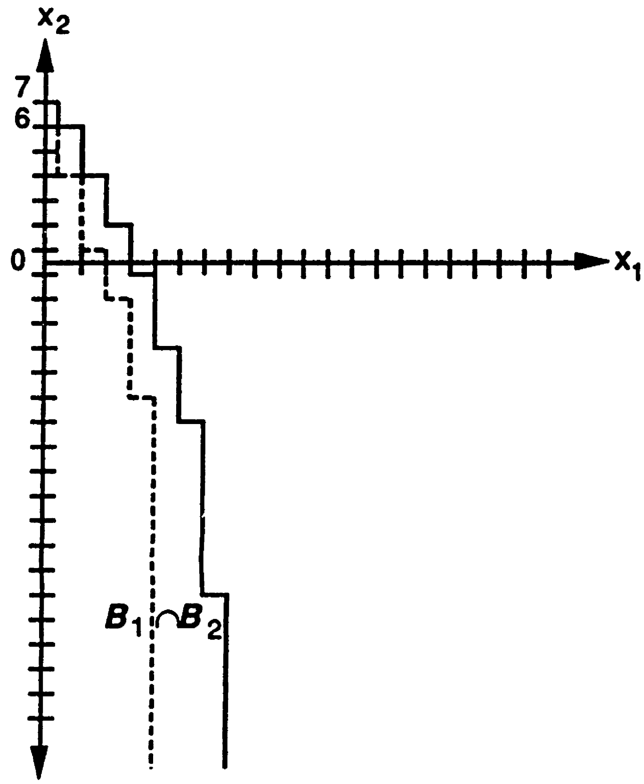

Figure 3-6: Optimal Policy for Case 2 (dashed line is case 6)

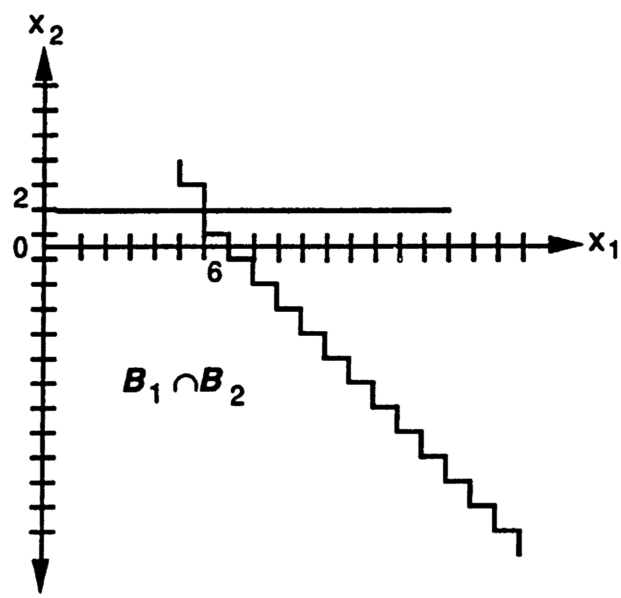

Figure 3-7: Optimal Policy for Case 3

well as kanban. These results are also explained by Figs. 3-5, 3-6, and 3-7, showing the optimal busy regions. Case 3 has the 45-degree line characteristic of base stock policies (although, by Theorem 5, this line must turn downward for large x_1), case 1 is nearly base stock, and case 2 is quite different.

Table 3.2 compares the policies in greater detail for additional test cases. The parameters c_1 and c_2 are the hedging point; i.e., the target levels of WIP and FG, respectively. Case 4 illustrates that much less stock is held and kanban is preferable when discounting is present (compare case 1). Case 5 shows that little stock is held when utilizations are low. In case 6, the combination of a faster downstream station and relatively high WIP holding costs creates a situation where significant FG but little WIP is held. The optimal policy for case 6, shown in Fig. 3-6, allows up to six units of FG to be held but has an even steeper switching curve than case 2.

The good performance of base stock policies suggests a way of quickly generating a nearly optimal policy: search over base stock policies to find the best one, evaluating V^π for each policy. Then revise the station 1 switching curve for this policy using V^π by setting $x \in \mathcal{B}_1$ if $\Delta_1(x) \leq 0$. Performance of this "revised base stock" policy is included in Table 3.1. As shown in Fig. 3-6, for case 1 this policy has nearly the same busy region as the optimal policy. Although we have used a slow, iterative algorithm to compute V^π in this study, a rapid solution should be possible because of the sparse structure of the linear system (16).

3.6 Appendix

<u>Proof of Theorem 1.</u>

In light of Lemma 1, it suffices to show (18-20) and $\Delta_{12}(1, x_2) \leq 0$, $x_2 < 0$.

(18). Consider the coupled process associated with $\Delta_1(0, x_2)$, $x_2 < 0$. Partition its possible states into the configurations listed in Table 3.3. It moves through these configurations according to Fig. 3-8, where the quantity next to each arc is the probability of moving along that arc. Let T_i be the time of first departure from configuration i, T_m be the merge time (merge occurs upon entering configuration 4), and $V(s, t)$ be

Table 3.3: Configurations of the Coupled Process in (18)

Configuration	States (X, Y)	Cost Rate	Departure Rate Λ
1	$(x, x+e_1)$	1	$\mu_1+\mu_2$
2	$(x, x+e_2)$	$-b$	$0 \leq \Lambda \leq \mu_1$
3	$(x, x-e_1+e_2)$	$-b-1$	$\mu_2 \leq \Lambda \leq \mu_1+\mu_2$
4	(x, x)	0	0

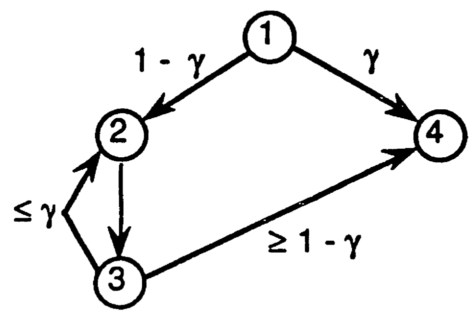

Figure 3-8: Transition Diagram for the Coupled Process in (18); $\gamma = \mu_1/(\mu_1+\mu_2)$.

Table 3.4: Configurations of the Coupled Process in (19)

Configuration	States (X, Y)	Cost Rate	Departure Rate Λ
1	$(x, x+e_1)$, $x_1 = 1$	1	μ_2
2	$(x, x+e_1)$, $x_1 = 0$	1	$\mu_2 \leq \Lambda \leq \mu_1+\mu_2$
3	$((0,0), e_2)$	h	λ
4	$(x, x+e_2)$, $x_2 < 0$	$-b$	-

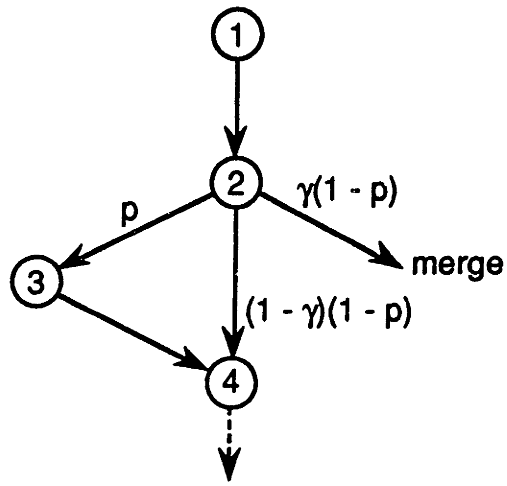

Figure 3-9: Transition Diagram for the Coupled Process in (19); $\gamma = \mu_1/(\mu_1+\mu_2)$.

Table 3.5: Configurations of the Coupled Process in (20)

Configuration	States (X, Y)	Cost Rate	Departure Rate Λ
1	$((0,0), e_1)$	1	$\lambda + \mu_2$
2	$((0,0), e_2)$	h	λ
3	$(x, x + e_1)$, $x_2 < 0$	1	$\mu_1 + \mu_2$
4	$(x, x + e_2)$, $x_2 < 0$	$-b$	-

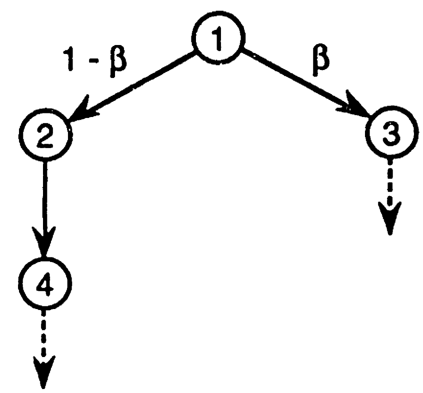

Figure 3-10: Transition Diagram for the Coupled Process in (20); $\beta = \lambda/(\lambda + \mu_2)$.

the cost incurred by the coupled process in the period $(s, t]$, given that the process has not merged by s. Then

$$\begin{aligned} V(T_1, T_3) &= E\left[-b\int_0^{T_2-T_1} e^{-\alpha t}dt \;-\; (b+1)e^{-\alpha(T_2-T_1)}\int_0^{T_3-T_2} e^{-\alpha t}dt\right] \\ &\le E\left[-\frac{b}{\alpha}(1 - e^{-\alpha(T_2-T_1)}) - \frac{b+1}{\mu_1+\mu_2+\alpha}e^{-\alpha(T_2-T_1)}\right] \\ &\le \max_{0\le t\le\infty}\left[-\frac{b}{\alpha}(1 - e^{-\alpha t}) - \frac{b+1}{\mu_1+\mu_2+\alpha}e^{-\alpha t}\right] \\ &= -\min\left\{\frac{b}{\alpha}, \frac{b+1}{\mu_1+\mu_2+\alpha}\right\}. \end{aligned} \tag{A.1}$$

The first integral was evaluated exactly; the second was bounded using the maximum departure rate. From (16) and Fig. 3-8,

$$\Delta_1(0, x_2) \le \frac{1}{\mu_1+\mu_2+\alpha}\left[1 - \mu_2 \min\left\{\frac{b}{\alpha}, \frac{b+1}{\mu_1+\mu_2+\alpha}\right\}\right], \tag{A.2}$$

where we have omitted $V(T_3, T_m) < 0$. The right side is nonpositive when (25) holds, i.e., (25) implies (18).

(19). For $\Delta_1(1,x_2)$, $x_2 < 0$, the relevant configurations are listed in Table 3.4, with the transitions in Fig. 3-9. Here p is the probability that $X(T_2^-) = (0,0)$ is the last state visited in configuration 2. A crude bound after entering configuration 4 is $V(T_3,T_m) \geq -(b+1)/\alpha$ (define $T_3 = T_2$ if configuration 3 is not visited). Denote the discount factor while in configuration i by $\alpha_i = E\{\exp[-\alpha(T_i - T_{i-1})]\}$. Using the maximum and minimum rates from Table 3.4 where applicable, and applying (16) repeatedly gives

$$\begin{aligned}
\Delta_1(1,x_2) &= V(0,T_1) + \alpha_1 \Big\{ V(T_1,T_2) + \alpha_2 \Big[pV(T_2,T_3) + p\alpha_3 V(T_3,T_m) \\
&\quad + \frac{\mu_2}{\mu_1+\mu_2}(1-p)V(T_3,T_m) \Big] \Big\} \\
&\geq \frac{1}{\mu_2+\alpha} \Big\{ 1 + \frac{\mu_2}{\mu_1+\mu_2+\alpha} + \frac{\mu_2^2 ph}{(\mu_2+\alpha)(\lambda+\alpha)} \\
&\quad - \left(\frac{\mu_2(\mu_1+\mu_2)}{\mu_1+\mu_2+\alpha}\right) \left[(1-p)\left(\frac{\mu_2(b+1)}{(\mu_1+\mu_2)\alpha}\right) + p\left(\frac{\lambda(b+1)}{(\lambda+\alpha)\alpha}\right) \right] \Big\}. \qquad \text{(A.3)}
\end{aligned}$$

Taking the minimum over $0 \leq p \leq 1$,

$$\Delta_1(1,x_2) \geq \frac{1}{\mu_2+\alpha}\left\{ 1 + \mu_2 \min\{\phi,0\} + \frac{\mu_2}{\mu_1+\mu_2+\alpha}\left[1 - \frac{\mu_2(b+1)}{\alpha}\right]\right\}, \qquad \text{(A.4)}$$

$$\text{where } \phi = \left(\frac{\mu_2}{\mu_2+\alpha}\right)\frac{h}{\lambda+\alpha} + \left(\frac{\mu_1+\mu_2}{\mu_1+\mu_2+\alpha}\right)\left(\frac{\mu_2}{\mu_1+\mu_2} - \frac{\lambda}{\lambda+\alpha}\right)\frac{b+1}{\alpha}.$$

The right side of (A.4) is nonnegative when (26) holds.

