# From Sleeping to Stockpiling: Energy Conservation via Stochastic Scheduling in Wireless Networks

by

David I Shuman

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Electrical Engineering: Systems)
in The University of Michigan
2010

Doctoral Committee:

Associate Professor Mingyan Liu, Chair
Professor Demosthenis Teneketzis
Associate Professor Achilleas Anastasopoulos
Assistant Professor Owen Wu

To my parents,
for their endless and unconditional love, support, and encouragement

# *Acknowledgements*

This work was supported in part by two grants from the National Aeronautics and Space Administration (NASA) Earth Science Technology Office (ESTO), Advanced Information Systems Technologies program, NNX06AD47G and NNX09AE91G, as well as a National Science Foundation (NSF) Career Award, ANI-0238035. I am also thankful to the University of Michigan and the Horace H. Rackham School of Graduate Studies for providing, amongst other resources, annual support to attend conferences to present this research and access to world-class library services.

I am especially grateful to my advisor Professor Mingyan Liu for her support and guidance. This thesis would not exist in its current form without her vision. She has been a fantastic mentor, dedicating considerable time to training me, while at the same time giving me the freedom to explore new directions. I have particularly benefited from her knowledge of wireless communications, her emphasis on the importance of balancing theory and application, and her astute advice on life in academia. I have truly enjoyed collaborating with her, and will take with me the many lessons she has taught me.

Three straight semesters of classes and many office hours with Professor Demosthenis Teneketzis brought me up to speed in a new field and revived my interest in mathematics. His passion and aptitude for teaching make him a role model to any aspiring teacher, and I treasure his advice throughout my time in Ann Arbor. I am also grateful to Professor Owen Wu for many insightful discussions on inventory theory that contributed especially to the research presented in Sections 5.5 and 8.1.3 of this thesis, and to Professor Achilleas Anastasopoulos for serving on my thesis committee and providing helpful suggestions.

I would also like to thank my labmates, colleagues, and members of the soil moisture smart sensor web research group led by Professor Mahta Moghaddam for many interesting conversations. In particular, I would like to thank Aditya Mahajan, Ashutosh Nayyar, and Shrutivandana Sharma for extremely helpful discussions that enhanced this research and my understanding of stochastic control.

I have had the great fortune of attending exceptional academic institutions - the Marblehead, MA public school system, Phillips Academy Andover, Stanford University, the University of Oxford, and the University of Michigan. The amazing teachers I have had during 23 years at these schools are too numerous to name individually, but I am forever grateful for their inspiration and all that they have taught me.

As always, I am thankful to all of my friends and family members for their advice, loyalty, and humor. I am lucky to have close friends who challenge me, make me laugh, and support me, none more so while I worked on this thesis than Vera Sacharin. My sisters, Lisa and Lauren, have always been there to support my endeavors; my grandmother, Charlotte Egoodkin, continues to be a beacon of joy; and my great-uncle George Cowan provided support at a critical time in my intellectual development, for which I am still grateful. Finally, I wish to express my sincere gratitude to my parents, Sandra and Arnold Shuman, who instilled in me at an early age the love of learning that brought me back to academia for the Ph.D. program, and whose work ethic and generosity I can only aspire to.

# *Table of Contents*

# *List of Figures*

**Figure**

# *List of Appendices*

**Appendix**

# *Abstract*

Motivated by the need to conserve energy in wireless networks, we study three stochastic dynamic scheduling problems.

In the first problem, we consider a wireless sensor node that can turn its radio off for fixed durations of time in order to conserve energy. We formulate finite horizon expected cost and infinite horizon average expected cost problems to model the fundamental tradeoff between packet delay and energy consumption. Through analysis of the dynamic programming equations, we derive structural results on the optimal policies for both formulations. For the infinite horizon problem, we identify a threshold decision rule to determine the optimal control action when the queue is empty.

In the second problem, we consider a sensor node with an inaccurate timer in the ultra-low power sleep mode. The loss in timing accuracy in the sleep mode can result in unnecessary energy consumption from two unsynchronized devices trying to communicate. We develop a novel method for the node to calibrate its timer: occasionally waking up to measure the ambient temperature, upon which the timer speed depends. The objective is to dynamically schedule a limited number of temperature measurements in a manner most useful to improving the accuracy of the timer. We formulate optimization problems with both continuous and discrete underlying time scales, and implement a numerical solution to an equivalent reduction of the second

formulation.

In the third problem, we consider a single source transmitting data to one or more receivers over a shared wireless channel. Each receiver has a buffer to store received packets before they are drained. The transmitter's goal is to minimize total power consumption by exploiting the temporal and spatial variation of the channel, while preventing the receivers' buffers from emptying. In the case of a single receiver, we show that modified base-stock and finite generalized base-stock policies are optimal when the power-rate curves are linear and piecewise-linear convex, respectively. We also present the sequences of critical numbers that complete the characterizations of the optimal policies when additional technical conditions are satisfied. We then analyze the structure of the optimal policy for the case of two receivers.

# *Chapter 1*

## Introduction

Energy conservation is well-recognized as a key design issue in wireless networks in general, and specifically in wireless sensor networks [37]. This is primarily because such networks are often intended to operate for long periods of time without human intervention, despite relying only on battery power or energy harvesting. Thus, energy-efficient design can help to prolong the lifetime of the network, and reduce cost by avoiding the need for more expensive batteries. A second reason to conserve energy by transmitting with lower power is to limit potential interference to other users in the network.

Motivated by this issue, there have been numerous studies on methods to effectively manage energy consumption, while minimizing adverse effects on other quality of service (QoS) requirements, such as connectivity, coverage, throughput, and packet delay. Our focus here is on energy-efficiency in the design of network protocols, as opposed to hardware design. Broadly speaking, energy-efficient network design studies consider two different aims with regards to energy consumption. The first is to minimize total energy consumption, and the second is to balance energy consumption across the network. We are primarily concerned with minimizing total energy consumption.

## 1.1 Overview of Common Energy Conservation Techniques

Some of the most common energy conservation techniques, which may of course be combined in various ways, include:

1. ***Limiting the idle time of a radio***. When operating in ad hoc mode, a wireless network node consumes nearly as much energy when idle as it does when transmitting or receiving [49]. Thus, one way to conserve energy is to turn a node's radio off when it is not needed for communication.

2. ***Limiting repeated retransmissions***. Classic ARQ protocols retransmit unsuccessfully transmitted packets, even when the current channel condition is poor. In [174], Zorzi and Rao analyze a more energy-efficient error control strategy, where the main idea is to be less persistent in retransmitting data.

3. ***Adjusting transmission powers***. In general, an energy-constrained wireless network node would like to transmit and receive at as low power as possible; however, competing QoS constraints may force it to do otherwise. In the presence of such QoS constraints, a designer can conserve energy by cleverly adjusting the transmission powers over time, taking into account factors such as the statistics and current states of the channels and data. Reference [14] provides an introduction to some issues that arise in power control in wireless networks, and we discuss the topic further below.

4. ***Aggregating data***. One technique used to reduce the overall amount of communication in the network is for a node to combine its data and that of other local sensors into a compressed set of meaningful information. This technique is referred to as *data aggregation* or *data fusion*, and examples of protocols using

it include [66] and [78]. Aggregating data may consume extra energy in computation or data processing, but the idea is that there will still be a net savings from reducing the total traffic.

5. ***Adjusting routing***. Studies such as [27], [66], [136], and [147] examine how to adjust the routing paths in a multi-hop network, so as to minimize the energy consumed per packet sent, maximize the lifetime of the network, or some combination thereof. A common technique used to balance energy consumption and maximize the network's lifetime is to form local clusters of nodes that communicate to a rotating *cluster-head*, whose job is to transmit data to the common destination [66]. Energy can also be conserved by increasing cache efficiency, in order to reduce the routing overhead. For instance, [136] suggests limiting, based on the stability of a route, the amount of time the route is stored in the cache.

6. ***Sporadic sensing***. The above methods all conserve energy in the process of communicating data. However, one can also try to limit the energy consumed in sensing. Namely, sporadically sensing might consume less energy than constant monitoring, depending on the application and the sensing equipment [3]. One example of such an application that controls when sensor measurements are taken is a smart sensor web technology for measurements of soil moisture, presented in [108], [109], [110], and [141].

## 1.2 Scope and Organization of the Thesis

The problems we consider in this thesis all fall into the broader class of *resource allocation* problems, where a centralized decision maker or scheduler must allocate a scarce and/or costly resource in a judicious manner. The primary resource of concern in our problems is energy. We formulate mathematical models of wireless communication systems that feature: (i) inherent uncertainty; and (ii) variables that can be controlled over time by the decision maker in order to optimize pre-determined performance criteria. This analytical approach results in *stochastic dynamic optimization* problems, which we analyze using tools from *Markov decision theory* (see, e.g., [89]).

Of the six energy conservation techniques mentioned in the previous section, we focus in this thesis first on limiting the amount of time a radio is powered on but not actively transmitting or receiving, and second on adjusting transmission powers.

### 1.2.1 Limiting the Idle Time of a Radio

One obvious cause of radio idling is a lack of data to be communicated. Accordingly, many studies have examined the possibility of conserving energy by turning nodes on and off periodically, a technique commonly referred to as duty-cycling. Of particular note, GAF [167] makes use of geographic location information provided for example by GPS to decide which nodes to turn off; ASCENT [26] programs the nodes to self-configure based on the local traffic and routing backbone; Span [31] is a distributed duty-cycling algorithm featuring local coordinators; and PEAS [168] is another protocol to dynamically adjust sleep periods that is specifically intended for nodes with constrained computing resources that operate in harsh or hostile environments. While the salient features of these studies are quite different, the analytical approach is similar. For the most part, they discuss the qualitative features of the algorithm, and then perform numerical experiments to arrive at an energy savings percentage over some baseline system.

In Chapter 2, we also consider a wireless sensor network whose nodes sleep periodically; however, rather than evaluating the system with a given sleep control policy, we impose a cost structure and search for an optimal policy amongst a class of policies. In order to approach the problem in this manner, we need to consider a far simpler system than those used in the aforementioned studies. Thus, we consider only a single sensor node and focus on the tradeoff between energy consumption and packet delay. This sleep scheduling work also appears in [138] and [139].

A second cause of radio idling is lack of synchronization. Consider the following simple example. Two nodes in a wireless sensor network are trying to communicate. The clocks on each sensor node are inaccurate, and their *drifts* depend on environmental influences, such as temperature and supply voltage variations. The first node turns its radio on to send data, but it has woken up before the scheduled meeting time. The sender's radio idles while waiting for the second node to turn on its radio. The result of their poor synchronization is higher energy consumption, as the sender's radio is powered on for longer than necessary. A similar but more detailed example, with specific estimates of typical drifts is given in [44].

Examples of studies that present synchronization methods specifically targeted at wireless sensor networks include [43], [123], and [146]. In these and other such algorithms, nodes synchronize by using the content and timing of exchanged messages to determine the difference in their local times. For an overview of such synchronization algorithms, as well as a comparison between those synchronization techniques designed for traditional networks and those designed for wireless sensor networks, see [44], [148], and [153].

In Chapter 3, we examine a novel method to improve synchronization. Namely, we configure a wireless sensor node to use ambient temperature measurements to calibrate its own clock. This method is complementary to the existing work in synchronization, and could potentially be used in combination with existing synchronization algorithms built around message exchanging. This work also appears in [143].

### 1.2.2 Adjusting Transmission Powers

Due to random fading, wireless channel conditions vary with time and from user to user. The amount of power required to transmit a fixed amount of data is a function of the condition of the channel. Specifically, a better channel condition allows the sender to reliably transmit the same amount of data with less power. The key realization from a transmission scheduling perspective is that these channel variations are not a drawback, but rather a feature to be beneficially exploited. Namely, transmitting more data when the channel between the sender and receiver is in a “good” state, and less data when the channel is in a “bad” state increases system throughput and reduces total energy consumption. Doing so is commonly referred to as *opportunistic scheduling*.[1]

In Chapter 4, we provide an introduction to opportunistic scheduling, discuss key modeling issues, and review the existing literature. In Chapter 5, we consider the problem of energy-efficient transmission scheduling subject to strict underflow constraints. We then compare one instance of this problem to related energy-efficient transmission scheduling problems with strict deadlines in Chapter 6. The wireless communication model of Chapter 5 corresponds closely with models from inventory theory. In Chapter 7, we take a closer look at the role of stochastic prices in inventory theory. Work on this energy-efficient transmission scheduling problem also appears in [140], [144], [142], and [145].

[1]Opportunistic scheduling problems are also referred to as multi-user variable channel scheduling problems [7].

## 1.3 Contribution of the Thesis

In this section, we discuss the main contributions of this thesis. We formulate the problems discussed in the previous section as Markov decision processes (MDPs), and the main tool we use to analyze them is dynamic programming. Most of the problems we consider feature a countably or uncountably infinite state and/or action space. Therefore, unlike *finite* MDPs, they cannot in general be numerically solved exactly via dynamic programming, and suffer from the well-known *curse of dimensionality* [35, 124]. In any case, our primary aim in tackling these problems is to analyze the dynamic programming equations in order to (i) determine if there are circumstances under which we can analytically derive optimal solutions; and (ii) leverage our mathematical analysis and results on the structures of the optimal scheduling policies to improve our intuitive understanding of the problems.

The mathematical abstractions we construct are all developed with specific wireless communications applications in mind; however, the formulations are general enough that our results and techniques may find applications in other domains as well.

As discussed in Section 1.2, the problems we consider all fall into the class of stochastic dynamic resource allocation problems featuring energy as the primary resource of concern. However, the competing QoS interests are different in each of the problems. In the sleep scheduling problem, the competing QoS interest is average packet delay, which is likely to increase with the amount of time the node spends conserving energy in the sleep mode. Packet delay is also the competing QoS interest in the transmission scheduling problem; however, rather than average delay, we consider strict deadline constraints by which the packets must be transmitted. In

Chapter 3, the competing QoS interest is clock calibration accuracy.

The nature of the contribution also differs across the problems we consider. Whereas the bulk of the contribution in the sleep scheduling and transmission scheduling problems lies in the technical analysis of the models and resulting insights, the main contributions in the clock calibration problem are the novel engineering approach and problem formulation.

We now elaborate on the specific contributions of each chapter. In Chapter 2, we consider a single wireless sensor network node that can be put to sleep to conserve energy. We formulate finite horizon expected cost and infinite horizon average expected cost problems to model the fundamental tradeoff between packet delay and energy consumption. For the infinite horizon problem, we completely characterize the optimal stationary policy, and show that the optimal control at the boundary state (when the queue is empty) is determined by a threshold decision rule that is a function of the problem parameters. For the finite horizon problem, we characterize the optimal control action when the queue is non-empty and when the queue is empty towards the end of the time horizon. We also show that due to "end-of-horizon" effects, the optimal control action at the boundary state may not be monotonic in time; however, we conjecture that it has one of three simple structures, depending on the problem parameters.

In Chapter 3, we consider an ultra-low power sensor node with an inaccurate timer in the sleep mode. We develop a novel method for the node to calibrate its own clock by occasionally waking up to take ambient temperature measurements. The objective is to dynamically schedule a limited number of temperature measurements in a manner most useful to improving the accuracy of the timer. We formulate the temperature measurement scheduling as both a partially observed semi-Markov decision process (POSMDP) with a continuous underlying time scale and a partially observed Markov decision process (POMDP) with a discrete underlying time scale (when extra technical assumptions can be made), and show that both problems can be reduced to finite state, finite action, finite horizon MDPs. We then implement a numerical solution to the MDP resulting from the second formulation. The numerical solution computes the optimal scheduling policy and resulting expected energy costs for low dimensional instances of the problem.

In Chapter 5, we formulate the task of energy-efficient transmission scheduling subject to strict underflow constraints as three different Markov decision problems, with the finite horizon discounted expected cost, infinite horizon discounted expected cost, and infinite horizon average expected cost criteria, respectively.

We begin by showing that in the case of a single receiver under linear power-rate curves, the optimal policy is an easily-implementable *modified base-stock* policy. In each time slot, it is optimal for the sender to transmit so as to bring the number of packets in the receiver's buffer level after transmission as close as possible to a target level or critical number. The target level depends on the current channel condition, with a better channel condition corresponding to a higher target level. We also show in Chapter 6 that the strict underflow constraints may cause the scheduler to be less opportunistic than it otherwise would be, and transmit more packets under "medium" channel conditions in anticipation of deadline constraints in future time slots.

We generalize this result in two different directions. First, we relax the assumption that the power-rate curves under each channel condition are linear, and model them as piecewise-linear convex to better approximate more realistic convex power-rate curves. Under piecewise-linear power-rate curves, we show the optimal policy is a *finite generalized base-stock* policy, and provide an intuitive explanation of this structure in terms of multiple target levels in each time slot. In addition to the structural results on the optimal policy for the case of a single receiver under either linear or piecewise-linear convex power-rate curves, we provide an efficient method to calculate the critical numbers that complete the characterization of the optimal policy when certain technical conditions are satisfied.

The second generalization of the single receiver model under linear power-rate curves is to a single user transmitting to two receivers over a shared wireless channel. In this case, we state and prove the structure of the optimal policy, and show how the peak power constraint in each slot couples the optimal scheduling of the two receivers' packet streams.

In all three setups, we prove that the structure of the optimal policy in the finite horizon discounted expected cost problem extends to the infinite horizon discounted and average expected cost problems.

Throughout the analysis, we make a novel connection with inventory models that may prove useful in other wireless transmission scheduling problems. Because the inventory models corresponding to our wireless communication models have not been previously examined, our results also represent a contribution to the inventory theory literature. Specifically, this is the first work we are aware of to consider multi-item inventory models with joint resources constraints and random ordering costs. In Chapter 7, we discuss why this class of models merits its own analysis, as compared to the more commonly considered multi-item inventory models with joint resource constraints and *deterministic* ordering costs.