(20). For $\Delta_1(0,0)$, a partial list of configurations and their transitions are given in Table 3.5 and Fig. 3-10. The first transition gives

$$\Delta_1(0,0) = \frac{1}{\lambda+\mu_2+\alpha}\left[1 + \lambda\Delta_1(0,-1) + \mu_2\Delta_2(0,0)\right]. \qquad \text{(A.5)}$$

Change (A.2) to a lower bound using $V(T_2,T_m) \geq -(b+1)/\alpha$, giving

$$\Delta_1(0,-1) \geq \frac{1}{\mu_1+\mu_2+\alpha}\left[1 - \frac{\mu_2(b+1)}{\alpha}\right]. \qquad \text{(A.6)}$$

Table 3.6: Configurations of the Coupled Process in (15)

Configuration	States (X,Y)	Cost Rate	Departure Rate Λ
1A	$(x, x-e_1+e_2)$, $x_2=0$	$h-1$	λ
1B	$(x, x+e_2)$, $x_2=0$	h	λ
2A	$(x, x-e_1+e_2)$, $x_2=-1$	$-b-1$	$\lambda+\mu_2$
2B	$(x, x+e_2)$, $x_2=-1$	$-b$	$\lambda+\mu_2$
3	$(x, x+e_1)$, $x_2=0$	1	$\lambda \le \Lambda \le \lambda+\mu_1$
4	$(x, x+e_1)$, $x_2=-1$	1	$\lambda \le \Lambda \le \lambda+\mu_1+\mu_2$
5	(x,y), $x_2<-1$	-	-

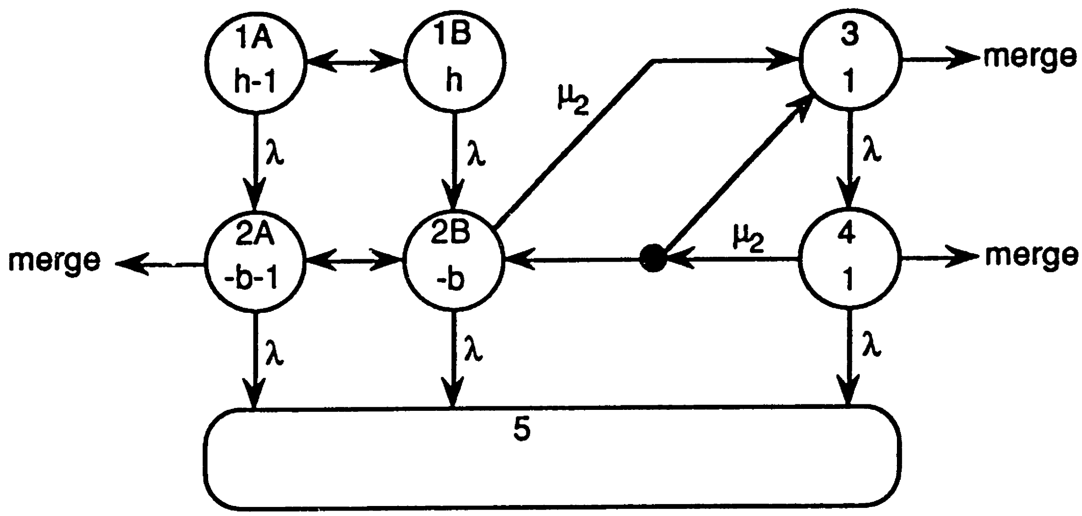

Figure 3-11: Transition Diagram for the Coupled Process in (15)

Expanding $\Delta_2(0,0)$ using the same lower bound, (A.5) can be written

$$\Delta_1(0,0) \ge \frac{1}{\lambda+\mu_2+\alpha}\left\{1+\frac{\lambda}{\mu_1+\mu_2+\alpha}\left[1-\frac{\mu_2(b+1)}{\alpha}\right]+\frac{\mu_2}{\lambda+\alpha}\left[h-\frac{\lambda(b+1)}{\alpha}\right]\right\}. \tag{A.7}$$

The right side is nonnegative when (27) holds.

$\Delta_{12}(1,x_2)$, $x_2<0$. The cost rate of the corresponding coupled process is negative, $-b$ or $-b-1$, until merging. Hence, $\Delta_{12}(1,x_2)\le 0$. □

Proof of Theorem 2.

By assumption, (13) holds. The proof of (14) is identical to that of Theorem 3 except that $x_2<0$ and the cost rates differ by $-b-1<0$, since x_2 can only decrease until the next job is processed at station 2. In light of Lemma 2, it remains

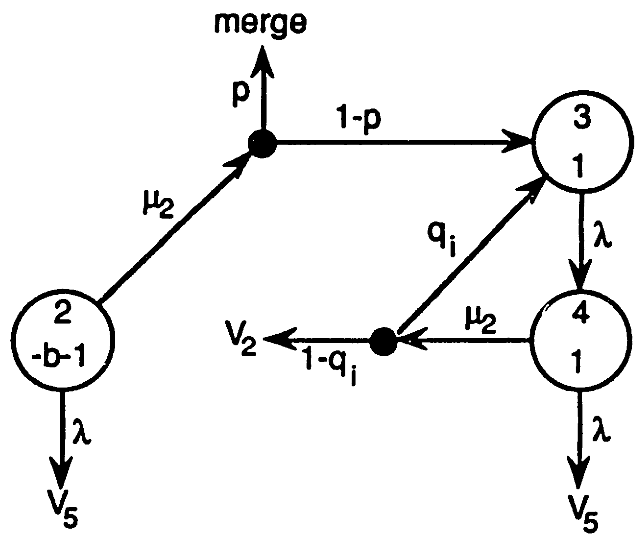

Figure 3-12: Lower Bound Process

to show (15) for $x_2 = 0$. Consider the configurations in Table 3.6 and the transitions in Fig. 3-11 for the coupled process corresponding to $\Delta_{12}(x_1, 0)$. Here the cost rate appears in each configuration and the transition rate appears next to each arc that has a constant rate (recall that the configurations are collections of states, so that the configuration does not evolve as a Markov process). The definition of these configurations is motivated by the fact that, as μ_2 increases, the probability of entering configuration 5 ($x_2 < -1$) decreases and more time is spent in configuration 3 with a positive cost. Adopt the notation $\mathcal{C}_i = \{(x, y) : (x, y) \in \text{configuration } i\}$, $V(x, y) = V^\pi(y) - V^\pi(x)$ (the cost incurred by the coupled process under the no-FG policy π), and $V_i = \min_{(x,y)\in\mathcal{C}_i} V(x, y)$. Then

$$\Delta_{12}(x_1, 0) = V((x_1, 0), (x_1 - 1, 1)) \geq V_{1A} \geq \frac{1}{\lambda + \alpha}(h - 1 + \lambda V_2). \qquad \text{(A.8)}$$

To bound V_2, we will approximate Fig. 3-11 with the process of Fig. 3-12. For a given state $(x, y) \in \mathcal{C}_2$, let p be the probability of merging before returning to $\mathcal{C}_2$, given that the system leaves $\mathcal{C}_2$ by a μ_2 transition, and let q_i be the probability of returning to configuration 3 upon leaving configuration 4 (by a μ_2 transition) for the ith time. We claim that, for this p and q_i, the approximate process is a lower bound in the sense that $V(x, y) \geq V_2'$, where V_2' is the value of the approximate process in state 2. Applying this bound for all states in C_2 gives

$$V_2 \geq V_2'. \tag{A.9}$$

To establish that the approximate process is a lower bound, make the following sequence of changes to Fig. 3-11, each of which decreases (or does not change) V_2. Eliminate configuration 1, replace the cost upon entering configurations 2B and 5 with their lower bounds V_2 and V_5, change the cost in 2B to $-b-1$, move the merges after configurations 3 and 4 to after configuration 2B (with an equivalent probability of merging; the only effect is to reduce the time spent in 3 and 4), and combine 2A and 2B into 2 (the equivalent probability of merging is p, defined above).

Analyzing Fig. 3-12,

$$V_2' = \frac{1}{\lambda + \mu_2 + \alpha}[-b - 1 + \lambda V_5 + (1-p)\mu_2 V_3']. \tag{A.10}$$

To establish bounds, use $V_5 \geq -(b+1)/\alpha$ and consider the values $p = 0$ and 1, one of which will be a worst case. If $p = 1$,

$$\Delta_{12}(x_1, 0) \geq \frac{1}{\lambda + \alpha}\left(h - 1 - \frac{\lambda}{\lambda + \mu_2 + \alpha}\left[b + 1 + \frac{\lambda(b+1)}{\alpha}\right]\right). \tag{A.11}$$

The right side is nonnegative if (29) holds.

Now consider $p = 0$. Either $q_i = 0$ for all i or $q_i = 1$ for all i is a worst case (other than variation in the transition probabilities q_i, the system is Markov, so that the minimal cost is achieved by a constant q). If $q_i = 1$,

$$V_3' = \frac{1}{\lambda + \alpha}\left[1 + \frac{\lambda}{\lambda + \mu_2 + \alpha}(1 + \lambda V_5 + \mu_2 V_3')\right]. \tag{A.12}$$

Solving for V_3' and dropping the prime notation gives (32). Combining (A.8-10) and requiring the right side to be nonnegative gives (30). Hence, for the case $q_i = 1$, (30) implies (15). If $q_i = 0$, V_2 replaces V_3' in the right side of (A.12). In light of (A.9), we can replace V_2 with V_2', substitute into (A.10), and solve for V_2' to obtain the lower bound (35). Since (31) requires the right side of (A.8) to be nonnegative, it implies (15) for the case $q_i = 0$. □

Chapter 4

Multi-Product Production/Inventory System

In a make-to-stock production facility with multiple products, one of the goals of the scheduling policy is to regulate finished goods inventory. Too small of an inventory risks incurring backorder or lost sales costs, while too large of an inventory increases holding costs. The target inventory level, called the base or safety stock, is vitally linked to randomness in the system and capacity constraints that limit the ability to respond to unexpected demand. Accordingly, a realistic model of the make-to-stock system should include queueing effects. The queueing framework combines the dynamic stochastic nature of the scheduling problem, often studied in inventory systems, with the capacity constraint, usually dealt with through deterministic production scheduling.

Although make-to-order environments, where production occurs after customer orders are received, have been analyzed extensively as queueing control problems (see, for example, Klimov 1974), little work has been done on the problem of scheduling a multiclass make-to-stock queue. Wein (1991) develops a scheduling policy for the make-to-stock system based on a heavy traffic approximation that results in a Brownian motion control problem. Zheng and Zipkin (1990) and Zipkin (1992) propose and analyze a simple "longest queue" policy that is optimal for a system with identical product classes operating under independent base stock policies. Zipkin

(1990) proposes yet another policy, attempting a compromise between the very different Brownian and longest queue policies. While promising computational results have been reported for the Brownian and longest queue policies, no performance guarantees are available and both have obvious deficiencies in the structure of the control.

This paper develops and tests several new scheduling policies. The system considered is a multiclass $M/M/1$ make-to-stock queue. Preemptive resume scheduling is allowed and there are no set-up costs or times when switching classes. In the backorder version of the problem, the objective is to minimize holding and backorder costs. A lost sales problem is also considered, where demands that cannot be met from inventory are lost and a cost incurred. We focus on a long-run average cost criterion, but also consider discounted costs.

Three index policies are considered, where an index is computed for each class as a function of the inventory in that class. The class with the smallest index is served; if all indices are positive, the server is idle. Index policies have the advantage of being computationally tractable even for a large number of classes. One of the index policies is taken from Zipkin (1990). The most innovative, and one that performs well, is a "restless bandit" index, defined for a general problem in Whittle (1988). This index has the property that it is asymptotically optimal as the number of classes goes to infinity (and the utilization of each class goes to zero). We also show how to compute the Gittens index for this problem, neglecting its "restlessness," and point out a connection between the Gittens and restless bandit indices. A third index is developed by computing the value function for the system under a naive random policy.

We will show that index policies perform well at determining which class to serve, but do poorly at deciding when to idle. The problem is that each index is computed without knowledge of the other classes, and hence, without knowing the system utilization. Several other approximations are proposed for when to idle. One method decomposes the system into single-class subproblems with the same utilization as the original system. Another aggregates the system into a single product class. The most elegant, and the most accurate, idling decision is derived using a diffusion approximation. We develop a Brownian motion (diffusion) policy for the lost sales case, complementing the backorder case treated in Wein (1991). A fourth idleness policy is derived by computing the inventory distribution assuming that the longest queue policy is used, and then decomposing into single-class subproblems.

Numerical results are presented that compare all of the proposed policies with optimal policies for two- and three-product problems. Combinations of index and idleness policies are found that perform well, particularly for the lost sales problem. The structure of optimal policies is also investigated.

The rest of the paper is organized as follows. Section 1 formulates the problem mathematically and discusses the structure of scheduling policies. Index and idleness policies are introduced in Section 2. The two policies of the most theoretical interest, which also turn out to be quite effective, the restless bandit index and the Brownian policy, are derived in Sections 3 and 4, respectively. All other policies are developed in Section 5. Numerical results are presented in Section 6. Some suggestions for further research are made in Section 7.

4.1 Dynamic Scheduling Problem

Consider a multiclass, make-to-stock M/M/1 queueing system: a server can produce K different classes of items; each finished item is placed in its respective inventory, $X_k(t)$ for class k, $k = 1, \ldots, K$ at time t; this inventory services an exogenous demand. In the backorder version of the problem, demand that cannot be met from inventory is backordered and recorded as negative inventory. In the lost sales problem, class k demands that occur when $X_k(t) = 0$ are ignored and a cost incurred.

The demands for each class are independent Poisson processes with rates λ_k and the service times for class k items are independent and exponentially distributed with mean $m_k = 1/\mu_k$. In Section 4 we will briefly consider general interdemand and service time distributions. For the backorder problem, stability of the system requires that $\rho = \sum \rho_k < 1$, where $\rho_k = \lambda_k/\mu_k$ (all indices range over $1, \ldots, K$ unless otherwise noted).

The scheduling decision is whether to produce product $1, \ldots, K$ or to idle at each time t. An admissible scheduling policy π is a function $\zeta(X,t)$ that takes on the values $0, 1, \ldots, K$ (zero denoting idle) and is nonanticipating with respect to X. Let Π denote the class of admissible policies. Production of an item can be interrupted and resumed; no set-up costs or times are incurred when switching from one class to another. Because the system is memoryless, a Markov policy, depending only on the current state $X(t) = (X_1(t), \ldots, X_K(t))$, will be optimal. Consequently, the policy space can be enlarged to include multiple servers or the allocation of partial service effort without changing the problem.

The objective is to minimize holding and backorder costs, incurred at the rate

$$c^{BO}(x) = \sum (h_k x_k^+ + b_k x_k^-)$$

in state x, for the backorder problem, or holding and lost sales costs,

$$c^{LS}(x) = \sum (h_k x_k^+ + s_k 1_{\{x_k = 0\}}),$$

for the lost sales problem. Note that s_k is the cost of a stockout. The expected cost per lost sale is s_k/λ_k. The infinite-horizon cost, discounted at the rate $\alpha > 0$, is

$$V^\pi(x) = E_x \int_0^\infty e^{-\alpha t} c(X(t)) dt.$$

Here E_x denotes expectation given the initial state $X(0) = x$ and policy π. Let $\overline{\mu} = \max_k\{\mu_k\}$ and $\Lambda = \sum \lambda_k + \overline{\mu} + \alpha$. The optimal cost function, $V(x) = \min_{\pi \in \Pi} V^\pi(x)$, satisfies the dynamic programming optimality equations

$$V(x) = \mathbf{T}V(x), \tag{1.1}$$

$$\mathbf{T}V(x) = \frac{1}{\Lambda}\Big[c(x) + \sum \lambda_k V(x - e_k)$$
$$+ \min\{\overline{\mu}V(x), \min_k\{\mu_k V(x + e_k) + (\overline{\mu} - \mu_k)V(x)\}\}\Big], \tag{1.2}$$

where e_k is the unit vector with kth component equal to one. For the lost sales problem, replace $x - e_k$ with x in (1.2) when $x_k = 0$. An undiscounted, long-run average cost criterion will also be considered. In this case $\alpha = 0$; the optimal average cost rate (gain) g and relative value function $V(x)$ satisfy

$$V(x) + g = \mathbf{T}V(x), \tag{1.3}$$

where we arbitrarily set $V(0) = 0$.

The following characterization of Markov policies will be useful. Let $\mathcal{B}_k$ be the set of states in which class k is produced and $\mathcal{I}$ the set in which the server is idle. We consider only policies having the monotonicity property that $x + ae_k \in \mathcal{B}_k$ if and only if $a < A$, for some $-\infty \leq A \leq \infty$. For such policies, each pair $\mathcal{B}_j, \mathcal{B}_k$ can be separated by a *switching surface* that is nondecreasing (when viewed as a function of any $K - 1$ of the state variables x_k). Roughly speaking, the intersection of all of the switching surfaces forms a *switching curve*. To be more precise, define the switching curve as a sequence of points $\{x_n, n = 1, \ldots, N\} \in \mathbf{Z}^k$ such that

1. $x^{n+1} = x^n + e_k$, where $x^n \in \mathcal{B}_k$,

2. $x^N \in \mathcal{I}$, and

3. for every $x^n \in \mathcal{B}_k$ and $j \neq k$, either $x^n - e_j \in \mathcal{B}_j$ or, for the lost sales problem, $x_j^n = 0$.

The final point in the switching curve is called the ***hedging point***, denoted $x^* = x^N$; its components x_k^* are the target or base *stock levels*. Starting from any point not in $\mathcal{I}$, the system without arrivals moves toward the switching curve, then along it to the hedging point, and then stops.

On its recurrent states, a policy can be defined by the switching surfaces and hedging point. To see the role of each more clearly, eliminate the idleness option and augment $\mathcal{B}_k$ to include the (previously idle) states in which class k is served, defining new switching surfaces and an unbounded switching curve ($N = \infty$). These surfaces define the preference among classes. The hedging point (on the switching curve) defines the idleness region. The combination of switching surfaces (just a switching curve for two classes) and hedging point fully specifies a policy.

Optimal policies for this problem can only be found numerically, and only when the number of classes is small. Hence, we are led to consider approximate policies. Several classes of policies have been considered in other papers. One possible set of switching surfaces is the non-Markov policy considered by Zheng and Zipkin (1990) that services demands FCFS given a hedging point. Another policy proposed for the backorder problem in Wein (1991) is the $b\mu$ and $h\mu$ rule: if there are backorders, serve the class with largest $b_k\mu_k$ of the backordered classes; otherwise, serve the class with smallest $h_k\mu_k$ for which $X_k(t) < x_k^*$. A hedging point approximation is obtained as the solution to a Brownian control problem. Zheng and Zipkin (1990) also consider a longest queue (LQ) policy that serves the class with the most demands in the demand queue. They consider two identical products, so that the LQ policy (which is symmetrical for $x_1^* = x_2^*$) is optimal, and find that LQ is only marginally better than FCFS. For two non-identical products, we call a switching curve that lies along $x_1 = x_2$ for $x_k \leq \min\{x_1^*, x_2^*\}$, then extends vertically or horizontally to the hedging point, an LQ switching curve. An *offset* LQ switching curve lies along $x_1 = x_2 + (x_1^* - x_2^*)$.

The primary goal of this paper is to develop more sophisticated, yet easily computable, switching curves/surfaces and additional hedging point approximations that perform better than the policies described above.

4.2 Index Policies

The class of index policies is attractive because its computational complexity only grows linearly in the number of classes. An index policy is defined by assigning an index $\nu_k(x_k)$ to class k and serving the class with smallest index. If the indices are nondecreasing, the policy has the monotonicity property of Section 1 and has a switching curve. A hedging point must also be defined. We will call a policy *pure index* if the idleness region is $\mathcal{I} = \{x : \nu_k(x_k) \geq 0 \text{ for all } k\}$; the hedging point is the

Table 4.1: Proposed Indices and Hedging Points

	Backorders	Lost Sales
Index		
Value function approx. ($\mu\Delta V$)	√	√
Service time look-ahead (STLA)	√	√
Restless bandit	—	√
Hedging Point		
Server allocation	hedging point	hedging point
Aggregate product	workload	workload
Longest queue (LQ) hedging point	hedging point	—
Brownian motion	workload	workload

first point on the switching curve that is in $\mathcal{I}$. Otherwise, the hedging point will be specified in addition to the index.

Observing that the minimum in (1.2) is achieved by a class k for which $\mu_k \Delta_k V(x)$ is minimal, where $\Delta_k V(x) = V(x+e_k) - V(x)$, the goal of any index policy is to rank-order $\{\mu_k \Delta_k V(x)\}$. Several indices, listed in Table 4.1, are developed and tested in this paper. The most obvious approach is to approximate the optimal value function $V(x)$. In Section 5.1, $\Delta_k V(x)$ is approximated using the value function for separate single-product problems where service capacity is allocated across product classes. A hedging point must be specified before this index can be computed. The service time look-ahead (STLA) index of Section 5.2, proposed by Zipkin (1990), replaces $V(x)$ with the expected cost rate after one service time. It is related to the fully myopic policy of using the cost rate after one transition, which produces the $b\mu$ and $h\mu$ rule. The last index is based on a Lagrangian approach to the multi-armed "restless" bandit problem in Whittle (1988).

Several hedging point approximations, listed in Table 4.1, are also developed and tested. Recall that one can use either a pure index policy or an index policy plus a hedging point approximation. The server allocation method (Section 5.3) allocates service capacity across classes and then solves each single-product problem to find its base stock level x_k^*. A second method is to aggregate into a single product. Instead of stock levels, a threshold is found on the total workload, or expected service time, represented by the stock. Given a switching curve, the workload threshold defines a hedging point. A third method (Section 5.4) uses the analysis of the LQ policy for the backorder case in Zipkin (1992). For each product, the variance of the demand queue under an LQ policy is approximated, a distribution fit to the variance, and a stock level computed. A workload threshold can also be found through Brownian motion analysis. The backorder problem is analyzed in Wein (1991); we treat the lost sales problem in Section 4. Finally, if the performance of a policy can be evaluated, the best hedging point for a given index policy can be found by searching along the switching curve. A good starting point is the pure index hedging point.

We have several reasons for considering the lost sales problem. It has received less attention than the backorder problem in the literature, is at least as appropriate in many applications, provides an interesting example of Brownian motion analysis, and is the only version to which the restless bandit index applies. This index is unique in possessing some form of asymptotic optimality. We focus on undiscounted costs, but discounted costs are used in Section 3.1 to compute the Gittens index.

4.3 Restless Bandit Analysis

The scheduling problem defined in Section 1 can be viewed as a generalization of the multi-armed bandit problem called a restless bandit (see Whittle 1988 and Weber and Weiss 1990). This section begins by showing how to calculate the Gittens index, neglecting the restlessness, for the discounted backorder problem. Next, the restless bandit index is derived for the lost sales problem and its properties discussed.

4.3.1 Gittens Index

Consider the following single-product optimal retirement subproblem with discounting and backorders. At each time we decide whether to continue to produce class k or retire for a cost of M. The instantaneous cost function is $c_k(x) = h_k x^+ + b_k x^-$. Also let $\Lambda_k = \lambda_k + \mu_k + \alpha$. Dropping the subscript k, the optimal cost function $V(x|M)$

satisfies the optimality equations

$$V(x|M) = \frac{1}{\Lambda}\min\{M, c(x) + \lambda V(x-1) + \mu V(x+1)\}. \tag{3.1}$$

The Gittens index is usually defined as the indifference value

$$M(x) = \max\{M:\ M = V(x|M)\}. \tag{3.2}$$

It is the retirement value at which one is indifferent to retiring in state x. Katehakis and Veinott (1987) point out that $M(x)$ can be found by solving the following dynamic program, called a restart-to-x problem, for each state x: in the subproblem without retirement, one can choose at any time to transition immediately to state x. The optimal cost function for the restart-to-x problem, $v^x(\cdot)$, satisfies

$$v^x(y) = \min\{c(y)/\Lambda + qv^x(y-1) + pv^x(y+1), v^x(x)\},\ \ y \neq x,\ \text{ and} \tag{3.3a}$$

$$v^x(x) = c(x)/\Lambda + qv^x(x-1) + pv^x(x+1), \tag{3.3b}$$

where $p = \mu/\Lambda$ and $q = \lambda/\Lambda$. The Gittens index is

$$M(x) = v^x(x). \tag{3.4}$$

To solve (3.3), first consider the case $x \leq 0$. The optimal policy will restart from states $x-1$ and some $B > 0$. Dropping the superscript x, (3.3) becomes

$$v(x) = bx^-/\Lambda + qv(x) + pv(x+1), \tag{3.5a}$$

$$v(y) = c(y)/\Lambda + qv(y-1) + pv(y+1),\ \ x < y < B-1, \text{and} \tag{3.5b}$$

$$v(B-1) = h(B-1)/\Lambda + qv(B-2) + pv(x). \tag{3.5c}$$

Furthermore, B is the smallest state y in which restarting from y is preferable to restarting from $y+1$, i.e., B is minimal satisfying

$$v(x) < hB/\Lambda + qv(B-1) + pv(x). \tag{3.6}$$

Given B, (3.5) is a linear system with a unique solution $v(y)$. The following method can be used to solve (3.5) and find B. First, express $v(y)$ in terms of $v(x)$:

$$v(y) = a(y)v(x) + k(y).$$

Then

$$a(y) = (1/p)a(y-1) - (q/p)a(y-2) \text{ and} \tag{3.7a}$$

$$k(y) = -c(y-1)/\mu + (1/p)k(y-1) - (q/p)k(y-2), \tag{3.7b}$$

for $x+1 < y < B$, with initial conditions $a(x) = 1$, $a(x+1) = (1-q)/p$, $k(x) = 0$, and $k(x+1) = bx/\mu$. The following procedure is used to find B.

1. Compute $a(y)$ and $k(y)$, $y = x+2, \ldots, 1$ using (3.7). Set $y = 1$.

2. If $y+1 = B$, (3.5c) gives

$$v(x) = \frac{k(y) - qk(y-1) - hy/\Lambda}{p - a(y) + qa(y-1)}$$

and (3.6) becomes

$$[1 - p - qa(y)]v(x) < h(y+1)/\Lambda + qk(y). \tag{3.8}$$

If (3.8) holds then stop; $B = y+1$ and $M(x) = v(x)$.

3. Otherwise, increment y, compute $a(y)$ and $k(y)$ from (3.7), and go to step 2.

The same approach works for $x > 0$. The optimal policy restarts from states $x+1$ and $B < 0$, where B is maximal satisfying

$$v(x) < bB/\Lambda + qv(x) + pv(B+1). \tag{3.9}$$

The dynamic programming equations are

$$v(x) = hx/\Lambda + qv(x-1) + pv(x), \tag{3.10a}$$

$$v(y) = c(y)/\Lambda + qv(y-1) + pv(y+1), \quad x < y < B-1, \text{ and} \tag{3.10b}$$

$$v(B+1) = h(B+1)^{-}/\Lambda + qv(x) + pv(B+2). \tag{3.10c}$$

Then

$$a(y) = (1/q)a(y+1) - (p/q)a(y+2) \text{ and} \tag{3.11a}$$

$$k(y) = -c(y)+1)/\lambda + (1/q)k(y+1) - (p/q)k(y+2), \tag{3.11b}$$

for $B < y < x-1$, with initial conditions $a(x) = 1$, $a(x-1) = (1-p)/q$, $k(x) = 0$, and $k(x-1) = -hx/\lambda$. To find B:

1. Compute $a(y)$ and $k(y)$, $y = x-2, \ldots, 0$ using (3.11). Set $y = 0$.

2. If $y - 1 = B$, (3.10c) gives

$$v(x) = \frac{k(y) - pk(y+1) + by/\Lambda}{q - a(y) + pa(y+1)}$$

   and (3.9) becomes

$$[1 - q - pa(y)]v(x) < -b(y-1)/\Lambda + pk(y). \tag{3.12}$$

   If (3.12) holds then stop; $B = y - 1$ and $M(x) = v(x)$.