## 1.4 Notational Conventions

Before proceeding, we introduce some notation. We define $\mathbb{R}_+ := [0,\infty)$, $\mathbb{R}_{++} := (0,\infty)$, $\mathbb{Z}_+ := \{0,1,2,\ldots\}$, and $\mathbb{N} := \{1,2,\ldots\}$. The indicator function is defined as:

$$\mathbb{1}_{\{\text{statement}\}} := \begin{cases} 1, & \text{if the statement is true} \\ 0, & \text{otherwise} \end{cases} .$$

We also define the function:

$$[x]^+ := \begin{cases} x, & x \geq 0 \\ 0, & \text{otherwise} \end{cases} .$$

A single dot, as in $a \cdot b$, represents scalar multiplication. In general, we denote random variables and vectors by capital letters and their realizations by lowercase letters. We use bold font to denote column vectors, such as $\mathbf{w} = (w^1, w^2, \ldots, w^M)$. We include a transpose superscript whenever a vector is meant to be a row vector, such as $\mathbf{w}^\mathsf{T}$. The notations $\mathbf{w} \preceq \tilde{\mathbf{w}}$ and $\mathbf{w} \succeq \tilde{\mathbf{w}}$ denote component-wise inequalities; i.e., $w^m \leq$ (respectively, $\geq$) $\tilde{w}^m$, $\forall m$. Finally, we use the standard definitions of the meet and join of two vectors. Namely,

$$\begin{aligned} \mathbf{w} \wedge \tilde{\mathbf{w}} &= \left(w^1, w^2, \ldots, w^M\right) \wedge \left(\tilde{w}^1, \tilde{w}^2, \ldots, \tilde{w}^M\right) \\ &:= \left(\min\left\{w^1, \tilde{w}^1\right\}, \min\left\{w^2, \tilde{w}^2\right\}, \ldots, \min\left\{w^M, \tilde{w}^M\right\}\right) , \\ \text{and} \quad \mathbf{w} \vee \tilde{\mathbf{w}} &= \left(w^1, w^2, \ldots, w^M\right) \vee \left(\tilde{w}^1, \tilde{w}^2, \ldots, \tilde{w}^M\right) \\ &:= \left(\max\left\{w^1, \tilde{w}^1\right\}, \max\left\{w^2, \tilde{w}^2\right\}, \ldots, \max\left\{w^M, \tilde{w}^M\right\}\right) . \end{aligned}$$

In terms of notational consistency, the thesis is separated into three sections: Chapter 2, Chapter 3, and Chapters 4-7. Within these three groups, a symbol has the same meaning; however, the same symbol may have different meanings across the groups (for example, $d$ has a different meaning in Chapter 2 and Chapter 5).

# *Chapter 2*

# Optimal Sleep Scheduling for a Wireless Sensor Network Node

## 2.1 Introduction

In this chapter, we restrict our attention to a single node in a wireless sensor network. The node has the option of turning its transmitter and receiver off for fixed durations of time in order to conserve energy by limiting the idle time of the radio. Doing so results in additional packet delay. Thus, the focus of this study is on the fundamental tradeoff between delay and energy consumption. As such, we do not consider other quality of service measures such as connectivity or coverage. We attempt to identify the manner in which the optimal (to be defined in the following section) sleep schedule varies with the length of the sleep period, the statistics of arriving packets, and the charges assessed for packet delay and energy consumption.

The only other works we are aware of that take a similar approach are by Sarkar and Cruz, [128] and [129]. These studies consider a similar set of assumptions to our model, with the notable exceptions that a fixed cost is incurred for switching between the sleep and awake modes, and the duration of the sleep periods is flexible. The authors formulate an optimization problem and proceed to numerically solve the optimal duration and timing of sleep periods through a dynamic program.

Our model of the duty-cycling node falls into the general class of vacation models. Applicable to a wide range of problems from machine maintenance to polling systems, vacation models date back to the late 1950s. Many important results on vacation models in discrete time can be found in [4] and [154]. Reference [60] was the first study to analyze the steady-state distribution of the queue length and unfinished work of the Geo/D/1 queue, which is the uncontrolled analog to the controlled queue in our system. Reference [131] extends these results to the Geo/D/1 queue with priorities.

Within the class of vacation models, we are particularly interested in systems resulting from threshold policies; i.e., control policies that force the queue to empty out and then resume work after a vacation when either the queue length or the combined service time of jobs in queue (learned upon arrival of jobs to the system) reaches a critical threshold. The introduction of [47] provides a comprehensive overview of the results on different types of threshold policies. Of these models, [47] is the most relevant to our model, and we discuss it further in Section 2.4.5. The relevant discrete-time infinite horizon optimization results are covered in [12] and [133], and are discussed further in Section 2.4.1.

The rest of this chapter is organized as follows. In the next two sections, we describe the general system model and formulate the finite horizon expected cost and infinite horizon average expected cost optimization problems. In Section 2.4, we provide a brief review of some key results in average cost optimization theory for countable state spaces, and then characterize completely the optimal sleep policy for the infinite horizon problem. In Section 2.5, we partially characterize the optimal sleep policy for the finite horizon problem, and present two conjectures concerning the optimal control at the one state for which we have not yet specified the optimal action. Section 2.6 concludes the chapter.

## 2.2 Problem Description

The single node is modeled as a single-server queue that accepts packet arrivals, stores them in a buffer, and transmits them over a reliable channel. In order to conserve energy, the node goes to sleep (turns off its transceiver) from time to time. While asleep, the node is unable to transmit packets; however, packets continue to arrive at the node. This results in a queueing system with controlled vacations. We consider time evolution in discrete steps. Slot $t$ refers to the slot defined by the interval $[t, t+1)$.

In general, switching on and off is also an energy consuming process [137]. Therefore, we want to avoid changing modes too frequently. There are different ways to incorporate this goal into the model. One is to charge a switching cost whenever we turn on the node. In this study, instead of charging the node for switching, we require that when the node is put to sleep, it must remain asleep for $N$ time slots.

We model the packet arrival process, $\{A_t\}_{t=0,1,\ldots}$, as a Bernoulli process with success probability $p$. Packets arriving in one slot may not be transmitted until the following slot. Only one packet may be transmitted in each slot, and we assume transmission is successful with probability one (w.p.1).

Even while asleep (i.e., the radio is off), the node accurately learns its current queue size at each time $t$. When awake, the node decides whether to remain awake (and transmit a packet if the queue is non-empty) or go to sleep for $N$ slots. This decision is based on the current backlog information, as well as the current time slot. See Fig. 2.1 for a diagram of the system.

There are two objectives in determining a good sleep policy. One is to minimize

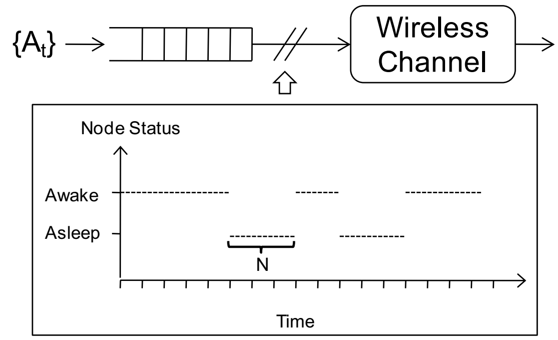


Figure 2.1. System model for the sleep scheduling problem.

the packet queueing delay and the other is to conserve energy in order to extend the node's lifetime. Accordingly, our model assesses a constant, positive cost $c$ to each backlogged packet in each slot, and a constant positive cost $d$ in each slot in which the node remains awake.

## 2.3 Problem Formulation

We consider two distinct problems. The first, Problem (**P2.1**), is the infinite horizon average expected cost problem. The second, Problem (**P2.2**), is the finite horizon expected cost problem. The two problems feature the same information state, action space, system dynamics, and cost structure, but different optimization criteria.

For both problems, the system dynamics are given by:

$$\mathbf{X}_{t+1} = \begin{cases} \begin{bmatrix} B_t + A_t \\ S_t - 1 \end{bmatrix}, & \text{if } S_t > 0 \\ \begin{bmatrix} B_t + A_t \\ N - 1 \end{bmatrix}, & \text{if } S_t = 0 \ \&\ U_t = 0 \\ \begin{bmatrix} [B_t - 1]^+ + A_t \\ 0 \end{bmatrix}, & \text{if } S_t = 0 \ \&\ U_t = 1 \end{cases} \tag{2.1}$$

Here, the information state, $\mathbf{X}_t$, is a two-dimensional vector that tracks both the current queue length, $B_t$, and the current sleep status, $S_t$, which denotes the number of slots remaining until the node wakes up ($S_t = 0$ means the node is awake). It belongs to the state space $\mathcal{X} := \mathbb{Z}_+ \times \{0, 1, \ldots, N-1\}$. Given the current state, $\mathbf{X}_t$, the probability of transition to the next state, $\mathbf{X}_{t+1}$, depends only on the random arrival, $A_t$, and the sleep decision, $U_t$. Thus, model (2.1) is a completely observed controlled Markov chain with a time-invariant matrix of transition probabilities, $\mathbf{P}(u)$.

Note that when the node is asleep ($S_t > 0$), the only available action is to continue to sleep ($U_t = 0$); however, when the node is awake ($S_t = 0$), both control actions are available, with $U_t = 0$ representing going to sleep for the next $N$ slots, and $U_t = 1$ representing staying awake. The feasible action space at state $\mathbf{x}$ is denoted by $\mathcal{U}(\mathbf{x})$.

At the beginning of the time horizon of $T$ slots, the node is awake and the initial queue length is given by the random variable $B_0$. We assume $B_0 \leq \bar{b}$, some arbitrary but fixed upper bound, with probability 1. We also assume the node's buffer size is infinite.

Finally, we present the optimization criterion for each problem. For Problem (**P2.1**), we wish to find a sleep control policy $\boldsymbol{\pi}$ that minimizes $J^{\boldsymbol{\pi}}$, defined as:

$$J^{\boldsymbol{\pi}} := \limsup_{T \to \infty} \frac{1}{T} \cdot \mathbb{E}^{\boldsymbol{\pi}} \left\{ \sum_{t=0}^{T-1} \{c \cdot B_t + d \cdot U_t\} \mid \mathcal{F}_0 \right\} . \tag{2.2}$$

In Problem (**P2.2**), the cost function for minimization is $J_0^{\boldsymbol{\pi}}$, where the expected cost-to-go at time $t$, $J_t^{\boldsymbol{\pi}}$, is defined as:

$$J_t^{\boldsymbol{\pi}} := \mathbb{E}^{\boldsymbol{\pi}} \left\{ \sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathcal{F}_t \right\} . \tag{2.3}$$

In both cases, we allow the sleep policy $\boldsymbol{\pi}$ to be chosen from the set of all randomized and deterministic control laws, $\boldsymbol{\Pi}$, such that $U_t = \pi_t\left(\mathbf{X}^t, U^{t-1}\right)$, $\forall t$, where $\mathbf{X}^t := (\mathbf{X}_0, \mathbf{X}_1, \ldots, \mathbf{X}_t)$ and $U^{t-1} := (U_0, U_1, \ldots, U_{t-1})$. $\mathcal{F}_t$ denotes all information available at time $t$.

In the next two sections, we study the infinite horizon (**P2.1**) and finite horizon (**P2.2**) problems, respectively.

## 2.4 Analysis of the Infinite Horizon Average Expected Cost Problem

In this section, we characterize the optimal sleep control policy $\boldsymbol{\pi}^*$ that minimizes (2.2). We begin by showing the existence of an optimal stationary Markov policy. We then show that the optimal policy is a threshold policy of the form: stay awake if and only if $S = 0$ and $B \geq \kappa^*$, where $\kappa^* = 0$ (never sleep) or $\kappa^* = 1$ (sleep only when the system empties out), depending on the parameters $N$, $p$, $c$, and $d$. As a matter of notation, we refer to the threshold policy with $\kappa^* = 0$, often called the "0-policy," as $\boldsymbol{\pi}_0$, and the threshold policy with $\kappa^* = 1$, often called the "1-policy," as $\boldsymbol{\pi}_1$ [47].

### 2.4.1 Conditions Guaranteeing the Existence of an Optimal Stationary Markov Policy

Due to the assumption of an infinite buffer size, the controlled Markov chain in Problem (**P2.1**) has a countably infinite state space. Recall that for such systems, an average cost optimal stationary policy is not guaranteed to exist. See [133, pp. 128–132] for such counterexamples. However, [133] also presents sufficient conditions for the existence of an average cost optimal stationary policy. We recall these conditions below and then show that the (BOR) set of assumptions is satisfied by Problem (**P2.1**) in the next subsection.

**Theorem 2.1** (Sennott)**.** *Assume that the following set (BOR) of assumptions holds (notations are explained following the theorem):*

**(BOR1)** *There exists a $z$-standard policy $\boldsymbol{g}$ with positive recurrent class $R^{\boldsymbol{g}}$.*

**(BOR2)** *There exists $\epsilon > 0$ such that $G = \{i | \bar{C}(i,u) \leq J^{\boldsymbol{g}} + \epsilon$ for some $u\}$ is a finite set.*

**(BOR3)** *Given $i \in \{G - R^{\boldsymbol{g}}\}$, there exists a policy $\boldsymbol{\theta}_i \in \Re^*(z,i)$.*

*Then there exists a finite constant $J$ and a finite function $h$, bounded below in $i$ such that:*

$$J + h(i) = \min_{u \in \mathcal{U}(i)} \left\{ \bar{C}(i,u) + \sum_j \boldsymbol{P}_{ij}(u) \cdot h(j) \right\} \ , \ \forall i \in \mathcal{X}.$$

*Moreover, a stationary policy $\boldsymbol{e}$ satisfying:*

$$\begin{aligned} \bar{C}(i,e(i)) + \sum_j \boldsymbol{P}_{ij}(e(i)) \cdot h(j) &= \min_{u \in \mathcal{U}(i)} \left\{ \bar{C}(i,u) + \sum_j \boldsymbol{P}_{ij}(u) \cdot h(j) \right\} \\ &= J + h(i), \quad \forall i \in \mathcal{X} \end{aligned}$$

*is average cost optimal.*

*Remarks on Theorem 2.1*: A Markov chain is said to be *z-standard* if there exists a distinguished state $z$ such that the expected first passage time and expected first passage cost from state $i$ to state $z$ are finite for all $i \in \mathcal{X}$. A (randomized or stationary) policy **g** is said to be a *z-standard policy* if it induces a z-standard Markov chain. $\bar{C}(i,u)$ is the one slot cost incurred at state $i$ under control action $u$. $\mathbf{P}_{i,j}(u) = \Pr(X_{t+1} = j \mid X_t = i, U_t = u)$. $J^{\mathbf{g}}$ is the average cost per unit time under policy **g**. $\Re^*(z,i)$, where $z$ refers to the distinguished state mentioned above, is the class of policies $\boldsymbol{\theta}$ such that:

(i) $\Pr^{\boldsymbol{\theta}}(X_t = i$, for some $t \geq 1 | X_0 = z) = 1$.

(ii) The expected time of first passage from $z$ to $i$ is finite.

(iii) The expected cost of first passage from $z$ to $i$ is finite.

The constant $J$ represents the minimum average cost per unit time. Note that under the (BOR) assumptions, the minimum average cost is constant and therefore independent of the initial state. This is not true in general, even when an optimal policy exists. References [12] and [121] interpret the function $h$ as a rough measure of how much we would pay to stop the process, but continue to incur a cost of $J$ per slot thereafter. In this manner, $h$ can be viewed as a cost potential function.

### 2.4.2 Existence of an Optimal Stationary Markov Policy in Problem (P2.1)

We now show that the hypotheses of Theorem 2.1 are met by our model.

**Lemma 2.2.** *Problem (**P2.1**) satisfies the (BOR) assumptions of Theorem 2.1, and therefore, there exists an optimal stationary policy $\boldsymbol{\pi}^*$ that minimizes (2.2).*

*Proof.* Let the distinguished state $z$ be $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ (the node is awake and the queue is empty). Consider the policy $\boldsymbol{\pi}_0$ of never sleeping. Given a fixed but arbitrary initial state $\left[\begin{smallmatrix}b_0\\0\end{smallmatrix}\right]$, the policy $\boldsymbol{\pi}_0$ induces a finite state Markov chain with a single positive recurrent class. In particular, the finite set of transient states is

$$T^{\boldsymbol{\pi}_0} = \left\{ \begin{bmatrix} b_0 \\ 0 \end{bmatrix}, \begin{bmatrix} b_0 - 1 \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 2 \\ 0 \end{bmatrix} \right\},$$

the set of recurrent states is

$$R^{\boldsymbol{\pi}_0} = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\},$$

and the transition diagram is shown in Figure 2.2.

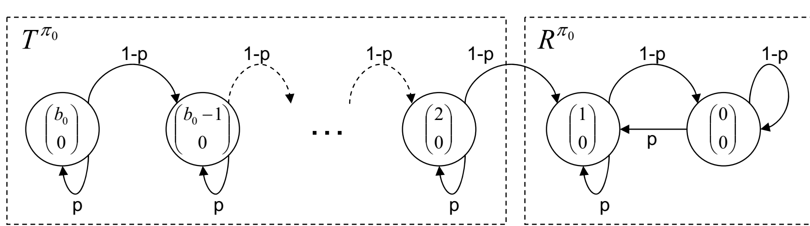


Figure 2.2. Transition diagram induced by policy $\boldsymbol{\pi}_0$.

For finite state Markov chains with a single positve recurrent class, the following three basic facts are true (see for example [24], [73]):

**(i)** The process enters the positive recurrent class (exits the transient states) in finite time with probability 1, and subsequently reaches each state in the recurrent class in finite time with probability 1.

**(ii)** There exists a unique stationary distribution, $\boldsymbol{\rho}^{\mathbf{g}}$, with

$$\boldsymbol{\rho}^{\mathbf{g}^{\mathrm{T}}} = \boldsymbol{\rho}^{\mathbf{g}^{\mathrm{T}}} \mathbf{P}^{\mathbf{g}} \text{ and } \sum_{\mathbf{x} \in \mathcal{X}} \boldsymbol{\rho}^{\mathbf{g}}(\mathbf{x}) = 1.$$

**(iii)** The long run average cost $J^{\mathbf{g}}$ is equal to

$$\sum_{\mathbf{x} \in \mathcal{X}} \boldsymbol{\rho}^{\mathbf{g}}(\mathbf{x}) \cdot \bar{C}\big(\mathbf{x}, g(\mathbf{x})\big) \ ,$$

where $\bar{C}\big(\mathbf{x}, g(\mathbf{x})\big)$ is the instantaneous cost incurred at state $\mathbf{x}$ and control action $g(\mathbf{x})$.

Thus, the first passage time from any state in the Markov chain induced by policy $\boldsymbol{\pi}_0$ to state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right] \in R^{\boldsymbol{\pi}_0}$ is finite w.p.1 by (i) above. A finite sum of bounded one slot costs is finite, and it therefore follows that the expected first passage cost from any state to $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is also finite under $\boldsymbol{\pi}_0$. We conclude $\boldsymbol{\pi}_0$ is a $z$-standard policy with positive recurrent class $R^{\boldsymbol{\pi}_0}$, and (BOR1) is satisfied.