3. Otherwise, decrement y, compute $a(y)$ and $k(y)$ from (3.11), and go to step 2.

As discussed in Section 2, the Gittens index is not adequate for the scheduling problem because of restlessness: the arms (classes) change state while passive (not being served) due to arrivals. In fact, another Gittens index can be computed for each class when it is idle. The previous analysis is modified by setting $\mu_k = 0$. We use the subscript A for the previous "active arms" and P for these "passive arms."

For $x \leq 0$, the optimal policy restarts from $x - 1$, and

$$v(x) = bx^-/\Lambda + qv(x);$$

$$M_P(x) = v(x) = bx^-/(1-q), \quad x \leq 0, \tag{3.13}$$

where now $q = \lambda/(\lambda + \alpha)$. For $x > 0$, the optimal policy restarts from the maximal $B < 0$ for which

$$v(x) < bB^-/\Lambda + qv(x),$$

namely

$$B = -\lceil \alpha v(x)/b \rceil, \tag{3.14}$$

and

$$v(x) = \sum_{i=0}^{x-B-1} q^i c(x-i)/\Lambda + q^{x-B} v(x). \tag{3.15}$$

A search procedure can be used to solve (3.14) and (3.15), giving $M_P(x) = v(x)$.

Neither the active or passive index is monotonic. For either index, as $|x| \to \infty$ the cost of not retiring is of order $c(x)/\alpha$, the discounted cost of remaining in state x forever. Hence, the index approaches this value as well. But $c(x)$ is decreasing for $x < 0$ and increasing for $x > 0$.

4.3.2 Restless Bandit Index

Whittle (1988) defines a restless bandit problem as a resource allocation problem similar to a multi-armed bandit except that the arms not being played, called passive, continue to change state according to a Markov law that is different than the law governing their transitions when active. Passive arms can also incur costs. In the scheduling problem, there are $K + 1$ arms, one for each class plus an idleness arm whose index is zero. There must be exactly one active arm at each time, where active means that the server is assigned to that class. An index is defined by considering a single-product subproblem that chooses between active (serving) or passive (idling) given a passive tax, or cost of not serving, ν. We will consider only the undiscounted, lost sales problem, so that $\Lambda_k = \lambda_k + \mu_k$. Dropping the subscript k, the optimal cost function $V(x,\nu)$ and gain $g(\nu)$ satisfy the optimality equations

$$V(x,\nu) + g(\nu) =$$

$$\frac{1}{\Lambda}\min\{c(x)+\lambda V(x-1,\nu)+\mu V(x+1,\nu), \nu+c(x)+\lambda V(x-1,\nu)+\mu V(x,\nu)\}. \quad (3.16)$$

When $x = 0$, replace $x-1$ with x in (3.16). The optimal policy will be of threshold form, active in states $x = 0,\ldots,B-1$ and passive when $x \geq B$. Denote this policy $\pi(B)$, with the corresponding gain $g^{\pi(B)}$. Under $\pi(B)$, $B - X$ is a finite queue with utilization ρ (all notation now refers to a single class). Assuming that $\rho < 1$, its steady-state distribution is

$$P_B(x) = \Pr\{X = x|\pi(B)\} = \rho^{B-x}P_B, \quad \text{where}$$

$$P_B = P_B(B) = \frac{1-\rho}{1-\rho^{B+1}}.$$

The restless bandit index $\nu(x)$ is defined as the passive tax that achieves indifference in the min in (3.16):

$$\Delta V(x,\nu(x)) = \nu(x)/\mu, \quad (3.17)$$

where $\Delta V(x,\nu) = V(x+1,\nu) - V(x,\nu)$. Indifference in (3.16) at $x = B-1$ implies that the policies $\pi(B-1)$ and $\pi(B)$ are both optimal. Hence, the index $\nu(B-1)$ can be found by solving for ν in

$$g^{\pi(B-1)}(\nu) = g^{\pi(B)}(\nu). \quad (3.18)$$

For a given policy, the contribution of ν to the gain (per transition) is

$$g^{\pi(B)}(\nu) = g^{\pi(B)}(0) + P_B\nu/\Lambda. \quad (3.19)$$

Combining (3.18) and (3.19),

$$\nu(B-1) = \frac{\Lambda[g^{\pi(B)}(0) - g^{\pi(B-1)}(0)]}{P_{B-1} - P_B}. \tag{3.20}$$

The gain is

$$\begin{aligned}
\Lambda g^{\pi(B)}(0) &= \sum_{x=0}^{B} P_B(x)c(x) \\
&= sP_B(0) + h\sum_{x=1}^{B} xP_B(x) \\
&= s\rho^B P_B + hP_B\rho^B \sum_{x=1}^{B} x(1/\rho)^x \\
&= s\rho^B P_B + hP_B\left[\frac{B-(B+1)\rho+\rho^{B+1}}{(1-\rho)^2}\right].
\end{aligned}$$

The denominator of (3.20) can be written

$$P_{B-1} - P_B = \frac{1-\rho}{1-\rho^B} - \frac{1-\rho}{1-\rho^{B+1}} = \rho^B P_{B-1}P_B.$$

After further simplification, (3.20) becomes

$$\nu(B-1) = -\frac{s}{\rho} + \frac{h[1-(B+1)\rho^B + B\rho^{B+1}]}{(1-\rho)^2\rho^B}. \tag{3.21}$$

Surprisingly, the lost sales term in (3.21) is constant, meaning that the penalty paid for lost sales in this index scheme is the same whether lost sales are being incurred ($x = 0$) or not ($x > 0$). As a result, the lost sales term tends to dominate the index for small inventory positions (less than the hedging point), and the class with minimal $s_k\mu_k/\lambda_k$ is produced when x is small. This min $s\mu/\lambda$ rule will be seen again in Section 4. For large inventory positions (beyond the hedging point), the switching curve is approximately a straight line with slope $\mathrm{d}x_i/\mathrm{d}x_j = \ln\rho_i/\ln\rho_j$, which is a weighted version of the LQ policy.

4.3.3 Asymptotic Optimality

Unlike the Gittens index for the standard multi-armed bandit problem, the restless bandit index does not give an optimal policy. Under certain conditions, however, an asymptotic optimality holds. Let $\mathbf{X}$ be the state space for an arm and $D(\nu) \subseteq \mathbf{X}$ be the set of states in which the optimal policy for tax ν is active.

Definition 1 *An arm is indexable if $D(\nu)$ increases monotonically from $\emptyset$ to $\mathbf{X}$ as ν increases from $-\infty$ to ∞.*

Consider a problem with the constraint that exactly m of n arms must be active at any one time, and a relaxed-constraint problem where a time average of m arms must be active. If all arms are indexable, then as $m, n \to \infty$ with m/n fixed, the optimal per-project gain is asymptotically the same as that for the relaxed-constraint problem. Furthermore, Whittle (1988) conjectures and Weber and Weiss (1990) prove under an additional technical condition that the index policy of Section 3.2 is asymptotically optimal for the relaxed-constraint problem. Hence, the index policy is asymptotically optimal for the exactly-m problem.

To establish indexability for the lost sales problem, one needs to show that $\nu(x)$ is nondecreasing and well-defined for all x. These properties follow from (3.21). However, the above result is for problems with $m \to \infty$ active arms, i.e., m servers. It says that, for an m-server K-class problem (with $K + m$ arms), the index policy is approximately optimal for a large number of servers and classes. The one-server problem differs from an m-server problem with service rates divided by m only in the higher service rate that can be applied to a class. If $K\rho_k$ is bounded for all k as $K \to \infty$, i.e., each product has a small utilization, then the service rate limit should be irrelevant and it is reasonable to expect that asymptotic optimality holds for the one-server problem with a large number of classes.

In contrast, the backorder problem is not indexable; $\nu(x)$ does not exist (i.e., equals $-\infty$) for all x. The difficulty is that ν is a Lagrange multiplier for the constraint on the time-average number of active arms. For the backorder problem, any stable policy must serve a time-average of ρ classes, so relaxing this constraint does not change the optimal value and the Lagrange multiplier does not exist.

A second characteristic of the restless bandit index is that the hedging point x^* defined by the pure index policy lies on the boundary of the optimal idleness region,

$$x_k^* = \min\{x_k : x \in \mathcal{I}\},$$

so that x^* is a lower bound on the optimal hedging point. Monotonicity arguments imply that $\mathcal{I}$ touches the "asymptotes" $x_k = x_k^*$ for large x.

4.4 Brownian Motion Analysis

Wein (1992) develops a policy for the backorder problem using a diffusion approximation. The server is idle when the workload $w(t) = \sum m_k X_k(t)$ exceeds the threshold

$$w = \frac{\sum \lambda_k m_k^2 (v_{sk}^2 + v_{dk}^2)}{2(1-\rho)} \ln(1 + b/h), \tag{4.1}$$

where $b = \min\{b_k \mu_k\}$, $h = \min\{h_k \mu_k\}$, and v_{sk} and v_{dk} are the service and inter-demand time coefficients of variation for class k. In the case of exponential distributions, $v_{sk} = v_{dk} = 1$. The switching curve is determined by the $b\mu$ and $h\mu$ rule, so that all of the workload is held in the class with minimal $h_k \mu_k$ (actually, a small discretionary amount can be held in every other class).

We derive an analogous policy for the lost sales problem. The essential difference is that the heavy traffic condition, ρ slightly less than one, is replaced by the conditions $\rho \approx 1$ and the ratio of holding to lost sales cost is small, $h/s \ll 1$. The approach taken is to (1) formulate the scheduling problem in terms of cumulative processes, (2) define an approximate Brownian motion control problem, (3) reformulate to give a one-dimensional control problem called the workload formulation, (4) solve the workload formulation for an initial throughput, and (5) calculate a new throughput from this solution and iterate until the throughput converges. For consistency with much of the heavy traffic scheduling literature, inventory will be denoted $Z_k(t)$, not $X_k(t)$, in this section.

4.4.1 The Scheduling Problem

Define the independent renewal counting processes

$S_k(t)$ = number of class k service completions after serving class k for time t,

$D_k(t)$ = number of class k demands that occur during the first t units of non-stockout time, i.e., times s when $Z_k(s) > 0$,

with rates and coefficients of variation μ_k, λ_k, v_{sk}, and v_{dk}, respectively. This representation assumes that the demand process is turned off during stockouts; it is equivalent to continuous demands for exponential distributions. The scheduling problem of Section 1 can be posed as finding processes T_k that are nonanticipating with respect to Z to

$$\min \limsup_{T \to \infty} \frac{1}{T} E\left[\int_0^T \sum h_k Z_k(t) \mathrm{d}t + \sum s_k \overline{A}_k(T)\right] \tag{4.2}$$

subject to

$$Z_k(t) = S_k(T_k(t)) - D_k(A_k(t)), \tag{4.3}$$

$$A_k(t) = \int_0^t 1_{\{Z_k(s)>0\}} \mathrm{d}s, \tag{4.4}$$

$$\overline{A}_k(t) = t - A_k(t), \tag{4.5}$$

$$I(t) = t - \sum T_k(t), \quad \text{and} \tag{4.6}$$

$$T_k \ \text{ and } \ I \ \text{ nondecreasing with } \ T_k(0) = 0. \tag{4.7}$$

Here $T_k(t)$ is the cumulative time that class k is served in $(0,t]$, $I(t)$ is the idle time in $(0,t]$, and $A_k(t)$ is the time in $(0,t]$ during which class k demands can arrive.

4.4.2 The Brownian Control Problem

Let $\alpha_k = \rho_k/\rho$. Following Wein (1991) and Dai and Harrison (1991), define

$$Y_k(t) = \alpha_k t - T_k(t) \ \text{ and}$$

$$X_k(t) = S_k(T_k(t)) - \mu_k T_k(t) - [D_k(A_k(t)) - \lambda_k A_k(t)] + (\mu_k \alpha_k - \lambda_k)t. \tag{4.8}$$

By (4.3), (4.5), and (4.8),

$$Z_k(t) = X_k(t) - \mu_k Y_k(t) + \lambda_k \overline{A}_k(t).$$

Note that $\overline{A}_k$ increases only when $Z_k = 0$. Letting $L_k(t) = \lambda_k \overline{A}_k(t)$, we see that L_k is a lower regulator for Z_k, and will use the regulator form below.

We will approximate X_k with independent Brownian motions; the approximation is informally justified in Section 4.6 using a limit argument. For simplicity, the parameters of the Brownian motion are obtained here by computing the moments of $X_k(t)$ and using them to define the drift and variance of the Brownian motion. The limiting Brownian motion constructed in Section 4.6 has the same parameters. Since several different limiting Brownian motions can be constructed for a given system (see, for example, the alternative variance formula in Harrison 1988), the fact that they agree should be viewed as validating the limit argument, not the moment-fitting. Clearly, when moments can be computed, a two-moment Brownian approximation should match them. However, computing the moments here does require the assumption that demand and service are renewal processes. Any process obeying a functional central limit theorem is allowed in the limit argument of Section 4.6.

From (4.8), the drift of the Brownian motion for X_k is

$$\lim_{t \to \infty} \frac{E[X_k(t)]}{t} = \mu_k \alpha_k - \lambda_k. \tag{4.9}$$

The variance depends on the policy. For a given policy, define

$$\gamma_k = \lim_{t \to \infty} \frac{T_k(t)}{t} \text{ and}$$

$$P_k = \lim_{t \to \infty} \frac{\overline{A}_k(t)}{t}.$$

Since γ_k is the fraction of time that class k is served, $\mu_k \gamma_k$ is the class k throughput; γ_k is called the workload throughput. The lost sales rate for class k is $\lambda_k P_k$, P_k is the probability of a stockout, and

$$\gamma_k = (1 - P_k)\rho_k \leq \rho_k. \tag{4.10}$$

Assume, as argued in Harrison (1988), that we can replace $T_k(t)$ and $A_k(t)$ in (4.8) by their mean values $\gamma_k t$ and $(1 - P_k)t$. Then, from (4.8) and a central limit theorem for renewal counting processes, the variance of the Brownian motion for X_k is

$$\begin{aligned} \lim_{t\to\infty} \frac{\text{Var}[X_k(t)]}{t} &= \mu_k \gamma_k v_{sk}^2 + \lambda_k(1 - P_k)v_{dk}^2 \\ &= \mu_k \gamma_k (v_{sk}^2 + v_{dk}^2). \end{aligned} \tag{4.11}$$

Limit arguments can also be used to justify dropping the requirement that T_k be nondecreasing.

The resulting Brownian control problem (BCP) is to find processes Y_k that are nonanticipating with respect to X to

$$\min \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T \sum h_k Z_k(t)\mathrm{d}t + \sum l_k L_k(T)\right] \tag{4.12}$$

subject to

$$Z_k(t) = X_k(t) - \mu_k Y_k(t) + L_k(t), \tag{4.13}$$

$$I(t) = \sum Y_k(t), \tag{4.14}$$

$$L_k(t) = -\inf_{0\leq s\leq t}\{X_k(s) - \mu_k Y_k(s)\}, \tag{4.15}$$

$$I \text{ nondecreasing and } Y_k(0) = 0. \tag{4.16}$$

Here X_k is a Brownian motion with parameters given by (4.9) and (4.11). The stockout cost s_k has been replaced by the cost per lost sale $l_k = s_k/\lambda_k$.

4.4.3 The Workload Formulation

To achieve a state space collapse, we reformulate as in Wein (1991). Let

$$B(t) = \sum m_k X_k(t),$$

a Brownian motion with drift and variance

$$\mu = \sum m_k(\mu_k \alpha_k - \lambda_k) = 1 - \rho \text{ and} \tag{4.17}$$

$$\sigma^2 = \sum m_k \gamma_k (v_{sk}^2 + v_{dk}^2). \tag{4.18}$$

The workload formulation (WF) is to find processes Z_k, I, and L_k that are non-anticipating with respect to B to minimize (4.12) subject to

$$\sum m_k Z_k(t) = B(t) - I(t) + \mathcal{L}(t), \tag{4.19}$$

$$\mathcal{L}(t) = \sum m_k L_k(t), \tag{4.20}$$

$$Z_k(t) \geq 0, \text{ and} \tag{4.21}$$

$$I \text{ and } L_k \text{ nondecreasing.} \tag{4.22}$$

As the following theorem asserts, WF is a relaxation of BCP with the same optimal objective function value, and we can solve WF instead of BCP.

Theorem 7 *(i) Every feasible policy Y for BCP corresponds to a feasible policy (Z, I, L) for WF of equal cost. (ii) Every optimal policy (Z, I, L) for WF corresponds to a feasible policy Y for LCP of equal cost.*

A proof is given in Section 4.4 after the optimal policy is derived.

4.4.4 Solving the Workload Formulation

The workload formulation will be solved in two steps. First, an optimal Z and L is found in terms of I, then an optimal I is found. Define $W(t) = \sum m_k Z_k(t)$ and classes i and j satisfying

$$h \equiv h_i\mu_i = \min\{h_k\mu_k\} \text{ and}$$

$$l \equiv l_j\mu_j = s_j\mu_j/\lambda_j = \min\{s_k\mu_k/\lambda_k\}.$$

It is optimal to set

$$\mathcal{L}(t) = -\inf_{0\le s\le t}\{B(t) - I(t)\},$$

since this is the minimal $\mathcal{L}$ that satisfies $W(t) \ge 0$, implied by (4.21), and cost is increasing in $\mathcal{L}$. Then the optimal Z at each t is a solution to the linear program

$$\begin{aligned} \min \quad & \sum h_k Z_k(t) \\ \text{subject to} \quad & \sum m_k Z_k(t) = B(t) - I(t) + \mathcal{L}(t) \text{ and} \\ & Z_k(t) \ge 0, \end{aligned}$$

namely, the $h\mu$ rule

$$Z_k^*(t) = \begin{cases} \mu_k W(t), & k = i \\ 0, & k \ne i. \end{cases}$$

The optimal cost is $hW(t)$.

Similarly, the optimal L at each t is a solution to

$$\begin{aligned} \min \quad & \sum l_k L_k(t) \\ \text{subject to} \quad & \sum m_k L_k(t) = \mathcal{L}(t) \text{ and} \\ & L_k(t) \ge 0, \end{aligned}$$

namely, the "$l\mu$" rule

$$L_k^*(t) = \begin{cases} \mu_k \mathcal{L}(t), & k = j \\ 0, & k \ne j. \end{cases}$$

Note that L^* is nondecreasing as required by (4.22). The optimal cost is $s\mathcal{L}(t)$.

Next we solve for I. Substituting Z^* and L^* into WF gives

$$\min \limsup_{T\to\infty} \frac{1}{T} E\left[\int_0^T hW(t)\mathrm{d}t + l\mathcal{L}(T)\right] \tag{4.23}$$

$$\text{subject to} \quad W(t) = B(t) - I(t) + \mathcal{L}(t) \text{ and} \tag{4.24}$$

$$\mathcal{L}(t) = -\inf_{0\le s\le t}\{B(t) - I(t)\}. \tag{4.25}$$

A natural choice for I is to keep W in the interval $[0, c]$; standard arguments can be used to show that such a policy is optimal. Let I be the unique function satisfying

$$I(t) = \sup_{0\le s\le t} [B(s) + \mathcal{L}(s) - c]^+,$$

(4.25), and I increasing only when $W(t) = c$. Then W is a regulated Brownian motion (RBM) on $[0, c]$ with the same parameters as B. From Harrison (1985), the steady-state p.d.f. of W for the case $\mu > 0$, corresponding to $\rho < 1$, is

$$p(x) = \frac{\nu e^{\nu x}}{e^{\nu c} - 1}, \quad 0 \le x \le c,$$

where $\nu = 2|\mu|/\sigma^2$. The lower control rate is

$$\beta \equiv \lim_{t\to\infty} \frac{\mathcal{L}(t)}{t} = \frac{\mu}{e^{\nu c} - 1}. \tag{4.26}$$

If we restrict ourselves to RBM policies, then (4.23) - (4.25) reduces to finding c to minimize

$$\begin{aligned}\phi(c) &= \int_0^c hxp(x)\mathrm{d}x + l\beta \\ &= \int_0^c \frac{h\nu x e^{\nu x}}{e^{\nu c} - 1}\mathrm{d}x + \frac{l\mu}{e^{\nu c} - 1} \\ &= \frac{l\mu + hce^{\nu c}}{e^{\nu c} - 1} - \frac{h}{\nu}.\end{aligned}$$

Setting $\phi'(c) = 0$ yields

$$e^{\nu c} - \nu c - 1 - \nu\mu l/h = 0, \tag{4.27}$$

which can be solved numerically for c.

Now consider the case $\mu < 0$, corresponding to $\rho > 1$. The p.d.f. of W is

$$p(x) = \frac{\nu e^{-\nu x}}{1 - e^{-\nu c}}, \quad 0 \le x \le c,$$

and the lower control rate is

$$\beta = \frac{|\mu|}{1 - e^{-\nu c}}. \tag{4.28}$$

The cost rate is

$$\begin{aligned} \phi(c) &= \int_0^c \frac{h\nu x e^{-\nu x}}{1 - e^{-\nu c}} \mathrm{d}x + \frac{l|\mu|}{1 - e^{-\nu c}} \\ &= \frac{l|\mu| - hce^{-\nu c}}{1 - e^{-\nu c}} + \frac{h}{\nu}, \end{aligned}$$

and is minimized at

$$e^{-\nu c} + \nu c - 1 - \nu |\mu| l/h = 0, \tag{4.29}$$

Although no limit argument is provided in Section 4.6 for the case $\mu = 0$, corresponding to $\rho = 1$, we can analyze it using moment-fiting. The p.d.f. of W is

$$p(x) = 1/c, \quad 0 \le x \le c,$$

and the lower control rate is

$$\beta = \frac{\sigma^2}{2c}.$$

The cost rate is

$$\phi(c) = \frac{hc}{2} + \frac{l\sigma^2}{2c},$$

and is minimized at

$$c = \sqrt{\sigma^2 l/h}. \tag{4.30}$$

These three results are consistent. As $\rho \to 1$, $\nu \to 0$ and the small-exponent approximation $e^x \approx 1 + x + x^2/2$ can be used in (4.27) and (4.29), giving (4.30).

We end this section by proving the theorem.

Proof of theorem.

(i) Given Y feasible for BCP, let L_k satisfy (4.15), $\mathcal{L}$ satisfy (4.20), and Z_k satisfy (4.13). Then

$$\begin{aligned}\sum m_k Z_k(t) &= \sum m_k X(t) - \sum Y_k(t) + \sum m_k L_k(t) \\ &= B(t) - I(t) + \mathcal{L}(t),\end{aligned}$$

i.e., (4.19) holds. Also, (4.13), (4.15), and (4.16) imply (4.21) - (4.22), and (Z, I, L) is feasible for WF.

(ii) Given (Z, I, L) optimal for WF, let Y satisfy (4.13):

$$Y_k(t) = m_k[X_k(t) - Z_k(t) + L_k(t)].$$

Then

$$\sum Y_k(t) = B_k(t) - \sum m_k Z_k(t) + \mathcal{L}(t) = I(t),$$

i.e., (4.14) holds. Substituting Y_k into the r.h.s. of (4.15) gives

$$-\inf_{0\le s\le t}\{Z_k(s) - L_k(s)\}. \tag{4.31}$$

For $k \neq j$, $L_k(s) = 0$, (4.31) reduces to $Z_k(0) = 0$, and (4.15) holds. Now consider $k = j$. If $j \neq i$, then $Z_j(s) = 0$ and, since L_j is nondecreasing, (4.31) is just $L_j(t)$, i.e., (4.15) holds. If $j = i$, then by WF optimality, L_j increases only at times s when $\mathcal{L}(s)$ is increasing. But $\mathcal{L}$ is a lower regulator for W, so at these times $W(s) = Z_j(s) = 0$. Since $Z_j(s) \geq 0$ and L_j is nondecreasing, it follows that (4.31) is the largest value of L_j, namely, $L_j(t)$, and (4.15) holds. Optimality also ensures that $I(0) = 0$ and (4.16) holds, so Y is feasible for BCP. □

4.4.5 Updating the Throughput

The Brownian motion variance σ^2 appearing in (4.27) and (4.29) depends on the unknown throughputs γ_k. As in Dai and Harrison (1991), we overcome this difficulty

Table 4.2: Throughput Iteration for Lost Sales Case 2

Iteration	Initial γ_2	c	Final γ_2
1	.45	10.8	.4069
2	.4069	10.5	.4084
3	.4084	10.5	.4083

by iteratively computing c and γ. A reasonable initial value is $\gamma_k = \rho_k$. Given γ_k, use (4.18), (4.26) or (4.28), and (4.27) or (4.29) to compute σ^2, the lower control rate β, and c. To update γ, recall that all lost sales are attributed to class j by the $l\mu$ rule, so that $\mathcal{L}(t) = m_j L_j(t)$ and the lost sales rate for class j is

$$\lambda_j P_j = \beta / m_j.$$

From (4.10) we obtain

$$\gamma_j = \rho_j - \beta. \tag{4.32}$$

It is possible for (4.32) to give $\gamma_j < 0$, meaning that there are more lost sales than class j arrivals. A reasonable allocation of these lost sales is to set $\gamma_j = 0$, $\beta = \beta - \rho_j$, and repeatedly apply (4.32) to the class with next smallest $l_k \mu_k$.

Using the new γ, the calculations can be repeated. Convergence is reached rapidly, as demonstrated in Table 4.2.