Next, we calculate the average cost per unit time under $\boldsymbol{\pi}_0$ and examine the set $G^{\boldsymbol{\pi}_0}$ defined in (BOR2). The unique stationary distribution under this policy is given by:

$$\boldsymbol{\rho}^{\boldsymbol{\pi}_0}(\mathbf{x}) = \begin{cases} 1-p, & \mathbf{x} = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right] \\ p, & \mathbf{x} = \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right] \\ 0\ , & \text{otherwise} \end{cases} . \tag{2.4}$$

Under the 0-policy, we have:

$$\bar{C}\big(\mathbf{x}, \pi_0(\mathbf{x})\big) = d + c \cdot b \ . \tag{2.5}$$

Combining (2.4), (2.5), and property (iii) above, we compute the average cost per unit time as:

$$\begin{aligned} J^{\boldsymbol{\pi}_0} &= \sum_{\mathbf{x}=(b,s)\in\mathcal{X}} \boldsymbol{\rho}^{\boldsymbol{\pi}_0}(\mathbf{x}) \cdot \bar{C}\big(\mathbf{x}, \pi_0(\mathbf{x})\big) \\ &= (1-p)\cdot d + p \cdot (d+c) \\ &= d + pc \ . \end{aligned} \tag{2.6}$$

Taking $\epsilon = \frac{1}{2}$, we have:

$$\begin{aligned} G^{\boldsymbol{\pi}_0} &= \big\{\mathbf{x} \mid \bar{C}(\mathbf{x},u) \leq J^{\boldsymbol{\pi}_0} + \epsilon \text{ for some } u\big\} \\ &= \left\{\mathbf{x} \in \mathcal{X} \mid c \cdot b \leq d + pc + \frac{1}{2}\right\} \\ &= \left\{\mathbf{x} \in \mathcal{X} \mid b \leq \frac{d}{c} + p + \frac{1}{2c}\right\} . \end{aligned}$$

Therefore, $G^{\boldsymbol{\pi}_0}$ is a finite set, and (BOR2) is satisfied.

Finally, let state $\mathbf{j} = (b_j, s_j) \in G^{\boldsymbol{\pi}_0}$ be arbitrary. Consider the policy $\boldsymbol{\theta}_{\mathbf{j}}$ of sleeping at state $\mathbf{x} = (b_x, s_x) \in \mathcal{X}$ if $b_x < b_j$ or if $s_x > 0$, and serving if $s_x = 0$ and $b_x \geq b_j$. Then, $\boldsymbol{\theta}_{\mathbf{j}} \in \Re^*\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right], \mathbf{j}\right)$, as the induced chain visits state $\mathbf{j}$ w.p.1, and the expected first passage time and cost from $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $\mathbf{j}$ under policy $\boldsymbol{\theta}_{\mathbf{j}}$ are both finite. So (BOR3) is also satisfied by Problem (**P2.1**), and we conclude that there exists an optimal stationary policy $\boldsymbol{\pi}^*$ that minimizes (2.2). □

### 2.4.3 Optimal Policy When Queue Is Non-Empty

We now begin to identify the optimal stationary policy at each state in the state space. Throughout the following sections, we rely heavily on *interchange arguments*, which are explained in more detail in [100] and [112].

**Lemma 2.3.** *The optimal control at state* $\boldsymbol{x} = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$ *is* $U^* = 1$*, for all* $b \geq 1$*.*

*Proof.* Let $b \geq 1$ be arbitrary. Assume the state at time $t$ is $\mathbf{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$. Consider the following three policies:

- $\hat{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and behave optimally thereafter.
- $\bar{\boldsymbol{\pi}}$ : go to sleep for $N$ slots, and behave optimally thereafter.
- $\tilde{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and then sleep; if $\bar{U}_{l_0} = 1$ (i.e. the node stays awake under $\bar{\boldsymbol{\pi}}$) at any time $l_0 \geq t + N$, then let $\tilde{U}_l = \bar{U}_l$, $\forall l > l_0$; otherwise, continue to sleep.

It is clear that $\hat{\boldsymbol{\pi}}$ is superior to $\tilde{\boldsymbol{\pi}}$ by construction, so we need to show that $\tilde{\boldsymbol{\pi}}$ is superior to $\bar{\boldsymbol{\pi}}$. If the node continues to sleep forever under $\bar{\boldsymbol{\pi}}$, the queue length grows *ad infinitum* since $p > 0$. This results in $J^{\bar{\boldsymbol{\pi}}} = \infty$, due to the linear holding cost structure. Yet, we have already shown there exists at least one policy, $\boldsymbol{\pi}_0$, with a finite average cost. Therefore, the policy of sleeping for all slots after time $t + N$ is suboptimal, and cannot occur under $\bar{\boldsymbol{\pi}}$. So eventually the node will awake under $\bar{\boldsymbol{\pi}}$.

Let $\tau$ denote the number of slots from time $t$ until the first time the node awakes under policy $\bar{\boldsymbol{\pi}}$. We now compare the evolution of the Markov chain under $\bar{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{\pi}}$. For all realizations, a single packet is transmitted $\tau$ slots later under $\bar{\boldsymbol{\pi}}$, and all other packets are transmitted at the same time under both policies. Thus, the total cost from time $t$ under $\bar{\boldsymbol{\pi}}$ is almost surely $\tau \cdot c$ greater than the total cost from time $t$ under $\tilde{\boldsymbol{\pi}}$, and we conclude $\tilde{\boldsymbol{\pi}}$ is superior to $\bar{\boldsymbol{\pi}}$. By transitivity, $\hat{\boldsymbol{\pi}}$ is superior to $\bar{\boldsymbol{\pi}}$. Therefore, it is optimal to stay awake and transmit at $\left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$, for all $b \geq 1$. $\square$

### 2.4.4 Complete Characterization of the Optimal Policy

We now present the main result of this section.

**Theorem 2.4.** *In Problem (**P2.2**), the optimal control at state* $X = \begin{bmatrix} b \\ 0 \end{bmatrix}$ *such that* $b > 0$*, is* $U^* = 1$*. At the boundary state* $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$*, the optimal control,* $U^*$*, is given by:*

$$\left(\frac{p}{1-p}\right)\left(\frac{N-1}{2}\right) \quad \overset{U^*=0}{\underset{U^*=1}{\lessgtr}} \quad \frac{d}{c}\,. \tag{2.7}$$

*Proof.* The first statement follows directly from Lemma 2.3. We showed in the proof of Lemma 2.2 that the average cost per unit time under the 0-policy (never sleep) is $d+pc$. To determine the optimal policy at the boundary state $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$, we must compare the average cost per unit time of the 0-policy with that of the 1-policy (stay awake if the queue is non-empty, and sleep otherwise). The transition diagram under $\boldsymbol{\pi}_1$ is shown in Figure 2.3, with $T^{\boldsymbol{\pi}_1}$ denoting the set of transient states, and $R^{\boldsymbol{\pi}_1}$ denoting the single positive recurrent class.

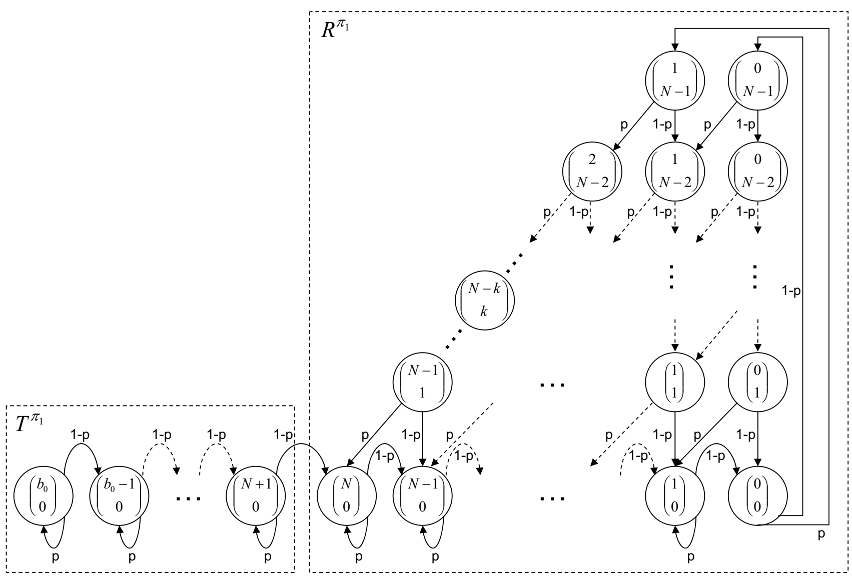


Figure 2.3. Transition diagram induced by policy $\boldsymbol{\pi}_1$.

Once again, this Markov chain has a unique stationary distribution, and it is straightforward to verify that the balance equations hold for the following stationary distribution:

$$
\boldsymbol{\rho}^{\boldsymbol{\pi}_1}(\mathbf{x}) = \begin{cases}
\frac{1-p}{N}, & \mathbf{x} = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right] \\[2ex]
\frac{1}{N} \cdot \sum_{m=0}^{N-j} \left\{ \binom{N}{m} \cdot (1-p)^m \cdot p^{N-m} \right\}, & \mathbf{x} = \left[\begin{smallmatrix} j \\ 0 \end{smallmatrix}\right], \qquad j = 1, 2, \ldots, N \\[2ex]
\frac{1-p}{N} \cdot \binom{k}{l} \cdot p^l \cdot (1-p)^{k-l}, & \mathbf{x} = \left[\begin{smallmatrix} l \\ N-k \end{smallmatrix}\right], \quad \begin{matrix} 1 \le k \le N-1 \\ 0 \le l \le k \end{matrix} \\[2ex]
0\ , & \text{otherwise}
\end{cases}
$$

From this stationary distribution, we compute the average cost per unit time:

$$
\begin{aligned}
J^{\boldsymbol{\pi}_1} &= \sum_{\mathbf{x}=(b,s)\in\mathcal{X}} \boldsymbol{\rho}^{\boldsymbol{\pi}_1}(\mathbf{x}) \cdot \Big( c \cdot b + d \cdot \pi_1(\mathbf{x}) \Big) \\
&= 0 \cdot \frac{1-p}{N} + \sum_{j=1}^{N} \left\{ (d + jc) \cdot \frac{1}{N} \cdot \sum_{m=0}^{N-j} \left\{ \binom{N}{m} \cdot (1-p)^m \cdot p^{N-m} \right\} \right\} \\
&\quad + \sum_{k=1}^{N-1} \sum_{l=0}^{k} \left\{ (lc) \cdot \frac{1-p}{N} \cdot \binom{k}{l} \cdot p^l \cdot (1-p)^{k-l} \right\} \\
&= \frac{d}{N} \cdot \sum_{j=1}^{N} \sum_{m=0}^{N-j} \left\{ \binom{N}{m} \cdot (1-p)^m \cdot p^{N-m} \right\} \\
&\quad + \frac{c}{N} \cdot \sum_{j=1}^{N} j \cdot \sum_{m=0}^{N-j} \left\{ \binom{N}{m} \cdot (1-p)^m \cdot p^{N-m} \right\} \\
&\quad + \frac{c(1-p)}{N} \cdot \sum_{k=1}^{N-1} \sum_{l=0}^{k} \left\{ l \cdot \binom{k}{l} \cdot p^l \cdot (1-p)^{k-l} \right\} \\
&= \frac{d}{N} \cdot pN + \frac{c}{N} \cdot \left( \frac{p^2 N(N-1)}{2} + pN \right) + \frac{c(1-p)}{N} \cdot \sum_{k=1}^{N-1} pk \\
&= pd + \frac{p^2 c(N-1)}{2} + pc + \frac{c(1-p)}{N} \cdot \frac{pN(N-1)}{2} \\
&= pd + pc + \frac{pc(N-1)}{2} \cdot (p + (1-p)) \\
&= pd + \frac{pc(N+1)}{2}\ . \qquad (2.8)
\end{aligned}
$$

Finally, combining (2.6) and (2.8), we compare the average costs for the two policies to determine the optimal policy at the boundary state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$:

$$J^{\boldsymbol{\pi}_1} = pd + \frac{pc(N+1)}{2} \quad \underset{U^*=1}{\overset{U^*=0}{\lesseqgtr}} \quad d + pc = J^{\boldsymbol{\pi}_0} \ . \tag{2.9}$$

Rearranging (2.9) gives (2.7). □

### 2.4.5 Related Work and Possible Extensions

The arguments presented above are quite similar to those applied to the embedded Markov chain model of [47]. In that paper, Federgruen and So consider an analogous problem in continuous time with compound Poisson arrivals. By formulating the problem as a semi-Markov decision process embedded at certain decision epochs, they show that either a no vacation policy or a threshold policy is optimal under a much weaker set of assumptions. Specifically, they allow general nondecreasing holding costs, multiple arrivals, fixed costs for switching between service and vacation modes, and general i.i.d. service and vacation times. It is quite possible that we could similarly relax our assumptions, and still retain the structural result that either a threshold policy or a no vacation policy is optimal. By imposing the extra assumptions, however, we have arrived at the more specific conclusion that if the optimal policy is an N-threshold policy, it is indeed a 1-policy; additionally, we have identified condition (2.7), distinguishing the parameter sets on which the 0-policy is optimal from those on which the 1-policy is optimal.

## 2.5 Analysis of the Finite Horizon Expected Cost Problem

In this section, we analyze the finite horizon problem, (**P2.2**), and attempt to characterize the optimal sleep control policy $\boldsymbol{\pi}^*$ that minimizes $J_0^{\boldsymbol{\pi}}$. Due to the finite time horizon and the assumption of a finite initial queue size, this problem features a finite state space (at most $\lceil \bar{b} + T \rceil \cdot N$ states). Additionally, we have a finite number of available control actions at each time slot. For such systems, we know the following from classical stochastic control theory (see, for example, [89, pp. 78–79]):

(i) There exists an optimal control policy; i.e., a policy $\boldsymbol{\pi}^*$ such that

$$J_0^{\boldsymbol{\pi}^*} = \inf_{\boldsymbol{\pi}} J_0^{\boldsymbol{\pi}},$$

where the infimum is taken over all randomized and deterministic history-dependent policies.

(ii) Furthermore, there exists an optimal deterministic Markov policy (a policy that depends only on the current state $\mathbf{X}_t$, not the past states $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \ldots$).

(iii) Define recursively the functions

$$\begin{aligned} V_T(\mathbf{x}) &:= c \cdot b \\ V_t(\mathbf{x}) &:= \min_{u \in \mathcal{U}(\mathbf{x})} \left\{ c \cdot b + u \cdot d + \mathbb{E}\left[ V_{t+1}(\mathbf{X}_{t+1}) \mid \mathbf{X}_t = \mathbf{x}, U_t = u \right] \right\}, \\ & \qquad t = 0, 1, \ldots, T-1. \end{aligned} \tag{2.10}$$

A deterministic Markov policy $\boldsymbol{\pi}$ is optimal if and only if the minimum in (2.10) is achieved by $\pi_t(\mathbf{x})$, for each state $\mathbf{x}$ at each time $t$.

We also define the "expected cost-to-go" associated with policy $\boldsymbol{\pi}$ over the time interval $[t, T]$, starting from state $\mathbf{x}$ as:

$$V_t^{\boldsymbol{\pi}}(\mathbf{x}) := \mathbb{E}^{\boldsymbol{\pi}} \left[ \sum_{k=t}^{T} \{ c \cdot B_k + d \cdot U_k \} \mid \mathbf{X}_t = \mathbf{x} \right].$$

While, in principle, we can compute the optimal policy through the dynamic program (2.10), we are more interested in deriving structural results on the optimal policy, e.g., by showing that the optimal policy satisfies certain properties or is of a certain simple form. In order to accomplish this, we use the above results throughout the section to identify the optimal control at each slot by comparing the expected cost-to-go under different deterministic Markov policies. Before proceeding, we note that for the remainder of this section, when we refer to the time $t$, we implicity assume $t \in \{0, 1, \ldots, T\}$.

### 2.5.1 Optimal Policy at the End of the Time Horizon

As with the infinite horizon problem, we identify the optimal policy in a piecewise manner, this time beginning with the slots at the end of the time horizon.

**Lemma 2.5.** *If $T - \frac{d}{c} \leq t \leq T$, the optimal policy to minimize $J_t^{\boldsymbol{\pi}}$ is $U_k^* = 0$ $\forall k \in \{t, t+1, \ldots, T\}$; i.e. sleep for the duration of the time horizon.*

*Proof.* We proceed by backward induction on $t$.

Base Case: $t = T$

This is trivial, as choosing $U_T = 1$ incurs an additional charge of $d$ without benefit.

Induction Step: We now assume it is optimal to sleep for the duration of the horizon at times $t = l+1, l+2, \ldots, T$, and show $U_l^* = 0$, where $l \geq T - \frac{d}{c}$. If the state at time $l$ is $\left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, staying awake incurs an energy cost of $d$, but provides no benefit, as it is optimal to sleep for the remainder of the horizon at time $l+1$, by the induction hypothesis. Thus, at $\mathbf{X}_l = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$, action $U_l = 0$ is optimal. If $\mathbf{X}_l = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$ for some $b > 0$, the net reward from staying awake in slot $l$ is $(T-l) \cdot c - d$, as the single packet transmitted in slot $l$ does not incur holding costs for the remainder of the horizon. However, by $l \geq T - \frac{d}{c}$, we have:

$$(T-l) \cdot c - d \leq \left[T - \left(T - \frac{d}{c}\right)\right] \cdot c - d = 0 \ .$$

Thus, $U_l = 0$ is also optimal at $\mathbf{X}_l = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$, completing the induction step and the proof of the lemma. $\square$

The simple intuition behind the above lemma is that the incremental cost of staying awake for an extra slot remains constant at $d$ throughout the time horizon; however, the benefit of doing so, as compared to sleeping for the duration of the horizon, diminishes as $t$ approaches $T$.

### 2.5.2 Optimal Policy When Queue Is Non-Empty Before the End of the Time Horizon

The following lemma characterizes the optimal sleep policy when the node is awake, the queue is non-empty, and the process is sufficiently far from the end of the time horizon.

**Lemma 2.6.** *If* $0 \leq t < T - \frac{d}{c}$ *and* $\boldsymbol{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right]$ *for some* $b > 0$*, the optimal control at slot* $t$ *to minimize* $J_t^{\boldsymbol{\pi}}$ *is* $U_t^* = 1$*; i.e., transmit a packet in slot* $[t, t+1)$*.*

*Proof.* We consider two separate cases.

Case 1: $t \geq T - N$.

Consider the following three policies:

- $\hat{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and behave optimally thereafter.
- $\bar{\boldsymbol{\pi}}$ : go to sleep (and remain asleep for duration of time horizon).
- $\tilde{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and sleep for duration of time horizon.

Define the expected rewards for following $\hat{\boldsymbol{\pi}}$ over $\bar{\boldsymbol{\pi}}$, $\tilde{\boldsymbol{\pi}}$ over $\bar{\boldsymbol{\pi}}$, and $\hat{\boldsymbol{\pi}}$ over $\tilde{\boldsymbol{\pi}}$:

$$
\begin{aligned}
R_t &:= V_t^{\bar{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right) - V_t^{\hat{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right), \\
R_t^1 &:= V_t^{\bar{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right) - V_t^{\tilde{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right), \text{ and} \\
R_t^2 &:= V_t^{\tilde{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right) - V_t^{\hat{\boldsymbol{\pi}}}\left(\begin{bmatrix} b \\ 0 \end{bmatrix}\right), \text{ respectively.}
\end{aligned}
$$

To show that $\hat{\boldsymbol{\pi}}$ is optimal in this case, it suffices to show:

$$
R_t = R_t^1 + R_t^2 \geq 0 \ . \tag{2.11}
$$

This is fairly straightforward as we have:

$$R_t^1 = c \cdot (T-t) - d \geq c \cdot \left[T - \left(T - \frac{d}{c}\right)\right] - d = 0 \ , \tag{2.12}$$

and $R_t^2 \geq 0$ by construction.