4.4.6 Diffusion Limit Argument

Following Krichagina, Lou and Taksar (1992), we construct a sequence of systems with $\rho \to 1$, $h_k = O(1-\rho)$ and $s_k = O((1-\rho)^{-1})$. Let

$$\begin{aligned} \hat{\lambda}_k &= \lambda_k / \rho, \\ \hat{h}_k &= h_k / |1-\rho|, \quad \text{and} \\ \hat{s}_k &= |1-\rho| s_k. \end{aligned}$$

Define the set of systems $\mathcal{S}_n^+$ for $\rho < 1$ or $\mathcal{S}_n^-$ for $\rho > 1$ with parameters

$$\begin{aligned}
\mu_k^{(n)} &= \mu_k, \\
\lambda_k^{(n)} &= \hat{\lambda}_k(1 \mp 1/\sqrt{n}), \\
h_k^{(n)} &= \hat{h}_k/\sqrt{n}, \text{ and} \\
s_k^{(n)} &= \sqrt{n}\hat{s}_k.
\end{aligned}$$

Note that $n = (1-\rho^{(n)})^{-2}$ and that the original system is $\mathcal{S} = \mathcal{S}^{\pm}_{(1-\rho)^{-2}}$. As argued in Harrison (1988), assuming that T_k and A_k in (4.8) can be replaced by their mean values, it can be shown that $X_k^{(n)}(nt)/\sqrt{n}$ converges weakly to a Brownian motion $\hat{X}_k$ as $n \to \infty$. Furthermore, the $\hat{X}$ are independent with drift

$$(\mu_k^{(n)}\alpha_k^{(n)} - \lambda_k^{(n)})\sqrt{n} = \pm\lambda_k/\rho$$

and variance given by (4.11).

In the long-run average sense of (4.2), $Z_k(t)$ can be replaced by $Z_k(nt)$ while $\overline{A}_k(t)$ can be replaced by $\overline{A}_k(nt)/n$, so

$$h_k^{(n)}Z_k^{(n)}(t) = \hat{h}_k Z_k^{(n)}(nt)/\sqrt{n} \Rightarrow \hat{h}_k\hat{Z}_k(t) \text{ and}$$

$$s_k^{(n)}\overline{A}_k^{(n)}(t) = \sqrt{n}\hat{s}_k\overline{A}_k^{(n)}(nt)/n \Rightarrow \hat{s}_k\hat{A}_k(t),$$

where $\hat{Z}$ and $\hat{A}$ are limiting processes for $Z^{(n)}$ and $\overline{A}^{(n)}$. The requirement (4.7) that T_k be nondecreasing can be dropped as $n \to \infty$ because, for any sequence of policies with throughput $\mu_k\gamma_k^{(n)}$ bounded away from zero, $T_k^{(n)}(nt)/\sqrt{n}$ increases arbitrarily quickly in t for large n. Also, replace A_k with a lower regulator $\hat{L}_k$ to enforce $\hat{Z}(t) \geq 0$. Then as $n \to \infty$, the scheduling problem (4.2) - (4.7) converges to the limiting control problem (LCP) (4.16) - (4.20) with a "$\hat{\ }$" added to h, s, λ, Z, L, and X. See Krichagina, Lou and Taksar (1992) for a rigorous definition of what is meant by convergence here.

Analogous to (4.17) and(4.18), the parameters of the Brownian motion $\hat{B}(t) =$

$\sum m_k \hat{X}_k(t)$ are

$$\hat{\mu} = \sum m_k(\pm\lambda_k/\rho) = \pm 1 \quad \text{and}$$

$$\hat{\sigma}^2 = \sigma^2.$$

The solution c to (4.27) or (4.29) (with ^'s added) is an upper control limit for the scaled workload $\hat{W} = \sum m_k \hat{Z}_k(t)$. We wish to interpret this control in terms of the workload in the original system. Since $W^{(n)}(nt)/\sqrt{n} \Rightarrow \hat{W}(t)$, the control $\hat{W}(t) \leq c$ corresponds to $W(t) \leq w$, where $w = \sqrt{n}c = c/|1-\rho|$. Substituting $c = |1-\rho|w$, $\hat{\mu} = \mu/|1-\rho|$, $\hat{\nu} = \nu/|1-\rho|$, $\hat{s}_k = |1-\rho|s_k$, and $\hat{h}_k = h_k/|1-\rho|$ into (4.27) (with ^'s added) gives

$$e^{\nu w} - \nu w - 1 - \nu\mu l/h = 0,$$

showing that the workload threshold w found in this section using the LCP is the same as the workload threshold c found earlier using the BCP .

4.5 Other Heuristics

4.5.1 Value Function Approximation

In this section we derive an index, for backorders and lost sales, using a value function approximation. For brevity, this index is called $\mu\Delta V$. Consider a single-product subproblem for class k with the service rate $(\rho_k/\rho)\mu_k$. In this problem the server's availability in any time interval has been allocated across classes in proportion to their utilization ρ_k. One motivation for this allocation is that it changes the subproblem utilization from ρ_k, which tends to zero as the number of classes increases, to the system utilization ρ. Another motivation is that this fixed allocation can be viewed as a crude static policy. Selecting the minimal $\mu\Delta V$ under this policy is equivalent to performing one value iteration, giving an improved policy.

We solve the allocated server subproblem for the undiscounted backorder case first. Drop the subscript k, let μ be the allocated service rate, and let $V(x)$ and g be the optimal relative value function and gain for this problem, satisfying (3.16) with

ν omitted. The optimal policy will be of threshold form $\pi(B)$ for some $B > 0$: the server is busy in states $x < B$ and idle in states $x \geq B$. Under $\pi(B)$, $B - X$ is an $M/M/1$ queue. Letting $f(\cdot)$ and $F(\cdot)$ be the steady-state p.d.f. and d.f. of $B - X$, the corresponding gain is

$$\begin{aligned} \Lambda g^{\pi(B)} &= \sum_{x=-\infty}^{B} c(x)\Pr\{X = x\} \\ &= \sum_{x=0}^{B-1} h(B - x)f(x) + \sum_{x=B+1}^{\infty} b(x - B)f(x). \end{aligned} \tag{5.1}$$

Differencing,

$$\begin{aligned} \Lambda(g^{\pi(B+1)} - g^{\pi(B)}) &= \sum_{x=0}^{B} hf(x) + \sum_{x=B+1}^{\infty} bf(x) \\ &= (h + b)F(B) - b. \end{aligned} \tag{5.2}$$

The d.f. F is nondecreasing, so $g^{\pi(B)}$ is convex. It attains its minimum at the smallest B for which (5.1) is positive:

$$B = \min\{x : F(x) > b/(h + b)\}. \tag{5.3}$$

Using the $M/M/1$ distribution $F(x) = 1 - \rho^{x+1}$,

$$B = \lceil \ln[h/(h + b)]/\ln \rho \rceil - 1. \tag{5.4}$$

We will also need the optimal gain. Evaluating (5.1) using $f(x) = (1 - \rho)\rho^x$ gives

$$\Lambda g = \frac{hB - h(B+1)\rho + (h+b)\rho^{B+1}}{1 - \rho}. \tag{5.5}$$

The dynamic programming equations are

$$V(x) + g = \frac{1}{\Lambda}[c(x) + \lambda V(x-1) + \mu V(x+1)], \quad x < B, \text{ and} \tag{5.6a}$$

$$V(x) + g = \frac{1}{\Lambda}[c(x) + \lambda V(x-1) + \mu V(x)], \quad x \geq B. \tag{5.6b}$$

We can recursively compute $\Delta V(x)$ from the equations

$$\Delta V(x) = [h(x+1) - \Lambda g]/\lambda, \quad x \geq B-1, \text{ and} \tag{5.7a}$$

$$\Delta V(x) = \frac{c(x+1) - \Lambda g}{\lambda} + \frac{\Delta V(x+1)}{\rho}, \quad x < B-1. \tag{5.7b}$$

The index is $\nu(x) = \mu \Delta V(x)$.

The same approach is used for the lost sales problem. The recursion obtained is

$$\Delta V(0) = (\Lambda g - s)/\mu \text{ and} \tag{5.8a}$$

$$\Delta V(x) = \frac{\Lambda g - hx}{\mu} + \rho \Delta V(x+1), \quad 0 < x < B, \tag{5.8b}$$

$$\text{where} \quad g = \left\{ s\rho^B P_B + hP_B \left[\frac{B - (B+1)\rho + \rho^{B+1}}{(1-\rho)^2} \right] \right\} / \Lambda \text{ and} \tag{5.9}$$

$$P_B = \frac{1-\rho}{1-\rho^{B+1}}. \tag{5.10}$$

Solving the recursion gives

$$\mu \Delta V(x) = -\frac{s\rho^x (1-\rho^{B-x})}{1-\rho^{B+1}} + \frac{h[B - x - (B+1)\rho^{x+1} + (x+1)\rho^{B+1}]}{(1-\rho^{B+1})(1-\rho)}. \tag{5.11}$$

4.5.2 Service Time Look-Ahead

Static priority policies such as the $h\mu$ rule are fully myopic in the sense that they minimize the cost rate $c(\cdot)$ of the next state. Zipkin (1991) proposes a service time look-ahead (STLA) policy that considers the expected cost rate after one service time. It can be viewed as a myopic policy with no preemption. No preemption forces some degree of looking ahead, and so is preferable to only considering the cost rate of the next state. This policy is reminiscent of the transportation time look-ahead policy of Miller (1974) for the decision of which base to send an item repaired at a central depot, where transportation time takes on the role of service time.

For a given product, let

$$g(x) = E[c(x - D(S))],$$

where S is the service time and $D(t)$ is the number of demands in the interval $(0, t]$. Then $g(x)$ is the expected cost rate after one service time if the server is idle. If the server is busy, the cost is $g(x+1)$; hence, the rate at which serving this class increases the expected cost rate, assuming no preemption, is $\nu(x) = \mu\Delta g(x)$. Zipkin's policy is to choose the class with smallest $\nu(x)$, so it is an index policy. We also consider the pure index policy of idling when $\nu(x) > 0$ for all classes.

Zipkin evaluates the index for the backorder problem using the fact that $D(S)+1$ has a geometric distribution with parameter $p = \mu/(\lambda+\mu)$. Letting $q = 1-p$,

$$\Pr\{D(S) = j\} = q^j p, \quad j \geq 0$$

and for $x \geq 0$

$$\begin{aligned} g(x) &= h\sum_{j=1}^{x} j q^{x-j} p + b\sum_{x=1}^{\infty} j q^{x+j} p \\ &= h(xp - q + q^{x+1})/p + bq^{x+1}/p. \end{aligned}$$

The index is

$$\mu\Delta g(x) = -b\mu q^{x+1} + h\mu(1 - q^{x+1}), \ \ x \geq 0. \tag{5.12}$$

For $x < 0$,

$$\begin{aligned} g(x) &= E[b(-x + D(S))] \\ &= b(-x + q/p), \end{aligned}$$

and the index is

$$\mu\Delta g(x) = -b\mu, \ \ x < 0. \tag{5.13}$$

Note that, as $x \to \infty$, (5.12) approaches $h\mu$, and for large inventories the STLA policy gives the $h\mu$ rule.

Turning to the lost sales problem,

$$
\begin{aligned}
g(x) &= s\sum_{j=x}^{\infty} q^j p + h\sum_{j=1}^{x} j q^{x-j} p \\
&= sq^x + h\left(\frac{xp - q + q^{x+1}}{p}\right).
\end{aligned}
$$

The index is

$$\mu\Delta g(x) = -s\mu p q^x + h\mu(1 - q^{x+1}). \tag{5.14}$$

4.5.3 Single-Product Methods for Hedging Points

In this section we develop two hedging point approximations that construct and solve single-product subproblems. The first method, called allocated server, creates a subproblem for each class by allocating the server as in Section 5.1. The subproblem parameters for class k are

$$
\begin{aligned}
\mu_k^{\text{all}} &= (\rho_k/\rho)\mu_k, \\
\rho_k^{\text{all}} &= \rho,
\end{aligned}
$$

and λ_k, h_k, b_k, and s_k unchanged. The second method, called aggregate product, aggregates all products into a single class. The subproblem parameters are

$$
\begin{aligned}
\lambda^{\text{agg}} &= \sum \lambda_k, \\
\mu^{\text{agg}} &= \sum \lambda_k/\rho, \\
\rho^{\text{agg}} &= \rho, \\
h^{\text{agg}} &= \sum(\rho_k/\rho)h_k, \\
b^{\text{agg}} &= \sum(\rho_k/\rho)b_k, \quad \text{and} \\
s^{\text{agg}} &= \sum(\rho_k/\rho)s_k.
\end{aligned}
$$

To interpret stock levels for this aggregate product, we use the workload concept of Section 4. A stock level of B represents an expected service time, or workload, of $w = B/\mu^{agg}$. This workload can then be combined with a switching curve to set stock levels for the original products.

Before solving these subproblems, we comment on the nature of the approximations. Decomposing the system into single-product problems with the same utilization is analogous to decomposing a multi-server queue into parallel single-server queues. Performance degrades, total queue length increases, and the optimal stock level increases to compensate. Thus, the allocated server hedging point should be larger than optimal. Aggregating the system into a single product neglects the variation in X_k given $\sum X_k$, i.e., the inability to maintain the desired allocation of inventory among classes. Because variability is neglected, this hedging point should be smaller than optimal.

We will solve a generic subproblem, dropping the superscripts and subscripts from the parameters. The backorder case has already been solved; the optimal stock level is given by (5.4). For the lost sales case, let

$$\Delta g(B) = g^{\pi(B+1)} - g^{\pi(B)},$$

where $g^{\pi(B)}$ is the gain under policy $\pi(B)$. From (5.9),

$$\Lambda \Delta g(B) = s\rho^B(\rho P_{B+1} - P_B)$$

$$+h\left\{\frac{P_{B+1}[B+1-(B+2)\rho+\rho^{B+2}] - P_B[B-(B+1)\rho+\rho^{B+1}]}{(1-\rho)^2}\right\}, \tag{5.15}$$

where P_B is given by (5.10). We wish to find the stock level B that minimizes $g^{\pi(B)}$. It can be shown that $g^{\pi(B)}$ is concave; hence,

$$B = \min\{x : \Delta g(x) \geq 0\}. \tag{5.16}$$

A search over the values $x = 0, 1, \ldots$ is used to find B.