<u>Case 2</u>: $t < T - N$.

Redefine the policies $\hat{\boldsymbol{\pi}}$, $\bar{\boldsymbol{\pi}}$, and $\tilde{\boldsymbol{\pi}}$ as follows:

$\hat{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and behave optimally thereafter.

$\bar{\boldsymbol{\pi}}$ : go to sleep for $N$ slots, and behave optimally thereafter.

$\tilde{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and then sleep for $N$ slots; define:

$$l_0 := \begin{cases} \min\{l : \bar{U}_l = 1\}, \text{ if } \bar{U}_l = 1 \text{ for some } l \in \{t+N, t+N+1, T-1\} \\ T, \text{ otherwise} \end{cases}$$

$$\text{and set } \tilde{U}_l := \begin{cases} 0, \ l = t+1, t+2, \ldots, l_0 \\ \bar{U}_l, \ l = l_0+1, l_0+2, \ldots T \end{cases} .$$

Let $R_t$, $R_t^1$, and $R_t^2$ be as in Case 1, and let

$$\mathbf{a}_{t,T-1} = [a_t, a_{t+1}, \ldots, a_{T-2}, a_{T-1}] \in \{0,1\}^{T-t}$$

be a sample path of the Bernoulli arrival process $\{A_t\}$ in slots $t$ through $T-1$. We split all possible such sample paths into the following two sets:

$$\mathcal{A}_0 := \left\{\mathbf{a}_{t,T-1} \in \{0,1\}^{T-t} : \bar{U}_l = 0, \forall l \geq t\right\}, \text{ and}$$
$$\mathcal{A}_1 := \left\{\mathbf{a}_{t,T-1} \in \{0,1\}^{T-t} : \bar{U}_{l_0} = 1, \text{ for some } l_0 \in \{t+N, t+N+1, \ldots, T-1\}\right\}.$$

That is, $\mathcal{A}_0$ comprises all sample paths of the arrival process that result in the node sleeping for the duration of the horizon under policy $\bar{\boldsymbol{\pi}}$, and $\mathcal{A}_1$ comprises all sample paths of the arrival process that result in the node staying awake for at least one slot before the end of the horizon under policy $\bar{\boldsymbol{\pi}}$. Then we have:

$$R_t^1 = \Pr\left(\mathbf{A}_{t,T-1} \in \mathcal{A}_0\right) \left\{ \begin{array}{l} \mathbb{E}^{\bar{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right], \mathbf{A}_{t,T-1} \in \mathcal{A}_0\right] \\ -\mathbb{E}^{\tilde{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right], \mathbf{A}_{t,T-1} \in \mathcal{A}_0\right] \end{array} \right\}$$
$$+ \Pr\left(\mathbf{A}_{t,T-1} \in \mathcal{A}_1\right) \left\{ \begin{array}{l} \mathbb{E}^{\bar{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right], \mathbf{A}_{t,T-1} \in \mathcal{A}_1\right] \\ -\mathbb{E}^{\tilde{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \left[\begin{smallmatrix} b \\ 0 \end{smallmatrix}\right], \mathbf{A}_{t,T-1} \in \mathcal{A}_1\right] \end{array} \right\} \tag{2.13}$$

By the same argument as (2.12), we have:

$$\begin{aligned} &\mathbb{E}^{\bar{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \begin{bmatrix} b \\ 0 \end{bmatrix}, \mathbf{A}_{t,T-1} \in \mathcal{A}_0\right] \\ &- \mathbb{E}^{\tilde{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \begin{bmatrix} b \\ 0 \end{bmatrix}, \mathbf{A}_{t,T-1} \in \mathcal{A}_0\right] \\ &= c \cdot (T - t) - d \geq 0 \ . \end{aligned} \tag{2.14}$$

Furthermore, we have:

$$\begin{aligned} &\mathbb{E}^{\bar{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \begin{bmatrix} b \\ 0 \end{bmatrix}, \mathbf{A}_{t,T-1} \in \mathcal{A}_1\right] \\ &- \mathbb{E}^{\tilde{\boldsymbol{\pi}}}\left[\sum_{k=t}^{T} \{c \cdot B_k + d \cdot U_k\} \mid \mathbf{X}_t = \begin{bmatrix} b \\ 0 \end{bmatrix}, \mathbf{A}_{t,T-1} \in \mathcal{A}_1\right] \\ &\geq c \cdot N \ . \end{aligned} \tag{2.15}$$

The logic behind (2.15) is as follows. For all realizations of $\mathbf{A}_{t,T-1}$ in $\mathcal{A}_1$, the node is awake for the same number slots under policies $\bar{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{\pi}}$. Moreover, all packets are transmitted in the same slot, except the packet transmitted in slot $t$ under policy $\tilde{\boldsymbol{\pi}}$, which is transmitted in slot $l_0$ under policy $\bar{\boldsymbol{\pi}}$. Thus, the additional holding cost incurred under policy $\bar{\boldsymbol{\pi}}$ is $c \cdot (l_0 - t) \geq c \cdot N$. Substituting (2.14) and (2.15) into (2.13) yields $R_t^1 \geq 0$. Since $R_t^2 \geq 0$ by construction, we conclude (2.11) is true, and policy $\hat{\boldsymbol{\pi}}$ is optimal. $\square$

### 2.5.3 Optimal Policy When Node Is Awake and Queue Is Empty (Boundary State)

We know from Lemma 2.5 that the optimal control at $\mathbf{X}_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is to sleep when $t \geq T - \frac{d}{c}$. We now examine the optimal control at this state when $t < T - \frac{d}{c}$.

**Slots $z^* - N + 1$ through $z^*$**

**Lemma 2.7.** *If $t = z^* := \left\lfloor T - \frac{d}{c} \right\rfloor$ and $\boldsymbol{X}_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, the optimal control policy to minimize $J_t^{\boldsymbol{\pi}}$ is to sleep for the duration of the time horizon.*

*Proof.* This is trivial as, due to Lemma 2.5, the optimal policy entails sleeping for the duration of the time horizon, beginning at the following time slot, $z^* + 1$. Therefore, staying awake in the $z^*$ time slot costs $d$ and does not provide any benefit, because the node will not transmit any packets for the remainder of the time horizon. □

**Lemma 2.8.** *If $z^* - N < t < z^*$ and $\boldsymbol{X}_t = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, the optimal control in slot $t$ to minimize $J_t^{\boldsymbol{\pi}}$ is described by the threshold decision rule:*

$$c \cdot \sum_{j=1}^{z^*-t} \left\{ p^j \left( T - t - j \right) \right\} - d \cdot \sum_{j=0}^{z^*-t} p^j \quad \underset{U_t^* = 1}{\overset{U_t^* = 0}{\lesseqgtr}} \quad 0 \; . \tag{2.16}$$

*Proof.* Redefine the policies $\hat{\boldsymbol{\pi}}$, $\bar{\boldsymbol{\pi}}$, and $\tilde{\boldsymbol{\pi}}$ once more as follows:

- $\hat{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot, and behave optimally thereafter.
- $\bar{\boldsymbol{\pi}}$ : go to sleep for $N$ slots, and behave optimally thereafter.
- $\tilde{\boldsymbol{\pi}}$ : stay awake for the $[t, t+1)$ slot. At each time $t+1, t+2, \ldots, z^*$ if there is a packet in the queue, transmit it; otherwise, go to sleep for the duration of the horizon.

We proceed by backward induction on $t$.

Base Case: $t = z^* - 1$

From Lemmas 2.6 and 2.7, we know that if there is a packet in the queue at $z^*$, the optimal action is to stay awake, but if there is not, the optimal action is to sleep. Furthermore, we know from Lemma 2.5 that the optimal policy is to sleep for the duration of the time horizon beginning at time $z^* + 1$, regardless of the queue size. This knowledge allows us to directly calculate the expected reward from following $\hat{\boldsymbol{\pi}}$ over $\bar{\boldsymbol{\pi}}$:

$$\begin{aligned} R_{z^*-1} &:= V_{z^*-1}^{\bar{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) - V_{z^*-1}^{\hat{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right), \\ &= -d + p \cdot [c \cdot (T - z^*) - d] \ . \end{aligned} \tag{2.17}$$

Note that for $t = z^* - 1$, the the RHS of (2.17) is equal to the LHS of (2.16). Therefore, if the LHS of (2.16) is greater than 0, we have $R_{z^*-1} > 0$, and the optimal action is $U_{z^*-1}^* = 1$. Alternatively, if the LHS of (2.16) is less than 0, we have $R_{z^*-1} < 0$, and the optimal action is $U_{z^*-1}^* = 0$. This completes the base case.

Induction Step: We now assume the optimal control action at state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ in slot $t$ is given by (2.16) for $t = l+1, l+2, \ldots, z^* - 1$, and show that the optimal control action at state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ in slot $l$ is also given by (2.16).

Define the index

$$w(t) := c \cdot \sum_{j=1}^{z^*-t} \left\{p^j \left(T - t - j\right)\right\} - d \cdot \sum_{j=0}^{z^*-t} p^j \ . \tag{2.18}$$

Then we have:

$$\begin{aligned} R_l &:= V_l^{\bar{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) - V_l^{\hat{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right), \\ &\geq V_l^{\bar{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) - V_l^{\tilde{\boldsymbol{\pi}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right), \\ &= w(l) \ . \end{aligned} \tag{2.19}$$

So, if $w(l) > 0$, then $R_l > 0$, and the optimal action is $U_l^* = 1$. Next, we consider the case of $w(l) < 0$. The following calculation demonstrates that $w(\cdot)$ is a nonincreasing function of $t$:

$$\begin{aligned}
w(t-1) - w(t) &= \left[ c \cdot \sum_{j=1}^{z^*-t+1} \left\{ p^j \left(T - t + 1 - j\right) \right\} - d \cdot \sum_{j=0}^{z^*-t+1} p^j \right] \\
&\quad - \left[ c \cdot \sum_{j=1}^{z^*-t} \left\{ p^j \left(T - t - j\right) \right\} - d \cdot \sum_{j=0}^{z^*-t} p^j \right] \\
&= p^{z^*-t+1} \cdot \left[ -d + c \cdot (T - z^*) \right] + c \cdot \sum_{j=1}^{z^*-t} p^j \\
&= p^{z^*-t+1} \cdot \left( c \cdot \left\lceil \frac{d}{c} \right\rceil - d \right) + c \cdot \sum_{j=1}^{z^*-t} p^j \\
&\geq 0, \ \forall t \in \left\{ z^* - N + 1, z^* - N + 2, \ldots, z^* - 1 \right\}. \quad (2.20)
\end{aligned}$$

From (2.20), it follows that

$$\begin{aligned}
&w(t_0) < 0 \text{ for some } t_0 \in \left\{ z^* - N + 1, z^* - N + 2, \ldots, z^* - 1 \right\} \\
&\quad \Rightarrow w(t) < 0 \ \forall t \in \left\{ t_0, t_0 + 1, \ldots, z^* - 1 \right\} \ . \quad (2.21)
\end{aligned}$$

By (2.21) and $w(l) < 0$, we know that $w(t) < 0$, for all $t \in \{l+1, l+2, \ldots, z^* - 1\}$. Thus, by the induction hypothesis, if $\mathbf{X}_t = \left[\begin{smallmatrix} 0 \\ 0 \end{smallmatrix}\right]$ for any $t \in \{l+1, l+2, \ldots, z^* - 1\}$, the node will sleep for the duration of the time horizon under the optimal policy, and the policies $\hat{\boldsymbol{\pi}}$ and $\tilde{\boldsymbol{\pi}}$ are equivalent. We conclude:

$$\begin{aligned}
R_l &:= V_l^{\bar{\boldsymbol{\pi}}}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) - V_l^{\hat{\boldsymbol{\pi}}}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \\
&= V_l^{\bar{\boldsymbol{\pi}}}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right) - V_l^{\tilde{\boldsymbol{\pi}}}\left( \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right), \\
&= w(l) < 0 \ ,
\end{aligned}$$

and the optimal action when $w(l) < 0$ is $U_l^* = 0$. □

### Is the Optimal Control Action at the Boundary State Nonincreasing Over Time?

Lemma 2.8 and its proof tell us that from slot $z^*-N+1$ until slot $z^*-1$, the optimal control action when the node is awake and the queue is empty is nonincreasing over time. We also know from Lemmas 2.5 and 2.7 that the optimal control is $U_t^* = 0$, for all $t \geq z^*$. Combining these, we know the optimal control action at $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is nonincreasing over time, from slot $z^* - N + 1$ until the end of the time horizon. The natural follow-up question to ask is whether or not the optimal control action at $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is necessarily monotonic over the entire duration of the time horizon. Intuitively, this might make sense if we extend the logic behind Lemma 2.5 to conclude that the marginal reward for serving a packet continues to increase as we move away from the end of the time horizon. However, as we explain further in Section 2.5.4, this intuition is not quite correct, as the following counterexample demonstrates.

**Counterexample 2.9.** Consider Problem (**P2.2**) with the parameters $T = 15$, $N = 3$, $c = 10$, $d = 21$, and $p = \frac{2}{3}$. The optimal sleep control policy at the boundary state $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$, computed through the dynamic program (2.10), is displayed in Figure 2.4. Clearly, the control action at the boundary state is not monotonic in time.

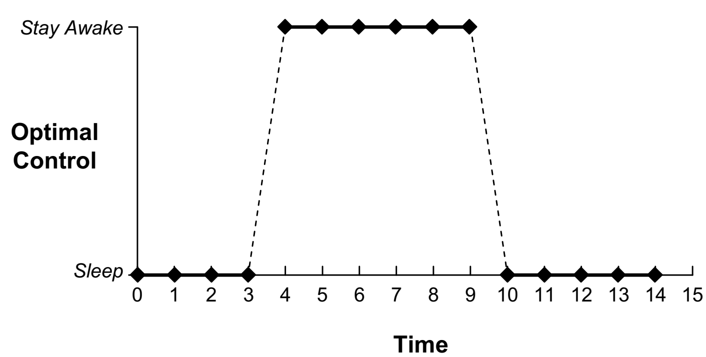


Figure 2.4. Optimal control actions at $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ when $T = 15$, $N = 3$, $c = 10$, $d = 21$, and $p = \frac{2}{3}$.

### Sufficient Condition Ensuring the Optimal Control Action at the Boundary State is Nonincreasing over Time

With such counterexamples in mind, we seek sufficient conditions for the optimal policy at the boundary state to be nonincreasing over the entire time horizon. Based on extensive numerical experiments, we believe the following conjecture is true, but have not yet been able to prove it.

**Conjecture 2.10.** *If the parameters of problem **(P2.2)** satisfy the following condition:*

$$\left(\frac{p}{1-p}\right) \cdot \left(\frac{N-1}{2}\right) > \frac{d}{c} \ , \tag{2.22}$$

*the optimal control action when the node is awake and the queue is empty is nonincreasing in time; i.e., if the expected cost-to-go $V_r\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]\right)$ is minimized by sleeping, then for all $t > r$, the expected cost-to-go $V_t\left(\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]\right)$ is minimized by sleeping.*

### Possible Optimal Policy Structures when the Optimal Control Action at the Boundary State is not Monotonic over Time

Assuming Conjecture 2.10 is true, we would also like to characterize the optimal policy at the boundary state when the parameters of Problem **(P2.2)** do not satisfy condition (2.22). One might think that the periodic nature of sleeping would lead to a periodic optimal policy at the boundary; however, based on numerical results, we believe the optimal control actions at the boundary state are still relatively "smooth," and can be characterized by the following conjecture.

**Conjecture 2.11.** *If the parameters of problem **(P2.2)** satisfy the following condition:*

$$\left(\frac{p}{1-p}\right) \cdot \left(\frac{N-1}{2}\right) < \frac{d}{c} \ , \tag{2.23}$$

*and if for some $k$, the optimal control at state $\boldsymbol{X}_k = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is $U_k^* = 0$ and the optimal control at state $\boldsymbol{X}_{k+1} = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is $U_{k+1}^* = 1$, then for all $0 \leq t < k$, the optimal control at state $\boldsymbol{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is $U_t^* = 0$.*

Conjecture 2.11 says that there can be at most one jump up in the optimal control from $U_t^* = 0$ at $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ to $U_{t+1}^* = 1$ at $\mathbf{X}_{t+1} = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$.

### 2.5.4 Discussion

In this section, we discuss the intuition behind Conjectures 2.10 and 2.11, their implications if they turn out to be true, and the challenges we face in proving them.

If Conjectures 2.10 and 2.11 turn out to be true, then they imply, in combination with Lemmas 2.5-2.8, that the sequence of optimal control actions at $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is of the form:

$$U_t^* = \begin{cases} 1 \text{ (stay awake)}, & \text{if } \lambda_1^* \leq t < \lambda_2^* \\ 0 \text{ (sleep)}, & \text{otherwise ,} \end{cases}$$

for some $\lambda_1^*, \lambda_2^* \in \{0, 1, \ldots, z^*\}$, with $\lambda_1^* \leq \lambda_2^*$. Specifically, only three structural forms of the optimal control action at the boundary state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ are possible. These are shown in Figure 2.5.

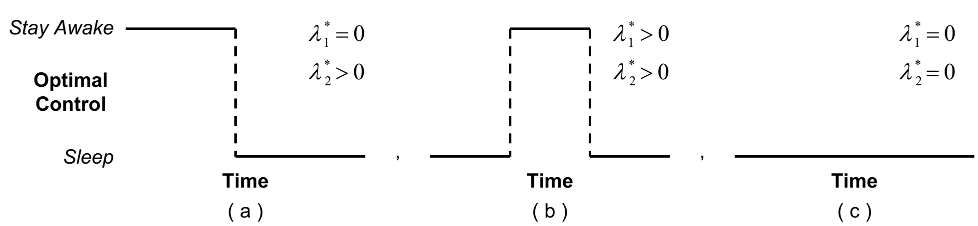


Figure 2.5. Possible structural forms for the optimal control actions over time at the boundary state $\mathbf{X}_t = \left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$.

Moreover, Conjecture 2.10 states that form (b) is not possible if condition (2.22) holds. Our numerical results not only support these conclusions, but also show the following:

**Observation 2.12.** If the time horizon is sufficiently long, then in fact the optimal control is of the form (a) if condition (2.22) holds, but of the form (b) or (c) if condition (2.23) holds.

We now attempt to provide some intuition as to why the optimal policy at the boundary state could be of form (b). The underlying tradeoff at the state $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ is between staying awake to reduce backlog costs and sleeping to avoid radio idling. In the infinite horizon problem, consider the two policies $\boldsymbol{\pi}_0$ (always awake) and $\boldsymbol{\pi}_1$ (sleep only at boundary state) described in Section 2.4, and assume the node is at state $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$ at some time $t$. In our model, the order in which packets are transmitted is of no importance (e.g. FIFO, LIFO). Therefore, let us assume that for every sample path, the packets arriving from time $t + N - 1$ onward are transmitted at exactly the same time under the two policies (by appropriate reordering of packets). Then the extra backlog charges incurred under $\boldsymbol{\pi}_1$ are entirely due to the packets arriving during $(t, t + N - 1)$. If there are $M$ arrivals during this period, the queue length at time $t + N$ under $\boldsymbol{\pi}_1$ is $M$ more than the queue length under $\boldsymbol{\pi}_0$. With each non-arrival after time $t + N - 1$, $\boldsymbol{\pi}_1$ "catches up" to $\boldsymbol{\pi}_0$ by one packet in the next slot. Eventually, after $M$ non-arrivals, the two policies will have transmitted the same number of packets, and both will end up back at the state $\begin{bmatrix} 0 \\ 0 \end{bmatrix}$. Figure 2.6 shows this comparison for a particular sample path in a problem with $N = 5$ and $M = 2$. If we compare the expected energy charges incurred by $\boldsymbol{\pi}_1$ during the $N$ unutilized slots of one such cycle to the expected extra backlog costs incurred by $\boldsymbol{\pi}_0$, we get (2.7), which describes the optimal stationary control action at the boundary state in the infinite horizon case.

Returning to the finite horizon problem, we see that (2.22) and (2.23) together are equivalent to (2.7). Let us now reconsider the two policies from the previous

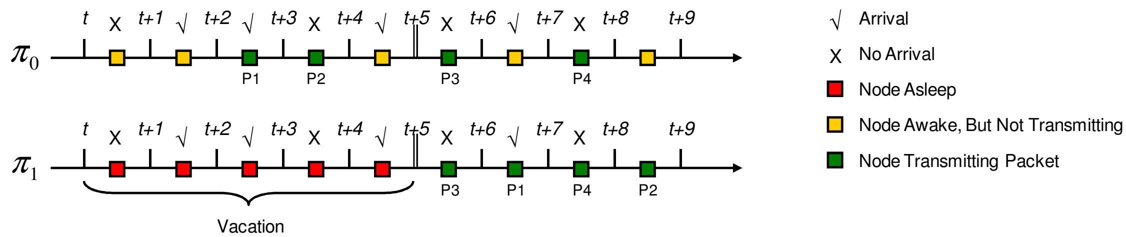


Figure 2.6. Comparison of policies $\boldsymbol{\pi}_0$ and $\boldsymbol{\pi}_1$ under a particular sample path. The sleep period length $N$ is equal to 5. Packets are reordered so that packets arriving after time $t+4$ are transmitted in the same slots under both policies. There are two arrivals between times $t$ and $t+4$, so the queue length under policy $\boldsymbol{\pi}_1$ catches up to the queue length under policy $\boldsymbol{\pi}_0$ in slot $t+8$, which is the slot following the second non-arrival after time $t+4$.

paragraph in the finite horizon context. The probability that the queue length under policy $\boldsymbol{\pi}_1$ catches up to the queue length under policy $\boldsymbol{\pi}_0$ before $z^*+1$, the time at which the node goes to sleep for good, increases as $t \to 0$. So Observation 2.12 makes intuitive sense as it just states that the optimal control at the boundary state in the finite horizon problem converges to the optimal control at the boundary state in the infinite horizon problem as we move farther and farther back from the end of the horizon.[1]

As we move closer to the end of the horizon (i.e., as $t$ increases), there is a higher probability of reaching time $z^*+1$ before the two policies reach the same state again. Any "extra" packets (amongst the $M$ arrivals in $(t, t+N-1)$) at $z^*+1$ will be charged for the rest of the time horizon, which has length $\left\lfloor \frac{d}{c} \right\rfloor$. This extra risk of going to sleep is likely the reason why form (b) is a possible form of the optimal policy. The middle bump in the policy plays the role of a "buffer zone" that incorporates the risk of untransmitted packets incurring charges throughout the entire shutdown zone at the end of the horizon.

[1]This is precisely the main idea of *turnpike* theorems [135]. It is common to search for a *planning horizon*, say $T_0$, such that for all times $t \leq T - T_0$, the optimal control policy is the same as the optimal stationary policy for the infinite horizon average expected cost problem. Hinderer and Hübner [70, 71, 75, 76] study such planning horizons for *undiscounted* stochastic dynamic programs, and provide technical conditions guaranteeing the existence of planning horizons and the validity of bounds on the planning horizons. Problem (**P2.2**) does not appear to satisfy these conditions; however, numerical experiments suggest a planning horizon does exist.

**Observation 2.13.** The structural forms in Figure 2.5 lie on a spectrum in the sense that changing one parameter at a time leads to a shift in the form of the optimal policy from either form (a) to form (b) to form (c), or from form (c) to form (b) to form (a). In particular, holding all other parameters constant, the form of the optimal policy shifts from (c) to (b) to (a) as we individually (or collectively) increase $p$, $N$, or $c$, but shifts from (a) to (b) to (c) as $d$ increases. Analogous statements can also be made concerning the dependence of the two thresholds, $\lambda_1^*$ and $\lambda_2^*$, on the problem parameters.

If they turn out to be true, the conjectures also lead to an efficient method to compute the optimal policy. Namely, if Conjecture 2.10 turns out to be true, then we can calculate the threshold $\lambda_2^*$ by computing the following index:

$$\tilde{w}(t) := V_t^{\bar{\bar{\boldsymbol{\pi}}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right) - V_t^{\hat{\tilde{\boldsymbol{\pi}}}}\left(\begin{bmatrix}0\\0\end{bmatrix}\right),$$

where the policies are defined as:

$\bar{\bar{\boldsymbol{\pi}}}$ : go to sleep for $N$ slots; in all subsequent slots, stay awake if and only if the queue is non-empty and the time is less than or equal to $z^*$.

$\hat{\tilde{\boldsymbol{\pi}}}$ : stay awake for the $[t, t+1)$ slot; in all subsequent slots, stay awake if and only if the queue is non-empty and the time is less than or equal to $z^*$.

Then, if $\tilde{w}(t) \leq 0$ for all $t$, let $\lambda_2^* := 0$; otherwise, let $\lambda_2^* := \max\{t : \tilde{w}(t) > 0\}$. Note that for $z^* - N + 1 \leq t < z^*$, the index $\tilde{w}(t)$ is the same as the index $w(t)$, defined in (2.18), because $\bar{\bar{\boldsymbol{\pi}}}$ is the same as $\bar{\boldsymbol{\pi}}$ and $\hat{\tilde{\boldsymbol{\pi}}}$ is the same as $\tilde{\boldsymbol{\pi}}$ in this region of time slots. For $z^* - 2N < t = z^* - N - l \leq z^* - N$, $\tilde{w}(t)$ can be computed as follows by conditioning on $m$, the number of arrivals before the first non-arrival, and $i$, the number of time slots after $z^* - l$ that the node goes to sleep for good under policy

$\bar{\bar{\boldsymbol{\pi}}}$:

$$\tilde{w}(t) := -d + p^{l+1}\left[c(l+1)(N-1)\right] + \sum_{j=2}^{N} p^{j+l}\left[-d + c\left(T - z^* + N - j\right)\right]$$
$$+ \sum_{m=0}^{l}\left\{p^m(1-p)\sum_{i=m}^{l+1}\left[\Psi_{l,i,m}\cdot\Gamma_{l,i,m}\right]\right\},$$

where

$$\Psi_{l,i,m} := \begin{cases} (1-p)^{N-1}p^{i-m}\left[\binom{N-1}{i-m} + \sum_{j=1}^{i-m-1}\left\{\binom{N-1}{j}\sum_{i=1}^{i-m-j}\binom{j}{i}\right\}\right], \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } i \in \{m, m+1, \dots, l\} \\ 1 - \sum_{i=m}^{l}\Psi_{l,i,m}, \ \text{if } i = l+1 \\ 0, \ \text{otherwise} \end{cases}$$

and

$$\Gamma_{l,i,m} := \begin{cases} mc(N-1) - c(i-m) + \sum_{j=1}^{l-i}\left\{p^j\left[-d + c\left(T - z^* + l - i - j\right)\right]\right\}, \\ \qquad\qquad\qquad\qquad\qquad\qquad \text{if } i \in \{m, m+1, \dots, l\} \\ mc(N-1) - c(l-m) + d - c\left(T - z^*\right), \ \text{if } i = l+1 \\ 0, \ \text{otherwise} \end{cases}.$$

We have not yet computed the index $\tilde{w}(t)$ in closed form for $t \leq z^* - 2N$.

If the sufficient condition (2.22) holds, calculating $\lambda_2^*$ in the above fashion completes the characterization of the optimal policy. When (2.22) does not hold, $\lambda_1^*$ can be calculated similarly by creating a second index that is a function of $t$ and $\lambda_2^*$. This methodology of leveraging the structural results to determine the optimal policy is much simpler computationally than computing the entire optimal policy through the dynamic program (2.10).

We now discuss briefly the challenges we face in proving Conjectures 2.10 and 2.11. In stochastic control problems, it is often the case that we can infer structural properties of the optimal control from certain properties of the value function, such as monotonicity, convexity, and supermodularity (see for example [72] and [149] for description of such techniques). In particular, supermodularity and submodularity are used throughout the queuing theory literature to prove the optimal control policy has a threshold form (for one such example, see [5]). However, the thresholds in these cases are usually thresholds in queue length (i.e., one control action is optimal if the queue length is above a critical number and another is optimal if it is below the critical number), as opposed to thresholds in time. In our model, such a result is true, but fairly trivial. We can see from Lemmas 2.5 and 2.6 that not only is the optimal control monotonic in queue length at each time $t$, but the threshold is always 0 (always stay awake), 1 (stay awake only if queue is non-empty), or $\infty$ (never stay awake). We are looking to strengthen this result by finding a sufficient condition for the optimal control to be monotonic in time; i.e., have those critical queue length numbers at each slot be nondecreasing over the entire time horizon. We are not aware of any previous works in which modularity properties are used to show the optimal control policy is monotonic in time.

Unfortunately, in our case, neither the value function nor its components display the nice properties we desire, even when we restrict the parameter sets to those satisfying (2.22). For instance, we can reduce part of the dynamic program (2.10) to the following form:

$$V_t \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \min \{\alpha_t, \beta_t\} \ , \tag{2.24}$$

where $\alpha_t$ is the expected cost-to-go under $U_t = 1$, and $\beta_t$ is the expected cost-to-go under $U_t = 0$. One way to show that the optimal control at the boundary state is of the form (a) or (c) (i.e., monotonic in time) when condition (2.22) is satisfied would be to show:

$$(2.22) \Rightarrow \beta_t - \beta_{t+1} < \alpha_t - \alpha_{t+1},\ \forall t \leq z^* - N \ . \tag{2.25}$$

Note that (2.25) would imply:

$$\beta_t < \alpha_t \Rightarrow \beta_{t+1} < \alpha_{t+1} \ , \tag{2.26}$$

which guarantees the optimal policy at state $\left[\begin{smallmatrix}0\\0\end{smallmatrix}\right]$ is nonincreasing in time. However, as we see in Figure 2.7, (2.25) is not necessarily true. We have tried numerous other approaches to prove Conjecture 2.10, to no avail.

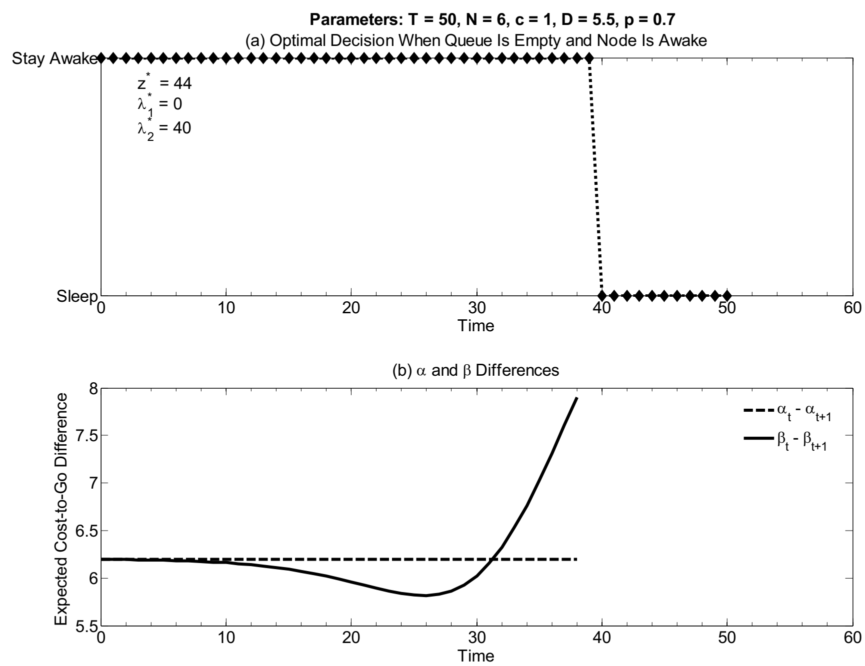


Figure 2.7. Expected cost-to-go differences under the two available control actions.

## 2.6 Summary

In this chapter we studied the problem of optimal sleep scheduling for a wireless sensor network node, and considered two separate discrete-time optimization problems. For the infinite horizon average expected cost problem, we demonstrated the existence of an optimal stationary Markov policy, and completely characterized the optimal control at each state in the state space. For the finite horizon expected cost problem, we completely characterized the optimal policy for all states except the boundary state where the node is awake and the queue is empty. One significant difference from the infinite horizon was the existence of a "shutdown" period at the end of the time horizon in which the queue stops serving packets, regardless of the queue size. We hypothesized a sufficient condition to guarantee the optimal control is nonincreasing over time when the queue is empty and the node is awake. Based on extensive numerical experiments, we also conjectured that even when this sufficient condition does not hold, there is at most one jump in the optimal control, providing a single "buffer zone."

# *Chapter 3*

## Dynamic Clock Calibration via Temperature Measurement

In this chapter, we study a clock calibration problem that arises from an ultra-low power sensor node platform. This sensor platform, built around the Phoenix Processor [64, 96], was initially designed as an intraocular pressure monitoring system, but could potentially be used in a host of applications from environmental monitoring to surveillance.

In this platform, energy consumption is managed through three modes, referred to as the *sleep*, *processor*, and *radio* modes. In the sleep mode, designed to conserve energy, it consumes on the order of 1-10 pW. In the processor and radio modes, it consumes on the order of 1 $\mu$W and 1 mW, respectively. Thus, when the sensor does not need to perform communication or sensing tasks, it is put into sleep mode, with only an *ultra-low power clock/timer* running. Typical operation is to stay in the sleep mode for extended periods of time (10-60 minutes), wake up briefly (less than a second), and go back to sleep. The ultra-low power clock essentially functions as an alarm clock to time out the desired sleep period and wake the node up at the appropriate time.

The power savings of the sleep mode, however, come at the expense of relatively low timing accuracy. Specifically, the accuracy of the ultra-low power clock is dependent on the ambient temperature and supply-voltage. The *processor clock* that is activated whenever the processor is turned on (i.e., when the node is in the processor mode) is more accurate, and the *radio (quartz) clock* that is activated when the radio transceiver is turned on (i.e., when the node is in the radio mode) is much more accurate.

Such inaccuracies in the ultra-low power clock can affect its scheduled wake up time (e.g., to take a measurement or to communicate with another node). In addition to taking measurements at the wrong time, this may lead to wasted energy consumption as a result of two unsynchronized devices trying to communicate. For instance, consider a node that turns on its radio to send data to a second node, but the sending node has woken up before its scheduled time. Its radio idles while waiting for the second node to turn on its radio, resulting in unnecessary energy consumption. Since the radio mode consumes a lot more power than the processor and sleep modes, even small improvements in clock accuracy in the sleep mode can result in significant energy savings in the radio mode.

It is therefore crucial to be able to accurately calibrate the ultra-low power clock while in the sleep mode. In this chapter, we examine a novel approach that exploits the temperature dependence of the ultra-low power clock by occasionally turning the processor on to take a temperature reading. Each temperature reading translates into a speed at which the ultra-low power clock ticks, a relationship that can be obtained fairly reliably in a lab setting. Such knowledge about the clock speed is then used in combination with the local clock time to obtain a better estimate of the real time that has elapsed. In essence, this approach trades a little extra energy consumption in taking temperature measurements for greater energy savings in communication. To the best of our knowledge, this is the first study on using sensing not as a means of data gathering for some higher level application, but as a way of self-improving the node's own performance (in this case its timing performance).

Because turning the processor on to take temperature measurements does consume energy, we would like to limit the number of such measurements. The problem arises as to how to dynamically schedule a limited number of temperature measurements in a manner most useful to improving the accuracy of the ultra-low power clock. We formulate this measurement scheduling problem as a stochastic control problem. Physically, this scheduling would be implemented in the processor, which would wake up, take a measurement, decide the number of clock ticks until the next wake-up time, and program the timer accordingly.

The remainder of the chapter is organized as follows. In Section 3.1, we present an abstraction of the problem and an illustrative example to motivate the decision between modeling the underlying time scale as continuous or discrete. In Section 3.2, we formulate an optimization problem based on a continuous underlying time scale. In Section 3.3, we formulate a second problem with a discrete underlying time scale. In Section 3.4, we compute the optimal control policy for a simple toy example. Section 3.5 concludes the chapter.

## 3.1 Problem Description

In this section, we present an abstraction of the synchronization problem outlined in the previous section. Our goal is to have the ultra low-power timer measure a fixed amount of time, $T$, as accurately as possible. In doing so, it is allowed to take up to $\bar{N}$ ambient temperature measurements. The control algorithm residing in the processor (also referred to below as the *controller* or *scheduler*) decides when these measurements are taken. We assume it knows the initial ambient temperature, as well as a statistical description of the stochastic temperature process.