4.5.4 Hedging Points Based on Longest Queue Policies

Zipkin (1992) analyzes the longest queue (LQ) policy for identical products in the undiscounted, backorder problem. In this section, we use his result to approximate the case of non-identical products. The corresponding policy is offset LQ. The steady-state distribution of X_k is then used to find the hedging point. This LQ hedging point involves five approximations:

1. Assume an offset LQ policy is optimal,
2. Use Zipkin's approximation of the inventory variance for identical products,
3. Adjust for non-identical products,
4. Fit a distribution to the mean and variance, and
5. Decompose the idleness decision into single-product subproblems.

Step (2) is exact for two products and (1) is exact for identical products. We also tested the accuracy of steps (3) and (4) for two products by computing the joint p.d.f. of the inventories as outlined in Section 8 of Zheng and Zipkin (1990). The j.p.d.f. method gave slightly larger hedging points but, due to compensating approximation errors, resulted in worse values of the gain g. The j.p.d.f. method is not described here or used in Section 6.

Zipkin's variance estimate is a function of the number of identical products. For products with $\lambda_j \neq \lambda_k$ or $\mu_j \neq \mu_k$, the fraction of utilization due to class k is $\alpha_k = \rho_k/\rho$. Define the effective number of products for class k as $K_k^{\text{eff}} = 1/\alpha_k$, i.e., the number of identical products needed to achieve the system utilization. Notice that for identical products $K_k^{\text{eff}} = K$. The steady-state variance of X_k is

$$\sigma_k^2 = \frac{\text{Var}(N) + E(D_k^2)}{(K_k^{\text{eff}})^2}, \tag{5.17}$$

where N is the aggregate demand queue for K_k^{eff} identical products, each with demand queue $N_k(t) = x_k^* - X_k(t)$, and $E(D_k^2)$ captures variation in N_k given N. Since N is the number of customers in an $M/M/1$ queue,

$$\mathrm{Var}(N) = \rho/(1-\rho)^2. \tag{5.18}$$

Zipkin uses the approximation

$$E(D_k^2) = (K_k^{\mathrm{eff}} - 1)\rho[1 + \rho + (1 - 2\alpha_k)\rho^2 + (1 - 2\alpha_k)^2\rho^3]. \tag{5.19}$$

Also estimate the mean of N_k as

$$E(N_k) = \frac{E(N)}{K_k^{\mathrm{eff}}} = \frac{\alpha_k \rho}{1-\rho} = \frac{\rho_k}{1-\rho}, \tag{5.20}$$

which is exact for identical products.

Given the first two moments of N_k, we will fit a distribution to them in order to calculate the upper percentiles on which the stock level depends. Since N is geometric, it is reasonable to use a distribution that has a geometric tail. For simplicity, we use the shifted geometric distribution

$$f(x) = pq^{(x-a)-1}, \quad x = a+1, a+2, \ldots \; .$$

Fitting to (5.17) and (5.20) gives

$$q = 1 - \frac{\sqrt{4\sigma_k^2 + 1} - 1}{2\sigma_k^2} \quad \text{and} \tag{5.21a}$$

$$a = E(N_k) - \frac{1}{1-q}. \tag{5.21b}$$

The optimal stock level has already been found for a geometric distribution; (5.4) gives

$$B = \lceil \ln[h/(b+h)]/\ln q \rceil - 1 + a. \tag{5.22}$$

4.6 Numerical Results

Dynamic programming value iteration was used to compute optimal policies for undiscounted problems with two and three products. The recurrent states are those below the hedging point, $x \leq x^*$. For the lost sales problem, the recurrent class is finite, $0 \leq x \leq x^*$. For the backorder problem, the state space was truncated. Larger and larger state spaces were tested until the results were insensitive to increasing the state space. State spaces up to about 30 by 30 and up to 2000 value iterations were required to achieve three digit accuracy. The lost sales problem generally ran much faster.

All compatible combinations of switching curves and hedging points listed in Table 4.1 were tested. These candidate policies were evaluated using a value-iteration scheme to avoid directly solving a large linear system. The LQ and offset LQ switching curves described in Section 1 were also tested, as were the hedging point approximations generated by pure STLA and restless bandit index policies. Finally, for STLA, restless bandit, and LQ switching curves, a one-dimensional search along the switching curve was conducted to find the best hedging point for that switching curve. This data point is used to determine how much of the suboptimality of a policy is due to the switching curve and how much is due to the hedging point. These three switching curves are combined with hedging points by converting the hedging point to a workload threshold (see Section 4), then finding the point on the switching curve that matches or exceeds this workload. The $\mu\Delta V$ and offset LQ switching curves require a K-dimensional hedging point, not just a a one-dimensional workload, to be specified. Best hedging points were not found for these switching curves.

4.6.1 Lost Sales

Most of the testing was devoted to the two-product lost sales problem. Five test cases are defined in Table 4.3. We begin by comparing hedging point approximations. Table 4.4 shows the hedging points for the test problems. Since $\mu_k = 1$ in these problems, the workload is just the sum of the hedging point coordinates, $w = \sum x_k^*$. The

Table 4.3: Lost Sales Cases ($\alpha = 0, \mu = 1$)

Case	λ			h			s		
1	.4	.5		1	1		60	80	
2	.45	.45		4	1		45	45	
3	.45	.45		1	1		90	360	
4	.3	.6		1	1		120	300	
5	.3	.4		2	1		60	40	
6	.2	.3	.4	1	1.25	1.5	40	75	120

Table 4.4: Lost Sales Hedging Point Approximations

		Pure Index									
	Opt.	Restless		STLA		Alloc.		Aggreg.		Brownian	
Cs.	x^*	x^*	sub	x^*	sub	x^*	sub	w	sub	w	sub
1	(6,7)	(4,5)	15%	(4,4)	21%	(8,10)	9%	9	12%	12.4	0%
2	(3,6)	(2,4)	7%	(2,3)	14%	(4,7)	2%	5	14%	10.5	2%
3	(7,10)	(5,6)	41%	(4,5)	54%	(10,18)	23%	15	5%	13.9	8%
4	(7,13)	(4,8)	48%	(4,6)	54%	(11,17)	15%	15	9%	17.9	1%
5	(3,5)	(3,4)	2%	(3,3)	5%	(5,5)	8%	6	5%	6.5	0%
6	(5,5,6)	(3,3,4)	29%	(2,3,4)	28%	(7,8,10)	27%	9	28%	13.9	2%

suboptimality, measured in terms of the gain per unit time g/Λ, is also shown. For convenience, all of the hedging points are combined with the STLA switching curve, except for the pure restless bandit index hedging point, which is combined with its own switching curve. However, the ranking of the suboptimalities was the same for most of the cases when other switching curves were used. Note that the allocated server hedging points shown must be shifted onto the STLA switching curves, maintaining the same workload, before evaluating the policy.

The results show the Brownian hedging point to be the clear winner, both in terms of accuracy of the workload threshold and the resulting gain. If the workload must be broken down into stock levels, the allocated server hedging point is the best candidate. As expected, the allocated server hedging point is too large and the aggregate product too small. The pure index hedges are too small; this was predicted for the restless bandit index in Section 3.3.

Table 4.5: Lost Sales Switching Curves

Case	LQ		Offset LQ		Restless		STLA		$\mu\Delta V$	
	w	sub	x^*	sub	x^*	sub	x^*	sub	x^*	sub
1	13	1%	(8,10)	11%	(6,7)	1%	(6,7)	0%	(8,10)	10%
	13	1%			(6,7)	1%	(6,7)	0%		
2	11	20%	(4,7)	13%	(5,6)	13%	(3,8)	2%	(4,7)	6%
	8	9%			(3,5)	1%	(3,6)	0%		
3	14	38%	(10,18)	35%	(7,7)	15%	(6,8)	8%	(10,18)	34%
	20	17%			(9,9)	6%	(8,10)	2%		
4	18	9%	(11,17)	20%	(6,12)	8%	(7,11)	1%	(11,17)	19%
	21	6%			(7,14)	6%	(8,12)	1%		
5	7	5%	(5,5)	16%	(3,4)	2%	(3,4)	0%	(5,5)	15%
	8	4%			(3,5)	1%	(3,5)	0%		
6	14	7%	(7,8,10)	26%	(4,4,6)	6%	(5,4,5)	2%	(7,8,10)	27%
	17	2%			(4,5,6)	4%	(6,4,5)	2%		

Next, using the best available hedging point (Brownian if only the workload is needed, allocated server if the stock levels are needed), the switching curves are compared in Table 4.5. The second row of figures within a case gives the best hedging point for the switching curve, where available. The best hedging point illustrates how much of the suboptimality is due to the switching curve.

The results suggest that the best policy is the STLA index combined with the Brownian hedging point. Its gain is within 8% of optimal for all five test cases. The restless bandit index combined with the Brownian hedging point also does well. Much of the suboptimality for these policies is due to the hedging point, even though we have used the most accurate of the approximations. It appears that finding a good hedging point is more difficult than finding a good switching curve. The potential savings from using index policies is best measured by comparing STLA to the LQ policy. Suboptimality is reduced from up to 38% to under 8%.

The offset LQ and $\mu\Delta V$ switching curves perform poorly, possibly because they must use the less accurate server allocation hedging point. The simple LQ switching curve performs slightly better, but has trouble with asymmetric products, such as case 3. Its suboptimality is due to both the switching curve and the Brownian hedging

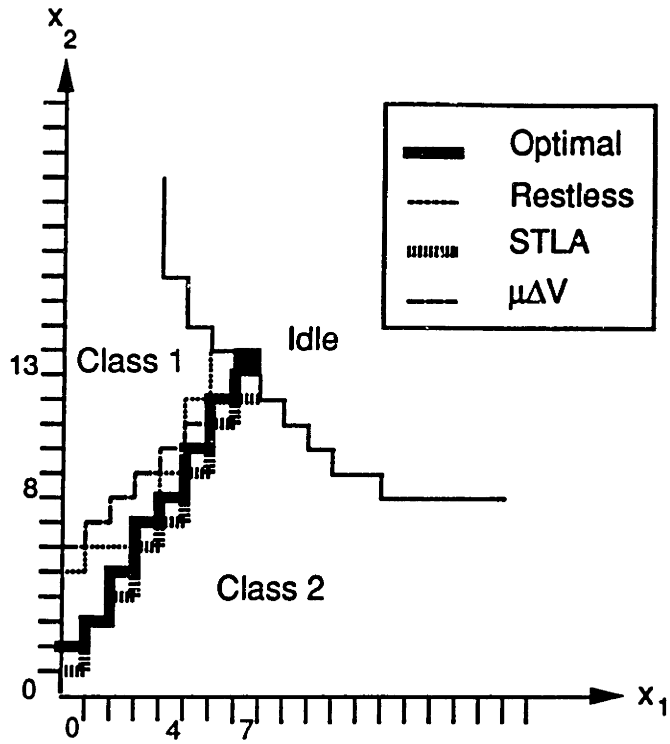

Figure 4-1: Shape of the Switching Curves— Lost Sales Case 4

point that is used with it.

The shapes of various switching curves for case 4 are shown in Fig. 4-1. The STLA curve is slightly closer to the line of symmetry $x_1 = x_2$, i.e., the LQ switching curve, than optimal. The restless bandit curve starts much farther from symmetry and initially is nearly horizontal, reflecting the dominance of the lost sales term in (3.21). The $\mu\Delta V$ curve is parallel to the symmetry line but shifted too far away. The optimal idleness region is also shown. The curvature of its boundary determines how far the optimal hedging point, (7,13), is from the the pure restless bandit index hedging point, (4,8), the latter lying on the aymptotes of the idleness region.

The optimal switching curve was found to be close to the symmetry line for fairly different products. The curve is particularly insensitive to s, with differences of a factor of two scarcely affecting the curve. Significantly different product parameters were found to have the following effects on the switching curve:

1. $s_1 \ll s_2$. The curve is offset toward class 2 ($x_1 < x_2$).

2. $h_1 \gg h_2$. The curve starts at the origin and moves toward $x_1 < x_2$ (slope steeper than one).

3. $\lambda_1 < \lambda_2$. The curve starts at $x_1 > x_2$ (below the symmetry line) and moves to $x_1 < x_2$ (slope steeper than one). If lost sales costs are constant, $s_1/\lambda_1 = s_2/\lambda_2$, the curve starts above the symmetry line.

4. Expensive product: $s_1 \ll s_2$ and $h_1/s_1 = h_2/s_2$. The curve starts at $x_1 < x_2$ (above the symmetry line) and is nearly horizontal.

We have also made a few runs with three products. Case 6 in Tables 4.4 and 4.5 is consistent with the results for two products. Recall that the restless bandit index is predicted to become more accurate for a large number of products. The three-product cases we have tested neither support nor refute this prediction.

4.6.2 Backorders

Three backorder cases are described in Table 4.6. The hedging point approximations and their suboptimality are shown in Table 4.7. Again, STLA switching curves are combined with the hedging points. The results suggest that the Brownian and LQ hedging points are the most accurate, with Brownian a little better on case 1. The aggregate product hedging point also performs fairly well. As expected, the allocated server hedging point is too large and the aggregate product too small.