Associated with each temperature is the frequency of the ultra-low power clock in terms of clock cycles (also referred to below as *clock ticks*) per unit of time. We assume the mapping $f : \mathcal{W} \rightarrow \hat{\mathcal{W}}$ describing the frequency associated with each temperature is known. Here, $\mathcal{W}$ is the space of possible temperatures, and $\hat{\mathcal{W}}$ is the space of possible frequencies. Figure 3.1 shows a plot of the function $\frac{1}{f}$.[1]

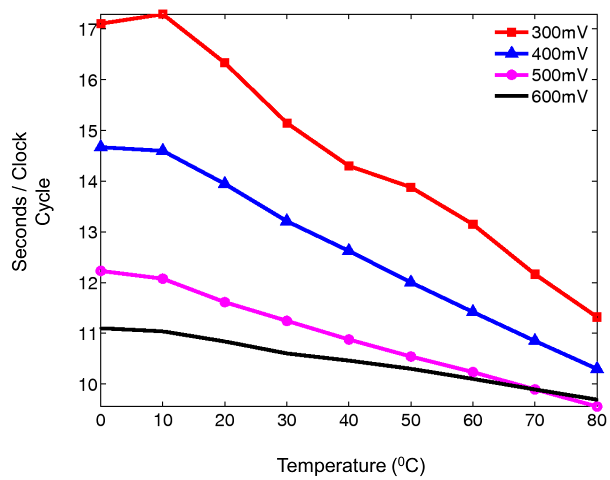


Figure 3.1. Functions mapping temperature to clock period at different supply voltages. The function $f$ mapping temperature to clock frequency is 1 divided by one such function.

At the beginning of the time horizon, the ultra-low power clock is synchronized with the actual time zero. In addition to the statistics of the temperature process, the initial temperature, and the function $f$ mapping temperature to frequency, the following information is available to the scheduler throughout the sleep period: (i) all prior scheduling decisions; (ii) all temperature measurements taken to date; and (iii) the number of clock cycles that have elapsed since time zero. The scheduler's tasks are to use this information to schedule each successive measurement, and to decide when to wake up and declare that $T$ units of actual time have elapsed. The performance criterion is a distortion function $\rho(T, \hat{T})$, where $\hat{T}$ is the (actual) time

[1] Data courtesy of Y. Lin, D. Blaauw, and D. Sylvester [96]. The timer consumes on the order of 1-10 pW ($10^{-12}$ to $10^{-11}$ W) at 300 mV supply voltage. Also note this is a very slow clock with one cycle per 10+ seconds.

at which the scheduler declares $T$ units of time have elapsed. The objective is to design measurement scheduling and declaration policies that minimize the expected value of this performance criterion.

The above description results in a decision problem. Decision problems commonly take time as a given, on which discrete-time and continuous-time models are built. The unusual feature of the problem at hand is that time is the very thing we are trying to estimate, which makes the formulation quite tricky. In particular, the environmental random process describing temperature evolution affects the frequency of the clock, which in turn affects the local time. The timing of the decisions is based on the local time, rather than the real time. This interplay between the temperature process (defined in real time) and the control process (defined in local time) results in significant conceptual and technical challenges.

Before proceeding to the mathematical formulation, we present an overly simple example to show that the most natural choice of underlying time scale is the continuous time scale. Consider an environment with only two possible temperatures, $w^1$ and $w^2$. When the temperature is $w^1$, the ultra-low power clock ticks once every 2 seconds. When the temperature is $w^2$, the clock ticks once every 4 seconds. Consider the following temperature realization: $w^1$ for 2 seconds, then $w^2$ for 5 seconds, and then $w^1$ for 5 seconds, as shown in the upper left graph in Figure 3.2. From the sample path of the temperature and the frequency mapping $f$, we determine the sample path of the clock frequency, shown in the lower left graph in Figure 3.2. Integrating the clock frequency (ticks per second) from 0 to $t$ yields the total number of clock ticks elapsed up to actual time $t$, shown in the graph on the right side of Figure 3.2. For this sample path of the temperature process, the first clock tick occurs at 2 seconds, and the second clock tick occurs at 6 seconds. When, after 7 total seconds, the temperature switches back to $w^1$, one fourth of the third clock tick has elapsed. Thus, the next clock tick occurs at 8.5 seconds, after another three fourths of a clock tick. The problem with using discrete time units of one second is that there is no such time as 8.5 seconds. We therefore start by considering time to be continuous, although we revisit the validity of a discrete-time model with some extra assumptions in Section 3.3.

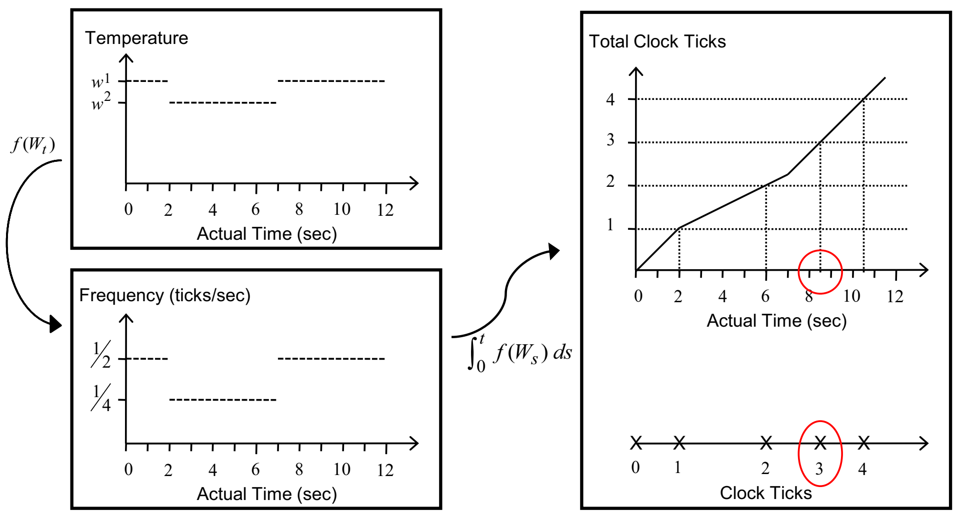


Figure 3.2. Illustrative example of how the clock ticks may not coincide with discrete time steps. The actual times of the clock ticks are determined from a sample path of the temperature process and the frequency mapping $f$, by converting temperature into frequency and integrating frequency over time.

## 3.2 Problem Formulation - Continuous Time

In this section, we take the underlying time scale to be continuous. We model the ambient temperature process, $\{W(t)\}_{t\geq 0}$, as a continuous-time homogeneous Markov process with finite state space $\mathcal{W}$, known initial temperature $w_0$, and known transition semigroup $\{\mathbf{P}(t)\}_{t\geq 0}$, where $\mathbf{P}(t) := \{p_{ij}(t)\}_{i,j\in\mathcal{W}}$ and

$$p_{ij}(t) := \Pr\left(W(t_0+t) = j \mid W(t_0) = i\right) \ .$$

We formulate the problem as a partially observed semi-Markov decision process (POSMDP). We then provide a high-level overview of how to reduce this POSMDP first to an equivalent partially observed Markov decision process (POMDP), and then to two equivalent discrete-time Markov decision processes (MDP's). Finally, we present the dynamic programming equations that solve the latter of these two MDP's, and highlight the most computationally intense steps in the dynamic program.

### 3.2.1 Formulation as a POSMDP

Recall that a semi-Markov decision process (SMDP) is a generalization of a discrete-time MDP that models the system evolution in continuous time, and allows the decision epochs to occur at random times. In our problem, the decision epochs of the SMDP occur at the times of the local clock ticks. We define a random process $\{C_t\}_{t\geq 0}$ by $C_t := \int_0^t f(W_s)\,ds$, which represents the (fractional) number of clock cycles that occur between the beginning of the time horizon and the actual time $t$. The decision epochs of the SMDP occur when $C_t \in \mathbb{Z}_+ := \{0, 1, 2, \ldots\}$. We represent the times of these clock ticks by the random variables $0 = \sigma_0 \leq \sigma_1 \leq \sigma_2 \ldots$, and let $\bar{\sigma}_k := \sigma_k - \sigma_{k-1},\ k = 1, 2, \ldots$ be random variables representing the inter-tick times. The conditional cumulative distribution function (cdf) of the real time of the $l^{th}$ clock tick, given the initial temperature, is:

$$\begin{aligned} F_{\sigma_l|W_0}(t \mid w) &:= \Pr\left(\sigma_l \leq t \mid W_0 = w\right) \\ &= \Pr\left(C_t \geq l \mid W_0 = w\right) \\ &= 1 - \Pr\left(\int_0^t f(W_s)\,ds < l \mid W_0 = w\right). \end{aligned}$$

For each $l \in \{1, 2, \ldots\}$, we denote the probability density function (pdf) induced by $F_{\sigma_l|W_0}(t \mid w)$ as $f_{\sigma_l|W_0}(t \mid w)$.

At $\sigma_k$, the random time of the $k^{th}$ clock tick, we define the state of the SMDP to be the triplet $\mathbf{S}_k := (X_k, W_k, N_k)$, where $X_k$ is the actual time elapsed; $W_k$ is the ambient temperature; and $N_k$ is the number of temperature measurements taken between the beginning of the horizon and the $k^{th}$ clock tick (inclusive of a measurement scheduled for the $k^{th}$ clock tick). The sample space of the triplet $\mathbf{S}_k$ is $\mathcal{S} := I\!R_+ \times \mathcal{W} \times \mathcal{N}$, where $\mathcal{W}$ is the finite space of possible temperatures, and $\mathcal{N} := \left\{0, 1, \ldots, \bar{N}\right\}$. At each tick $k$, the state corresponds to the state of the underlying continuous-time process, which Puterman [118] calls the "natural process;" i.e., $(X_k, W_k, N_k) = \left(X_{\sigma_k}, W_{\sigma_k}, N_{\sigma_k}\right), \forall k$.

Of course, this state is not perfectly observed by the controller, as the controller never observes the actual time $X_k$, and only observes the ambient temperature $W_k$ when it decides to take a measurement. The number of measurements taken and scheduled to date, $N_k$, is known perfectly by the controller, as we assume the controller remembers all past decisions. We represent the controller's observation at the $k^{th}$ clock tick by the random vector $\mathbf{Y}_k$, with sample space $\mathcal{Y} = \mathbb{Z}_+ \times \{\mathcal{W} \cup -1\}$. Here, the first element of the observation is the index of the clock tick, and the second element is the temperature measurement. We assume temperature measurements are correct with probability 1; i.e., $\mathbf{Y}_k = (k, W_k)$ if a measurement is taken at tick $k$. If no measurement is taken, then $\mathbf{Y}_k = (k, -1)$. Including the index of the clock tick in the observation space is a bit redundant; however, we do this to emphasize that i) the controller knows the number of clock ticks to date (indexed by $k$), but ii) the controller does not know the actual time at which each clock tick occurs (indexed by $t$).

The timing at each decision epoch, shown in Figure 3.3, is as follows. Immediately after the $k^{th}$ clock tick, the controller receives observation $\mathbf{Y}_k$. It then makes two decisions. First, it decides whether or not to declare that $T$ time units have elapsed (after $k$ clock ticks). If it decides to declare, the controlled sleep process is stopped, and the node wakes up. Otherwise, the controller also decides whether or not to take a temperature measurement at tick $k+1$. Thus, the decision space is

$$\mathcal{U} := \{1; (0,0)\,; (0,1)\}\,.$$

Here, 1 means "declare that $T$ time units have elapsed;" $(0,0)$ means "do not declare that $T$ time units have elapsed and do not take a measurement at clock tick $k+1$;" and $(0,1)$ means "do not declare that $T$ time units have elapsed and take a measurement at clock tick $k+1$." If $N_k$ is equal to $\bar{N}$, the maximum number of measurements allowed, the available decisions for $\mathbf{U}_k$ are $\bar{\mathcal{U}} := \{1; (0,0)\}$; otherwise, all decisions are available. We denote by $\mathcal{U}(\mathbf{s})$ the decisions that are available at state $\mathbf{s}$, and $U_k^2$ refers to the second component of the control decision (the measurement decision at the following tick).

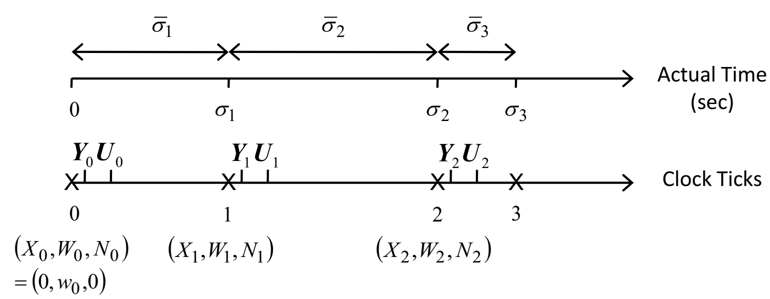


Figure 3.3. The timing of observations and decisions at each epoch. The decision $\mathbf{U}_k$ is made after observing $\mathbf{Y}_k$. $\mathbf{U}_k$ determines whether to declare that $T$ time units have elapsed at clock tick $k$, and (if the process is not stopped) whether to take a temperature measurement at the $(k+1)^{st}$ tick.

Next, we describe the probabilistic *state transition law*. If $\mathbf{U}_k = 1$, the process is stopped. Otherwise, the time, $\bar{\sigma}_{k+1}$, until the next clock tick, and the state, $\mathbf{S}_{k+1}$, at the next clock tick have the following joint distribution, conditioned on the current state and scheduling decision:

$$
\begin{aligned}
&Q\left(B_1, B_2, B_3, B_4 \mid x_k, w_k, n_k, \mathbf{u}_k\right) \\
&= \Pr\left(\begin{array}{c|c} X_{k+1} \in B_1, W_{k+1} \in B_2, & X_k = x_k, W_k = w_k, \\ N_{k+1} \in B_3, \bar{\sigma}_{k+1} \in B_4 & N_k = n_k, \mathbf{U}_k = \mathbf{u}_k \end{array}\right) && (3.1) \\
&= \sum_{w_{k+1} \in B_2} \sum_{n_{k+1} \in B_3} \int_{\bar{\sigma}_{k+1} \in B_4} \left(\begin{array}{l} \mathbb{1}_{\left\{n_{k+1} = n_k + u_k^2\right\}} \cdot \mathbb{1}_{\{x_k + \bar{\sigma}_{k+1} \in B_1\}} \\ \cdot\, p_{w_k, w_{k+1}}\left(\bar{\sigma}_{k+1}\right) \cdot f_{\sigma_1 | W_0}\left(\bar{\sigma}_{k+1} \mid w_k\right) d\bar{\sigma}_{k+1} \end{array}\right) && (3.2)
\end{aligned}
$$

for Borel sets $B_1 \in \mathcal{B}(I\!R_+), B_2 \in \mathcal{B}(\mathcal{W}), B_3 \in \mathcal{B}(\mathcal{N}), B_4 \in \mathcal{B}(I\!R_+)$, and for all $(\mathbf{s}_k, \mathbf{u}_k)$ such that $\mathbf{u}_k \in \mathcal{U}(\mathbf{s}_k)$. Equation (3.2) follows from the facts that i) $N_{k+1}$ is a function of $N_k$ and $\mathbf{U}_k$; ii) $X_{k+1}$ and $W_{k+1}$ are independent of $N_k$ and $\mathbf{U}_k$; and (iii) $f_{\sigma_1|W_0} = f_{\bar{\sigma}_{k+1}|W_k}$, by the homogeneity of the temperature process. Recall that $p_{w_k,w_{k+1}}\left(\bar{\sigma}_{k+1}\right)$ is the probability the temperature process jumps from $w_k$ to $w_{k+1}$ in $\bar{\sigma}_{k+1}$ real time units, and $f_{\sigma_1|W_0}\left(\cdot \mid w_k\right)$ is the distribution of the time between two consecutive ticks, given the initial temperature $w_k$.

At the beginning of the time horizon, the controller knows that the actual time is zero ($X_0 = 0$); no measurements have been taken ($N_0 = 0$); and the initial temperature is $w_0$. This assumption is reasonable, as the processer and radio were likely on at the end of the previously timed period.

We define an *observable history* up to the $k^{th}$ tick as:

$$\mathbf{h}_k := \big(\mathbf{y}_0, \mathbf{u}_0, \mathbf{y}_1, \mathbf{u}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{k-1}, \mathbf{u}_{k-1}, \mathbf{y}_k\big) \in \mathcal{H}_k \ ,$$

where $\mathcal{H}_k = \left(\mathcal{Y} \times \mathcal{U}\right)^k \times \mathcal{Y}$ is the space of possible histories up to the $k^{th}$ tick. We let $\mathcal{H}_0 = \mathcal{Y}$.

A *policy* is defined as a sequence $\boldsymbol{\gamma} := \{\gamma_k\}_{k=0}^{\infty}$, where for each $k$, $\gamma_k : \mathcal{H}_k \to \mathcal{P}\left(\mathcal{U}\right)$ maps the observable history up to the $k^{th}$ clock tick into the space of probability distributions on the decision space $\mathcal{U}$. A policy $\boldsymbol{\gamma}$ is *admissible* if for all $k$, $\gamma_k$ maps all histories $\mathbf{h}_k$ with $\sum_{i=0}^{k-1} u_k^2 = \bar{N}$ into probability distributions on $\bar{\mathcal{U}}$; i.e., if no measurements remain, the policy chooses control decisions 1 or $(0,0)$ with probability 1. We denote the space of all such admissible policies by $\Gamma$.

The quality of a temperature measurement scheduling and declaration policy is measured by a distortion function $\rho : I\!R_+^2 \to I\!R_+$ that determines the cost of declaring that $z$ time units have elapsed after $\hat{z}$ time units. For example, we could use the $L_1$ distortion function $\rho(z,\hat{z}) = |z - \hat{z}|$, or the square error distortion $\rho(z,\hat{z}) = (z-\hat{z})^2$. From this distortion, we define the cost of an admissable policy $\boldsymbol{\gamma}$ as:

$$J^{\gamma}(w_0) := I\!E^{\gamma}\left[\rho(T, X_\tau) \mid X_0 = 0, N_0 = 0, W_0 = w_0\right] \ ,$$

where $\tau$ is the random stopping time at which the controller declares that $T$ time units have elapsed. By assumption, the sample space $\mathcal{W}$ is finite. Accordingly, there exists a maximum frequency (clock cycles per unit time), which we denote by $\omega_{max} := \max_{w \in \mathcal{W}} \{f(w)\}$. The maximum number of clock cycles the controller needs to consider waiting before declaring $T$ time units have elapsed is therefore $\bar{K} := \lceil T \cdot \omega_{max} \rceil$. So we define the stopping time $\tau$ as:

$$\tau := \min\left\{\bar{K}, \min\{k : \mathbf{U}_k = 1\}\right\} \ .$$

This definition ensures there is a finite optimal stopping time.

We wish to find an optimal control policy $\boldsymbol{\gamma}^*$ such that:

$$J^{\gamma^*}(w_0) = J^*(w_0) := \inf_{\boldsymbol{\gamma} \in \Gamma} J^{\gamma}(w_0) \ , \ \forall w_0 \in \mathcal{W} \ . \tag{3.3}$$

We refer to the above problem as Problem (POSMDP).

### 3.2.2 Transformation to Equivalent Problems

We now provide a high-level overview of how to reduce Problem (POSMDP) to a series of equivalent problems. We start by defining a POMDP that describes the evolution of the system at the clock ticks. The only component of this POMDP, referred to as Problem (POMDP-1), that is different from Problem (POSMDP) is the probabilistic state transition law, which is given by $Q\left(B_1, B_2, B_3, I\!R_+ \mid x_k, w_k, n_k, \mathbf{u}_k\right)$, where $Q$ is defined in (3.1). The equivalence of Problems (POMDP-1) and (POSMDP) follows from the fact that all control decisions and cost assessments in Problem (POSMDP) occur at the clock ticks.

Next, we transform Problem (POMDP-1) into a completely observable MDP, which we call Problem (MDP-1), with state equal to the conditional probability distribution of the POMDP state $\mathbf{S}_k$, given all decisions and observations to date. We omit the detailed description of Problem (MDP-1), as the transformation is a standard procedure (see, e.g., [42, pp. 214-217], [67, pp. 86-90]).

The next transformation is to a second equivalent discrete-time MDP, which we refer to as Problem (MDP-2). The main idea underlying the transformation from the previous MDP to this one is as follows. If at clock tick $k$ in Problem (MDP-1), the controller decides not to declare that $T$ time units have elapsed and not to take a measurement at clock tick $k+1$, then it gains no useful information before having to choose its next control decision at clock tick $k+1$. Thus, it can choose the control decision for clock tick $k+1$ equally well at the current clock tick $k$. By extending the same logic, without loss of optimality, it can actually decide at the current clock tick $k$ how many clock ticks to wait before taking the next measurement or declaring $T$ time units have elapsed.

Accordingly, we define a new time scale, indexed by $m$, to be the number of measurements taken so far (note the difference from the above problems, where time $k$ is the number of clock ticks). Here, $m = 0$ denotes the start of the horizon, $m = 1$ denotes the process just after the first temperature measurement, and so forth. For

$m = 0, 1, \ldots, \bar{N}$, the state consists of the conditional distribution, $\pi_{\tilde{X}_m}$, of the actual time elapsed given the history, and the most recent temperature measurement, $\tilde{w}_m$. The state space is $\tilde{\mathcal{S}} := \mathcal{P}(I\!R_+) \times \mathcal{W}$. The decision space is $\tilde{\mathcal{U}} := \{0, 1, 2, \ldots, \bar{K}\}$. At all $m$, decision $\tilde{U} = 0$ means "declare that $T$ units of time have elapsed." For $m = 0, 1, \ldots, \bar{N} - 1$, decision $\tilde{U} = l$ for some $l \in \{1, 2, \ldots, \bar{K}\}$ means "wait $l$ clock ticks before taking the next temperature measurement." When $m = \bar{N}$, no temperature measurements remain, and decision $\tilde{U} = \bar{l}$ for some $\bar{l} \in \{1, 2, \ldots, \bar{K}\}$ means "wait $l$ clock ticks before declaring that $T$ time units have elapsed." Figure 3.4 compares the time scales for Problems (MDP-1) and (MDP-2).

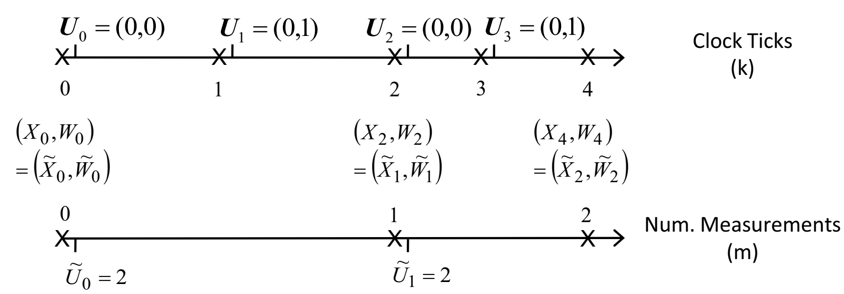


Figure 3.4. Example sample path to compare the time scales for Problems (MDP-1) and (MDP-2). The top timeline is based on $k$, the number of clock cycles elapsed, and the bottom timeline is based on $m$, the number of measurements taken. For both timelines, the first measurement is taken at clock tick 2, and the second measurement is taken at clock tick 4.

Note that while the states of Problem (MDP-2) belong to the continuous space $\mathcal{P}(I\!R_+) \times \mathcal{W}$, only a finite number of states in this space are reachable. This is due to the fact that the equivalent Problem (POMDP-1) has a finite horizon, finite decision space, and finite observation space. Thus, we can focus on this finite set of reachable states. By a standard result (see, e.g., [42, 67]), an optimal policy exists, and it can be found through the following dynamic program:

$$V_m\left(\pi_{\tilde{X}_m}, \tilde{w}_m\right) = \min \left\{ \begin{array}{l} \mathbb{E}\left[\rho\left(T, \tilde{X}_m\right) \mid \pi_{\tilde{X}_m}\right], \\ \min\limits_{l \in \left\{1,2\ldots,\bar{K}\right\}} \left\{ \mathbb{E}\left[ \begin{array}{l} V_{m+1}\left(\Pi_{\tilde{X}_{m+1}}, \tilde{W}_{m+1}\right) \mid \\ \Pi_{\tilde{X}_m} = \pi_{\tilde{X}_m}, \\ \tilde{W}_m = \tilde{w}_m, \tilde{U}_m = l \end{array} \right] \right\} \end{array} \right\},$$

$$m = 0, 1, \ldots, \bar{N} - 1,$$

$$V_{\bar{N}}\left(\pi_{\tilde{X}_{\bar{N}}}, \tilde{w}_{\bar{N}}\right) = \min \left\{ \begin{array}{l} \mathbb{E}\left[\rho\left(T, \tilde{X}_{\bar{N}}\right) \mid \pi_{\tilde{X}_{\bar{N}}}\right], \\ \min\limits_{\bar{l} \in \left\{1,2\ldots,\bar{K}\right\}} \left\{ \mathbb{E}\left[ \begin{array}{l} \rho\left(T, Z_{\bar{l}}\right) \mid \Pi_{\tilde{X}_{\bar{N}}} = \pi_{\tilde{X}_{\bar{N}}}, \\ \tilde{W}_{\bar{N}} = \tilde{w}_{\bar{N}}, \tilde{U}_{\bar{N}} = \bar{l} \end{array} \right] \right\} \end{array} \right\}.$$

$\tilde{V}_m$ represents the expected cost-to-go just after the $m^{th}$ measurement is taken; $\pi_{\tilde{X}_m}$ represents the conditional pdf of the actual time just after the $m^{th}$ measurement is taken; and $\tilde{w}_m$ represents the $m^{th}$ temperature reading. The first term in each outer minimization, $\mathbb{E}\left[\rho\left(T, \tilde{X}_m\right) \mid \pi_{\tilde{X}_m}\right]$, represents the conditional expected cost of stopping after the $m^{th}$ measurement. For $m = 0, 1, \ldots, \bar{N} - 1$, the second term in the outer minimization represents the expected cost if the scheduler waits $l$ clock ticks before taking the next measurement. For $m = \bar{N}$, the second term in the outer minimization represents the expected cost if the scheduler waits an additional $\bar{l}$ clock ticks before declaring that $T$ time units have elapsed. $Z_{\bar{l}}$ is a random variable describing the actual time $\bar{l}$ clock ticks after the $\bar{N}^{th}$ temperature measurement is taken; i.e., $Z_{\bar{l}} = \tilde{X}_{\bar{N}} + \sum_{i=k_{\bar{N}}+1}^{k_{\bar{N}}+\bar{l}} \bar{\sigma}_i$, where $k_{\bar{N}}$ is the clock tick at which the $\bar{N}^{th}$ measurement is taken. Due to the homogeneity of the Markov process $\{W(t)\}_{t \geq 0}$, $Z_{\bar{l}}$ has the following conditional pdf, given the realization $\tilde{w}_{\bar{N}}$ of the final temperature measurement and the conditional pdf, $\pi_{\tilde{X}_{\bar{N}}}$, of the time elapsed up to the $\bar{N}^{th}$

measurement:

$$f_{Z_{\bar{l}}|\Pi_{\tilde{X}_{\bar{N}}},\tilde{W}_{\bar{N}}}\left(z \mid \pi_{\tilde{X}_{\bar{N}}}, \tilde{w}_{\bar{N}}\right) = \pi_{\tilde{X}_{\bar{N}}}(x) * f_{\sigma_{\bar{l}}|W_0}\left(t \mid \tilde{w}_{\bar{N}}\right).$$

While the above dynamic program is conceptually straightforward, it is difficult to solve from a computational standpoint. The heart of the matter is in updating the conditional distribution of elapsed time after the $m^{th}$ temperature measurement, $\pi_{\tilde{X}_m}$, to the corresponding distribution after the $(m+1)^{st}$ temperature measurement, based on (i) $\tilde{w}_m$, the $m^{th}$ temperature reading; (ii) $\tilde{w}_{m+1}$, the $(m+1)^{st}$ temperature reading; and (iii) $\tilde{u}_m$, the number of clock ticks in between measurements, as chosen by the controller. Since the time elapsed up to the $m^{th}$ temperature measurement and the time elapsed between the $m^{th}$ and $(m+1)^{st}$ temperature measurements are conditionally independent given the temperature at the time of the $m^{th}$ measurement, we have $\pi_{\tilde{X}_{m+1}} = \pi_{\tilde{X}_m} * f_{\theta_{\tilde{u}_m}|W_0,W_f}\left(t \mid w_m, w_{m+1}\right)$. Here, $f_{\theta_{\tilde{u}_m}|W_0,W_f}\left(t \mid w_m, w_{m+1}\right)$ is the conditional distribution of the actual time elapsed over $\tilde{u}_m$ clock ticks, given the beginning and ending temperatures, $\tilde{w}_m$ and $\tilde{w}_{m+1}$, respectively. The difficulty lies in computing the distribution $f_{\theta_{\tilde{u}_m}|W_0,W_f}\left(t \mid w_m, w_{m+1}\right)$. At time $\bar{N}$, a similar difficulty arises in computing $f_{\sigma_{\bar{l}}|W_0}\left(t \mid \tilde{w}_{\bar{N}}\right)$, which is needed to compute a distribution on $Z_{\bar{l}}$.

## 3.3 Problem Formulation - Discrete Time

In Section 3.1, we argued that this problem is not immediately amenable to a discrete underlying time scale, because the clock ticks may not coincide with discrete time steps. However, by imposing constraints on the temperature process and possible durations of each clock cycle, it is possible to model the underlying time scale as discrete.

### 3.3.1 Toy Example

We start with another simple toy example where the temperature process is always in one of two states, $\hat{w}^1$ or $\hat{w}^2$. We assume the temperature can only change at integer multiples of two seconds; i.e., $2, 4, 6, \ldots$ seconds, and take the temperature process at these times to be a discrete-time homogeneous Markov process with transition matrix:

$$P = \begin{bmatrix} 0.9 & 0.1 \\ 0.3 & 0.7 \end{bmatrix}.$$

When the temperature is $\hat{w}^1$, the clock ticks once every second, and when the temperature is $\hat{w}^2$, the clock ticks once every two seconds. One possible sample path of the temperature process is shown in Figure 3.5. By repeating the calculations from Figure 3.2, we can determine the actual times of the clock ticks for this sample path. We make two observations about the timing of the clock ticks resulting from this temperature sample path that are actually true for all temperature sample paths: (i) there is a clock tick at every integer multiple of two seconds; and (ii) all clock ticks fall on discrete time steps of one second.

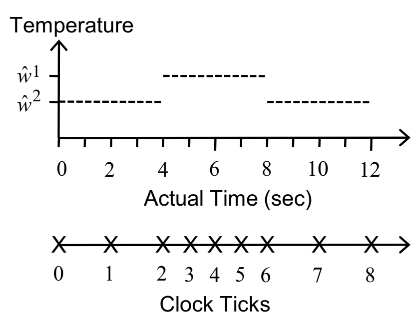


Figure 3.5. Example of a temperature process and frequency mapping satisfying a set of assumptions guaranteeing the clock ticks coincide with discrete time steps. The temperature only changes at integer multiples of two seconds. There are two possible temperatures, $\hat{w}^1$ and $\hat{w}^2$, and their associated frequencies, $f(\hat{w}^1)$ and $f(\hat{w}^2)$ are, respectively, 1 cycle per second and 1 cycle per 2 seconds. Note that the clocks ticks always occur at discrete time steps of one second.

### 3.3.2 Discrete Time Problem Formulation

By imposing additional assumptions on the temperature process and possible frequencies, we can generalize this example to ensure that the clock ticks occur at desired discrete time units. Assume that the underlying discrete time unit is $\Delta$, and that the transitions of the temperature process occur on a slower scale, say at $q\Delta, 2q\Delta, 3q\Delta$, and so forth, for some positive integer $q$. We model the ambient temperature process at these times, $\{\hat{W}_t\}_{t=0,q\Delta,2q\Delta,\ldots}$, as a discrete-time homogeneous Markov process with the same finite state space $\mathcal{W}$ as described for the continuous-time Markov process in Section 3.2, known initial temperature $\hat{w}_0$, and known matrix of transition probabilities $\hat{\mathbf{P}}$, where $\hat{\mathbf{P}} := \{\hat{p}_{ij}\}_{i,j\in\mathcal{W}}$ and $\hat{p}_{ij} := \Pr\left(W_{t+q\cdot\Delta} = j \mid W_t = i\right)$, for all $t$. The mapping $f : \mathcal{W} \to \hat{\mathcal{W}}$ describing the frequency associated with each temperature is the same as the continuous time problem. Assume also that for every $\hat{w} \in \mathcal{W}$, $q \cdot f(\hat{w}) \in \mathbb{Z}_+$ and $\frac{1}{f(\hat{w})} \in \mathbb{Z}_+$. Then, there is a clock tick every time the temperature changes, and every clock tick falls exactly on some integer multiple of $\Delta$.

With these assumptions in place, we formulate a new partially observed Markov decision process, which we refer to as Problem (POMDP-2). All components of Problem (POMDP-2) are the same as Problem (POMDP-1), except the state space is now $\hat{\mathcal{S}} := \mathcal{X} \times \mathcal{W} \times \mathcal{N}$, where $\mathcal{X} := \{0, 1, \ldots, \hat{x}_{\max}\}$. Here, $\hat{x}_{\max}$ is the maximum amount of actual time that could elapse in $\bar{K}$ clock ticks (i.e., if the temperature for the entire horizon was that temperature with the lowest associated frequency).

The state transition law is given by:

$$\Pr\left(\hat{X}_{k+1} = \hat{x}_{k+1}, \hat{W}_{k+1} = \hat{w}_{k+1}, \hat{N}_{k+1} = \hat{n}_{k+1} \middle| \hat{X}_k = \hat{x}_k, \hat{W}_k = \hat{w}_k, \hat{N}_k = \hat{n}_k, \hat{U}_k = \hat{u}_k\right)$$
$$= \Pr\left(\hat{X}_{k+1} = \hat{x}_{k+1}, \hat{W}_{k+1} = \hat{w}_{k+1} \middle| \hat{X}_k = \hat{x}_k, \hat{W}_k = \hat{w}_k\right) \cdot \mathbb{1}_{\left\{\hat{n}_{k+1} = \hat{n}_k + \hat{u}_k^2\right\}}.$$

Problem (POMDP-2) can be reduced in the same manner as Problem (POMDP-

1), resulting in a completely observed MDP whose time index is the number of measurements that have been taken. In fact, the resulting dynamic program has essentially the same form as the dynamic program for Problem (MDP-2). We remark on a few subtle differences. First, the resulting state, $\left(\pi_{\tilde{X}_m}, \tilde{w}_m\right)$, now comprises the most recent temperature and a *probability mass function (pmf)* on the finite space $\mathcal{X}$, rather than a *pdf* on $I\!R_+$. Additionally, recall that much of the computational difficulty of the continuous-time dynamic program arose from updating the conditional distribution of elapsed time after the $m^{th}$ temperature measurement, $\pi_{\tilde{X}_m}$, to the corresponding distribution after the $(m+1)^{st}$ temperature measurement. With an underlying discrete time scale, this calculation is considerably simpler, as there are a finite number of sample paths of the temperature process starting from a known temperature and ending in another known temperature that could result in a given number of clock ticks. One additional consideration for this discrete time case, however, is that the probabilities of these sample paths of the temperature process between the $m^{th}$ and $(m+1)^{st}$ measurements are not conditionally independent of the elapsed time up until the $m^{th}$ measurement, $\tilde{X}_m$, given the beginning temperature $\tilde{w}_m$, as was the case in Problem (MDP-2). This is due to the fixed times at which the temperature can change in Problem (POMDP-2). Nonetheless, one can still update the conditional distribution to $\pi_{\tilde{X}_{m+1}}$ by conditioning on $\tilde{X}_m$, and the net result is a considerably simpler computation.

In summary, we end up with another finite state, finite action, finite time MDP that can be solved through standard dynamic programming. However, computing the solution to this dynamic program is computationally simpler than computing the solution to the dynamic program resulting from the continuous time problem. The tradeoff is that the designer may need to make approximations at the modeling level in order to satisfy the additional assumptions on the temperature process and frequency range.

## 3.4 Computation of Optimal Control Policies for a Toy Example

In this section, we continue the toy example from Section 3.3.1 to show the benefit from scheduling temperature measurements. Let the temperature process and frequency mapping be as described in Section 3.3.1. The scheduler's objective is to time 12 seconds before waking up. The scheduler knows the initial temperature is $\hat{w}^2$. We use the $L_1$ distortion function, so the cost is the absolute value of the difference between 12 seconds and $\hat{X}_\tau$, the actual time at which the controller declares 12 seconds have elapsed.

If the temperature were $\hat{w}^1$ for the entire time horizon, then 12 local clock ticks would correspond to 12 seconds of real time; if the temperature were $\hat{w}^2$ for the entire time horizon, then 6 local clock ticks would correspond to 12 seconds of real time; and, if the temperature were to move between $\hat{w}^1$ and $\hat{w}^2$, then 12 seconds of real time would occur somewhere between the $6^{th}$ and $12^{th}$ clock tick. Moreover, the initial temperature is $\hat{w}^2$, so the first clock tick does not happen until 2 seconds of real time. Thus, *a priori*, the scheduler knows to declare 12 seconds have elapsed somewhere between the $6^{th}$ and $11^{th}$ clock tick.

For three different instances of the problem, with a limit of 0, 1, and 2 temperature measurements, respectively, we computed the optimal policy numerically. The three optimal policies and resulting expected distortions are shown in Figure 3.6. We observe from this toy example how the temperature measurements are used in combination with the local clock ticks to more accurately estimate elapsed real time.

Each additional measurement improves the controller's calibration of the local clock, thereby reducing the expected distortion.