To compare switching curves, the Brownian approximation is used when only a workload is needed; the LQ hedging point is used when stock levels are needed. Table 4.8 shows the results, where the second row within each case again gives the best hedging point for the switching curve. The STLA index combined with the Brownian hedging point appears to be the best candidate policy, with suboptimality of 7% or less. However, the relative accuracy of the policies varies significantly with the problem parameters. The STLA switching curve is much farther from symmetry ($x_1 = x_2$) than the optimal switching curve. The LQ switching curve (which is symmetrical) combined with the Brownian hedging point also performs well, but may do poorly on more asymmetric products. Case 3 suggests that the LQ hedging point gives an accurate workload, used with STLA in Table 4.7, but not accurate stock levels, used with $\mu\Delta V$ and offset LQ in Table 4.8.

Table 4.6: Two-Product Backorder Cases ($\alpha = 0, \mu = 1$)

Case	λ		h		b	
1	.3	.4	2	1	10	5
2	.45	.45	1	1	2	4
3	.45	.45	1.25	1	4	2

Table 4.7: Backorder Hedging Point Approximations

	Opt.	Pure STLA		Alloc		LQ		Aggreg		Brownian	
Case	x^*	x^*	sub	x^*	sub	x^*	sub	w	sub	w	sub
1	(1,3)	(1,1)	23%	(5,5)	46%	(1,2)	6%	5	0%	4.2	0%
2	(4,4)	(0,1)	55%	(10,15)	99%	(4,6)	5%	13	15%	9.9	5%
3	(3,5)	(1,0)	48%	(13,10)	78%	(6,4)	7%	12	13%	9.9	7%

Table 4.8: Backorder Switching Curves

	LQ		Offset LQ		STLA		$\mu\Delta V$	
Case	w	sub	x^*	sub	x^*	sub	x^*	sub
1	5	7%	(1,2)	8%	(1,4)	0%	(2,1)	11%
	4	3%			(1,3)	0%		
2	10	6%	(4,6)	7%	(5,5)	5%	(3,7)	12%
	9	5%			(4,5)	4%		
3	10	7%	(6,4)	10%	(2,8)	7%	(7,3)	15%
	8	5%			(2,6)	5%		

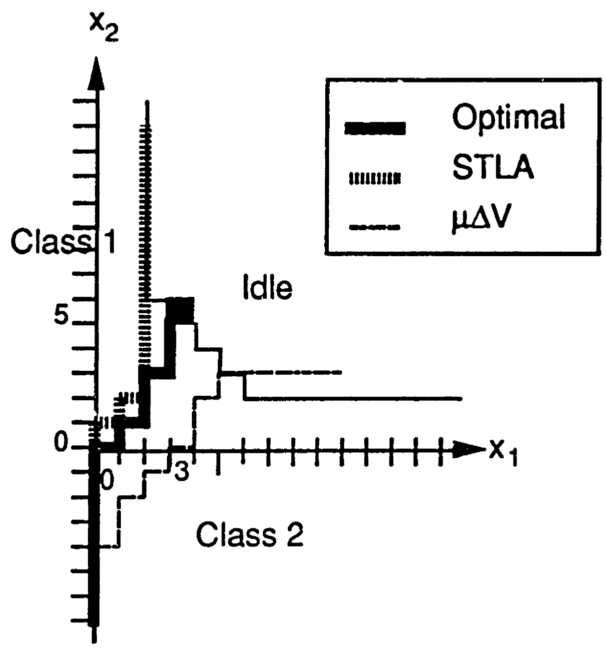

Figure 4-2: Shape of the Switching Curves— Backorder Case 3

For the backorder problem, index switching curves do not appear to significantly outperform the simpler LQ switching curve. We expect that the STLA index policy would be more robust over different parameter values, but this remains to be demonstrated. The results clearly show that choosing a good hedging point is very important.

Switching curves for case 3 are drawn in Fig. 4-2. The $\mu\Delta V$ curve is inaccurate primarily because of the LQ hedging point, (6,4), on which it is based. The STLA curve is slightly more asymmetrical than the optimal curve. In general, backorder curves are much more sensitive to asymmetric products, specifically to differences in λ and h. The 25% difference in holding costs in case 3 is responsible for moving the hedging point from symmetry to (3,5). Larger differences in h, such as in case 1, result in a hedging point near the axis. This explains the good performance of the "modified $h\mu$" rule reported in Wein (1991), where hedging points one unit away from the axis, e.g., $(x_1, 1)$, were used.

4.7 Further Research

An interesting area of further research would be to try adapting the Gittens index for standard (as opposed to restless) multi-armed bandits to this problem. The difficulty is that the Gittens index measures the value of playing an arm (serving a class) given that there are other arms of equal value, so that when its value drops we can "retire" to another arm of equal value. For restless problems, there is no constant retirement value. As a result, the Gittens index for our problem, derived in Section 3, is not monotonic. Whittle's restless bandit index assumes that the cost of the server is constant over time, and that each product can use the server whenever it is cost-effective. For our problem, these assumptions are only accurate when there are many classes and many servers (or one server and many low utilization classes). As discussed in Section 3.3, these are precisely the limiting conditions needed to apply Whittle's asymptotic optimality results.

It may be possible to improve the restless bandit index by making the cost μ of using the server a function of recent server usage by the same class. More recent usage would make the server more costly. The state space would have to be extended, and analytic results for the index seem unlikely.

Another approach is to compute a Gittens index using a variable retirement cost $M(x_k, \nu)$ that depends on the inventory x_k as well as the "base" retirement cost ν. The traditional Gittens index uses $M(x_k, \nu) = \nu$. Using $M(x_k, \nu) = V(x_k, \nu)$, the optimal value function for the single-product subproblem with server cost ν (see Section 3), gives the restless bandit index. In other words, the restless bandit index can be defined as a Gittens index with variable retirement cost. The question is how to specify $M(x_k, \nu)$ so that the Gittens index will be nondecreasing and produce accurate policies. Computing such indices numerically should be possible; one approach is found in Taylor (1968).

The restless bandit and generalized Gittens index may also be useful for attacking the related problem with set-up costs. When set-up costs are are added, the state of the dynamic program must be augmented with the class currently being produced.

The form of the optimal policy becomes much more complex, involving lot-sizing and scheduling, and good approximations have not been found. An index policy could be constructed by computing two indices for each class, measuring the value of starting and stopping production of that class.

References

Bertsekas, D.P. 1976. *Dynamic Programming and Stochastic Control.* Academic Press, New York.

Beutler, F. and D. Teneketzis 1989. Routing in Queueing Networks Under Imperfect Information: Stochastic Dominance and Thresholds, *Stoch. and Stoch. Reports* **26**, 81-100.

Bielecki, T. and P.R. Kumar 1988. Optimality of Zero-Inventory Policies for Unreliable Manufacturing Systems. *Oper. Res.* **36**, 532-541.

Buzacott, J., S. Price, and J. G. Shanthikumar 1991. Service Levels in Multistage MRP and Base Stock Controlled Production Systems, Working paper, U. of California, Berkeley.

Chen, H., P. Yang, and D. Yao 1991. Control and Scheduling in a Two-Station Queueing Network: Optimal Policies, Heuristics, and Conjectures, submitted to *Math. of Oper. Res.*

Cheng, D.W. and D.D. Yao 1991. Tandem Queues with Generalized Blocking: A Unified Model and Comparison Results. Working paper, New York University, New York.

Conway, R., W. Maxwell, J.O. McClain, and L.J. Thomas 1988. The Role of Work-in-Process Inventory in Serial Production Lines. *Oper. Res.* **36**, 229-241.

Dai, J.G. and J.M. Harrison 1991. Steady-State Analysis of RBM in a Rectangle: Numerical Methods and a Queueing Application *Annals Appl. Prob.* **1**, 16-35.

Ghoneim, H. and S. Stidham 1985. Control of Arrivals to Two Queues in Series, *Euro. J. of Oper. Res.* **21**, 399-409.

Glasserman, P. and D. Yao 1992. Monotone Optimal Control of Permutable GSMPs, Working paper, Columbia Univ., New York.

Graves, S.C. 1988. Safety Stock in Manufacturing Systems. *J. of Manufacturing and Oper. Mgmt.* **1**, 67-101.

Hajek, B. 1984. Optimal Control of Two Interacting Service Stations, *IEEE Trans. Auto. Control* **AC-29**, 491-499.

Hajek, B. 1985. Extremal Splittings of Point Processes, *Math. of Oper. Res.* **10**, 543-556.

Harrison, J.M. 1985. *Brownian Motion and Stochastic Flow Systems.* John Wiley and Sons, New York.

Harrison, J.M. 1988. Brownian Models of Queueing Networks with Heterogeneous Customer Populations, in W. Fleming and P.L. Lions (eds.), *Stochastic Differential Systems, Stochastic Control Theory and Applications*, IMA Vol. **10**, Springer-Verlag, New York, 147-186.

Hillier, F.S., R.W. Boling, and K.C. So 1986. Toward Characterizing the Optimal Allocation of Storage Space in a Production Line System with Variable Operating Times. Technical Report TR-76, Stanford University, Stanford.

Katehakis, M.N. and A.F. Veinott 1987. The Multi-Armed Bandit Problem: Decomposition and Computation. *Math. Oper. Res.* **12**, 262-268.

Klimov, G.P. 1974. Time Sharing Service Systems I. *Th. Appl. Prob.* **19**, 532-551.

Krichagina, E.V., S.X.C. Lou and M.I. Taksar 1992. Double Band Policy for Stochastic Manufacturing Systems in Heavy Traffic. Working paper, State Univ. of New York, Stony Brook.

Lee, Y.J. and P. Zipkin 1990. Tandem Queues with Planned Inventories. Working paper, Graduate School of Business, Columbia University, New York.

Lippman, S. 1975. Applying a New Device in the Optimization of Exponential Queueing Systems, *Oper. Res.* **23**, 687-710.

Lou, S.X.C. and G. van Ryzin 1989. Optimal Control Rules for Scheduling Job Shops. *Annals Oper. Res.* **17**, 233-248.

McClain, J.O. and D.R. Moodie 1991. A Comment on "Buffer Space Allocation in Automated Assembly Lines." *Oper. Res.* **39**, 857-860.

Mitra, D. and I. Mitrani 1990. Analysis of a Kanban Discipline for Cell Coordination in Production Lines, I. *Mgmt. Science* **36**, 1548-1566.

Mitra, D. and I. Mitrani 1991. Analysis of a Kanban Discipline for Cell Coordination in Production Lines, II: Stochastic Demands. *Oper. Res.* **39**, 807-823.

Muckstadt, J.A. and S.R. Tayur 1991. A Comparison of Alternative Kanban Control Mechanisms. Technical Report 962, School of Oper. Res. and Indust. Engr., Cornell University, Ithaca.

Perros, H.G. 1984. Queueing Networks with Blocking: A Bibliography. Performance Evaluation Review, *ACM Sigmetrics* **12**, 8-12.

Rosberg, Z., P. Varaiya, and J. Walrand 1982. Optimal Control of Service in Tandem Queues. *IEEE Trans. Auto. Control* **AC-27**, 600-610.

Schonberger, R.J. 1982. *Japanese Manufacturing Techniques.* Free Press, Detroit.

Smith, J.M. and S. Daskalaki 1988. Buffer Space Allocation in Automated Assembly Lines. *Oper. Res.* **36**, 343-358.

Stidham, S. 1985. Optimal Control of Admission to a Queueing System, *IEEE Trans. Auto. Control* **AC-30**, 705-713.

Stidham, S. 1988. Scheduling, Routing, and Flow Control in Stochastic Networks, in: *Stochastic Control Theory and Applications*, W. Fleming and P.L. Lions, eds., IMA Vol. 10, Springer-Verlag, New York.

Sugimori, Y., K. Kusunoki, F. Cho, S. Uchikawa 1977. Toyota Production Systems and Kanban System-Materialization of Just-in-Time and Respect-for-Human System. *Int. J. Prod. Res.* **15**, 553-564.

Taylor, H.M. 1968. Optimal Stopping of a Markov Process. *Annals Math. Stat.* **39**, 1333-1344.

Topkis, D. 1978. Minimizing a Submodular Function on a Lattice, *Oper. Res.* **26**, 305-321.

van Ryzin, G., S.X.C. Lou, and S.B. Gershwin 1991. Production Control for a Tandem Two-Machine System. Laboratory for Manufacturing and Productivity Report LMP-91-001, M.I.T., Cambridge.

Weber, R.R. and S. Stidham 1987. Optimal Control of Service Rates in Networks of Queues. *Adv. Appl. Prob.* **19**, 202-218.

Weber, R.R. and G. Weiss 1990. On an Index Policy for Restless Bandits. *J. Appl. Prob.* **27**, 637-648.

Wein, L.M. 1991. Dynamic Scheduling of a Multiclass Make-to-Stock Queue. Working paper, Sloan School of Management, MIT, Cambridge.

Whittle, P. 1988. Restless Bandits: Activity Allocation in a Changing World. In *A Celebration of Applied Probability*, ed. J. Gani, *J. Appl. Prob.* **25A**, 287-298.

Zheng, Y. and P. Zipkin 1990. A Queueing Model to Analyze the Value of Centralized Inventory Information. *Oper. Res.* **38**, 296-307.

Zipkin, P. 1990. An Alternative Dynamic Scheduling Policy. Working notes, Graduate School of Business, Columbia Univ., New York.

Zipkin, P. 1992. Performance of the Smallest-Inventory (or Longest-Queue) Policy. Working paper, Graduate School of Business, Columbia Univ., New York.