***Open-loop (no measurements)***

- Optimal to declare 12 seconds have elapsed after 9 clock ticks
- Resulting expected distortion is **1.85**

***One measurement allowed***

- Optimal to take the measurement at the $4^{\text{th}}$ clock tick
- If measurement is $w^1$, wait 6 more ticks before declaring (at $10^{\text{th}}$ tick)
- If measurement is $w^2$, wait 2 more ticks before declaring (at $6^{\text{th}}$ tick)
- Resulting expected distortion is **1.00**

***Two measurements allowed***

- Optimal to take the $1^{\text{st}}$ measurement at the $2^{\text{nd}}$ clock tick
- If $1^{\text{st}}$ measurement is $w^1$, wait 4 more ticks before $2^{\text{nd}}$ measurement
  - If $2^{\text{nd}}$ measurement is $w^1$, wait 4 more ticks before declaring (at $10^{\text{th}}$ tick)
  - If $2^{\text{nd}}$ measurement is $w^2$, wait 2 more ticks before declaring (at $8^{\text{th}}$ tick)
- If $1^{\text{st}}$ measurement is $w^2$, wait 2 more ticks before $2^{\text{nd}}$ measurement
  - If $2^{\text{nd}}$ measurement is $w^1$, wait 4 more ticks before declaring (at $8^{\text{th}}$ tick)
  - If $2^{\text{nd}}$ measurement is $w^2$, wait 2 more ticks before declaring (at $6^{\text{th}}$ tick)
- Resulting expected distortion is **0.57**

Figure 3.6. Optimal policies and resulting expected distortions of three different instances of the toy example.

## 3.5 Summary

We considered the problem of dynamically scheduling a limited number of temperature measurements in a manner most useful to improving the accuracy of an ultra-low power clock. We formulated two different optimization problems, with continuous and discrete underlying time scales, respectively. We reduced both problems to finite state, finite horizon, finite action Markov decision processes that can be solved numerically through standard dynamic programming. Modeling the underlying time as discrete is advantageous in terms of computational complexity, but requires extra conditions on the temperature process and frequency range.

# *Chapter 4*

# Introduction to Opportunistic Scheduling

In this chapter, we introduce opportunistic scheduling problems, where the common theme is exploiting the temporal and spatial variation of the wireless channel. In Section 4.1, we highlight the basic ideas of opportunistic scheduling through some motivating examples. We then provide a literature survey and more detailed introduction to the most common modeling issues in opportunistic scheduling problems in Section 4.2. Throughout, we consider a single source transmitting data to one or more users over a wireless channel.

## 4.1 Motivating Examples

**Example 4.1.** Consider a channel that can be in one of $M$ channel conditions, with probabilities $p^1, p^2, \ldots, p^M$, respectively. Associated with each channel condition is a known convex, increasing, differentiable power-rate function, $f_1(z), f_2(z), \ldots, f_M(z)$, respectively, describing the power required to transmit $z$ packets in a discrete time slot. The objective is to minimize the average power consumed over an infinite horizon, subject to a minimum average rate constraint, $\bar{R}$. This problem reduces to the following convex optimization problem:

$$\begin{array}{cl} \min\limits_{(z^1,z^2,\ldots,z^M)\in \mathbb{R}_+^M} & \sum_{i=1}^{M} p^i \cdot f_i\left(z^i\right) \\ \text{s.t.} & \sum_{i=1}^{M} p^i \cdot z^i \geq \bar{R} \ , \end{array} \tag{4.1}$$

where $z^i$ represents the number of packets transmitted when the channel is in condition $i$. The solution to (4.1) is found by reducing (in the same manner as [23, Example 5.2, p. 245]) the Karush-Kuhn-Tucker (KKT) conditions to:

$$z^{i^*} \geq 0,\ z^{i^*} \cdot p^i \cdot \left(\nu^* + f_i'(z^{i^*})\right) = 0,\ \text{and } \nu^* + f_i'(z^{i^*}) \geq 0,\ \forall i \in \{1,2,\ldots,M\}\ ,$$
$$\text{and } \ \mathbf{p}^{\mathrm{T}}\mathbf{z}^* = \bar{R}\ ,$$

where $\nu^*$ is the Lagrange multiplier associated with the rate constraint. Graphically, the so-called "inverse water-filling" solution is found by fixing the slope of a tangent line, and setting the number of packets to be transmitted under condition $i$ to be a $z^i$ such that $f_i'(z^i)$ is equal to the slope, or zero if $f_i'(z^i)$ is greater than the slope for all $z^i \geq 0$. This process is continuously repeated as the slope of the tangent line is gradually increased until $\sum_{i=1}^{M} p^i \cdot z^i = \bar{R}$. The resulting optimal solution $\mathbf{z}^*$ has the property that for every channel $i$, the optimum number of packets $z^{i^*}$ is either equal to zero or satisfies $f_i'(z^i) = -\nu^*$, where $-\nu^*$ is the slope of the final tangent line. See Figure 4.1 for a diagram of this solution.

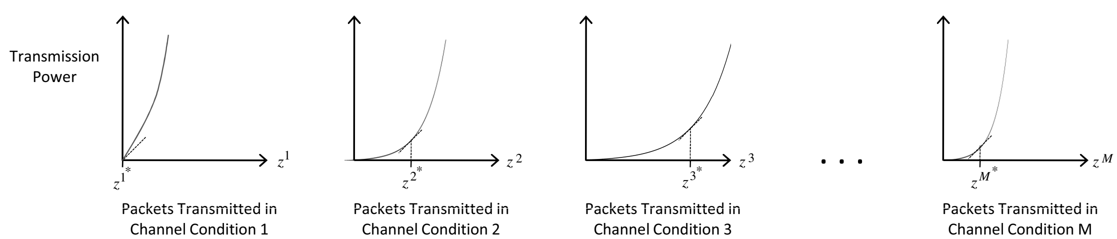


Figure 4.1. Pictorial representation of the solution to Example 4.1. The vector $\mathbf{z}^*$ of the optimal number of packets to transmit under each channel condition has the property that $f_i'(z^{i^*})$ is the same for all channel conditions $i$ such that $z^{i^*} > 0$.

**Example 4.2.** Next, we consider the same infinite horizon average cost problem as (4.1), with the additional stipulations that (i) the power-rate function in each channel condition is linear, with slope $\phi^i$; and (ii) there is a power constraint $P$ in each slot. In other words,

$$f_i\left(z^i\right) = \begin{cases} \phi^i \cdot z^i, & \text{if } \; z^i \leq \frac{P}{\phi^i} \\ \infty, & \text{if } \; z^i > \frac{P}{\phi^i} \end{cases} .$$

We assume without loss of generality that $\phi^1 \leq \phi^2 \leq \ldots \leq \phi^M$ (i.e., $\phi^1$ is the slope of the power-rate function under the best channel condition and $\phi^M$ is slope under the worst condition). With these assumptions, the problem becomes:

$$\begin{array}{lll} \displaystyle\min_{(z^1,z^2,\ldots,z^M)\in \mathbb{R}_+^M} & \sum_{i=1}^M p^i \cdot \phi^i \cdot z^i & \\ \text{s.t.} & \sum_{i=1}^M p^i \cdot z^i \geq \bar{R} & , \\ \text{and} & z^i \leq \frac{P}{\phi^i} \;,\; \forall i \in \{1,2,\ldots,M\} & \end{array} \tag{4.2}$$

where $z^i$ represents the number of packets transmitted when the channel is in condition $i$. The solution to (4.2) is found by defining:

$$j^* := \min\left\{ j \in \{1,2,\ldots,M\} : \sum_{m=1}^{j} p^m \cdot \frac{P}{\phi^m} \geq \bar{R} \right\} .$$

Then the optimal amount of data to send under each channel condition is given by:

$$z^{m^*} := \begin{cases} \frac{P}{\phi^m}, & \text{if } \; m < j^* \\ \frac{\bar{R} - \sum_{m=1}^{j^*-1} p^m \cdot \frac{P}{\phi^m}}{p^{j^*}}, & \text{if } \; m = j^* \\ 0, & \text{if } \; m > j^* \end{cases} . \tag{4.3}$$

See Figure 4.2 for a diagram of this solution.

Examples 4.1 and 4.2 illustrate the main idea of exploiting the temporal variation of the channel via opportunistic scheduling. Namely, we can reduce energy consumption by sending more data when the channel is in a “good” state, and less data when the channel is in a “bad” state. Much of the challenge for the scheduler lies in determining how good or bad a channel condition is, and how much data to send accordingly.

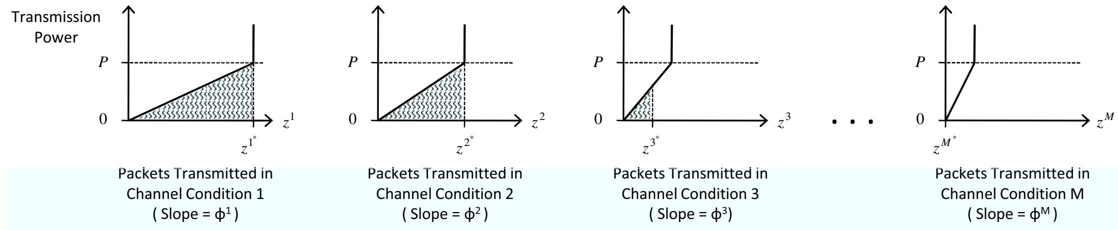


Figure 4.2. Pictorial representation of the solution to Example 4.2. Each plot represents the power-rate curve under a different channel condition. The full power available is used for transmission when the channel is in its best condition(s), and no packets are transmitted when the channel is in its worst condition(s).

In Examples 4.1 and 4.2, the sender is transmitting packets to a single receiver, but it is often the case in wireless communication networks that a single source sends data to multiple users over a shared channel. Such a *downlink* system model is shown in Figure 4.3. In this situation, the scheduler can exploit both the temporal variation

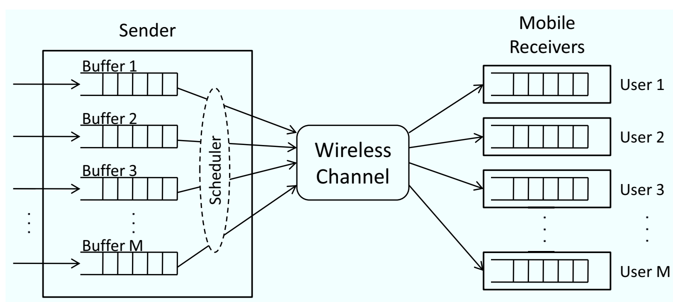


Figure 4.3. Multiuser downlink system model. A single source transmits data to multiple users over a shared wireless channel.

*and the spatial variation* of the channel by sending data to the receivers with the best conditions in each time slot. The benefit of increasing system throughput and reducing total energy consumption through such a joint resource allocation policy is commonly referred to as the *multiuser diversity gain* (see, e.g., [160, Ch. 6]). It was introduced in the context of the analogous *uplink* problem where multiple sources transmit to a single destination (e.g., the base station)[88].

## 4.2 Modeling Issues and Literature Review

There is a wide range of literature on opportunistic scheduling problems in wireless communications. This section is by no means intended to be an exhaustive survey of problems that have been examined, but rather an introduction to some of the most common modeling issues. For more complete surveys of opportunistic scheduling studies in wireless networks, see [98] and [99].

### 4.2.1 Wireless Channel

Modeling the wireless channel deserves an entire book in its own right. For a good introduction to the topic, see [160]. Here, we restrict attention to modeling the wireless channel at the simplest level required for opportunistic scheduling problems, without considering any specific modulation or coding schemes. In this context, the condition of the time-varying wireless channel is usually modeled either as (i) independently and identically distributed (IID) over time; or (ii) a discrete-time Markov process. In the case of multiple receivers, as shown in Figure 4.3, the channels between the sender and each receiver may or may not be correlated in each time slot. For a detailed introduction to modeling fading channels as Markov processes, see [126].

In general, the transmitter can reliably send data across the channel at a higher rate by increasing transmission power. For each possible channel condition, there is a corresponding power-rate curve that describes how much power is required to transmit at a given rate. In the low signal-to-noise ratio (SNR) regime, this power-rate curve is commonly taken to be linear and strictly increasing. In the high SNR regime, the power-rate curve is commonly taken to be convex and strictly increasing [160, Section 5.2]. For a justification of the convex assumption, see [162]. Specific convex power-rate curves that have been considered in the literature include: (i) $c(z,s) = \frac{2^z-1}{\alpha_1(s)}$ (see, e.g., [93]), motivated by the capacity of a discrete-time additive white Gaussian noise (AWGN) channel; and (ii) $c(z,s) = \frac{z^\mu}{\alpha_2(s)}$ (see, e.g., [92]), where in both cases, $c(z,s)$ is the power required to transmit at rate $z$ under channel condition $s$, $\mu$ is a fixed parameter, and the $\alpha_i$'s are parameters that may depend on the channel condition.

### 4.2.2 Channel State Information

In this chapter, we assume that, through a feedback channel, the transmission scheduler learns perfectly (and for free) the state of the channel between the sender and each receiver at the beginning of every time slot. Thus, its scheduling decisions are based on all past and current states of the channel(s), but none of the future channel realizations. This set of assumptions is commonly referred to as *causal* or *full* channel state information. Some papers such as [34] and [161] also refer to problems resulting from this assumption on the scheduler's information as *online scheduling* problems, to differentiate from *offline scheduling* problems, where the scheduler learns all future channel realizations at the beginning of the time horizon.

For a recent survey of research on systems with limited feedback, which may cause the channel state information to be outdated or suffering from errors, see [101]. References [53], [57], [127], [152], and [172] also discuss ways to deal with restrictions on the timing and amount of feedback in an opportunistic scheduling context.

A second relaxation of the perfect channel state information assumption is to force the scheduler to decide whether or not to attain channel state information at some cost, which represents the time and energy consumed in learning the channel state. The process of learning the channel is often referred to as *probing*, and [28], [29],

[30], [61], [62], [80], [113], and [125] are all examples of studies that examine the best joint strategies for probing and transmission in different contexts.

### 4.2.3 Data

The simplest and often most tractable way to model the data is that the sender has an infinite backlog of data to send to each receiver. Analysis under this assumption gives a bound on the maximum achievable performance of a system in terms of throughput. Alternatively, one can assume data arrives to the sender's buffer over time, and explicitly model the arrival process. The arrival process may be deterministic (often the case in *offline scheduling* problems, where the scheduler is assumed to learn the times of all future arrivals at the beginning of the horizon), an IID sequence of random variables (as in [36]), a Poisson process (as in [13]), a discrete-time Markov process (as in [6]) or just about any other stochastic process appropriate for a given application. With an arriving packet model, the scheduler's control policies often depend on both the current queue length of packets backlogged at the sender and the statistics of future arrivals. It may also be the case, as in [13], that the sender's buffer to store arriving packets is of finite length. If so, the scheduler must take care to avoid having to drop packets due to buffer overflow. Finally, the opportunistic scheduling literature is divided on the treatment of a "packet." Some studies take the data to be some integer number of packets that cannot be split, while others consider a *fluid* packet model that allows packets to be split, with the receiver reassembling fractional packets.

### 4.2.4 Performance Objectives

Broadly speaking, opportunistic scheduling problems in wireless networks focus on the tradeoffs between energy-efficiency, throughput, and delay. With some exceptions (e.g., [18]), delay is usually modeled as a QoS constraint (a maximum acceptable level of delay), rather than a quantity to be directly minimized. In many opportunistic scheduling problems, delay is not even considered, with the justification that some applications are not delay-sensitive. We discuss delay further in the next section.

Thus, two most basic setups are (i) to maximize throughput, subject to a constraint on the maximum average or total energy expended; and (ii) to minimize energy consumption, subject to a constraint on the minimum average or total throughput (as in Examples 4.1 and 4.2). These two problems are dual to each other, and many similar techniques can therefore be used to solve both problems. Examples of studies that solve both problems for a similar setup and relate their solutions are [52] and [92].

### 4.2.5 Resource and Quality of Service Constraints

Sending more data when the channel is in a good state can increase system throughput and/or reduce total energy consumption; however, in opportunistic scheduling problems, it is often the case that the transmission scheduler has competing resource and quality of service (QoS) interests. In this section, we provide a brief introduction to some common resource and QoS constraints.

#### Transmission Power

Due to hardware and/or regulatory constraints, a limit is often placed on the sender's transmission power in each slot. Some models allow the sender to transmit to multiple users in a slot, with the total transmission power not exceeding a limit, while others only allow the sender to transmit data to a single user in each slot. This power constraint is often left out of problems where the power-rate curve is strictly convex, as the increasing marginal power required to increase the transmission rate prevents the scheduler from wanting to increase transmission power too much. However, the absence of a power constraint in a problem with a linear power-rate curve would often result in the scheduler wanting to increase transmission power well beyond a reasonable limit in order to send a large amount of data when the channel condition is very good (see, e.g., [52]).

### Delay

Delay is an important QoS constraint in many applications. Different notions of delay have been incorporated into opportunistic scheduling problems. One proxy for delay is the *stability* of all of the sender's queues for arriving packets awaiting transmission. The motivation for this criterion is that if none of these queues blows up, then the delay is not "too bad." With stability as an objective, it is common to restrict attention to *throughput optimal* policies, which are scheduling policies that ensure the sender's queues are stable, as long as this is possible for the given arrival process and channel model. References [6], [114], [134], and [156] present such throughput optimal scheduling algorithms, and examine conditions guaranteeing stabilizability in different settings.

When an arriving packet model is used for the data, then one can also define *end-to-end delay* as the time between a packet's arrival at the sender's buffer and its decoding by the receiver. A number of opportunistic scheduling studies have considered the *average* end-to-end delay of all packets over a long horizon. For instance, [2], [18], [19], [21], [36], [41], [58], [59], [87], [86], [120], and [165] all consider average delay, either as a constraint or by incorporating it directly into the objective function to be minimized. However, the average delay criterion allows for the possibility of long delays (albeit with small probability); thus, for many delay-sensitive applications, *strict* end-to-end delay is often a more appropriate consideration for studies with arriving packet models. In [33] and [34], Chen, Mitra, and Neely place strict constraints on the end-to-end delay of each packet in a point-to-point system, examine the optimal scheduling policy assuming all future channel conditions are known, and suggest heuristics based on this optimal *offline* scheduling policy for the more realistic *online* case where the scheduler only learns the channel conditions in a causal fashion. Rajan, Sabharwal, and Aazhang also consider strict constraints on the end-to-end delay in an arriving packet model in [120, Section IV].

A strict constraint on the end-to-end delay of each packet is one particular form of a *deadline* constraint, as each arriving packet has a deadline by which it must be transmitted (which happens to be a fixed number of slots after its arrival). This notion can be generalized to impose individual deadlines on each packet, whether the packets are arriving over time or are all in the sender's buffer from the beginning, as with the case of infinite backlog. Studies that impose such individual packet deadlines include [34] and [144].

In [51], [52], [90], [91], [92], [93], [155], and [161] the individual deadlines coincide, so that all the packets must be received by some common deadline (usually the end of the time horizon under consideration). We examine further the role of these deadline constraints in the next two chapters.

**Fairness**

If, in the multiuser setting shown in Figure 4.3, the scheduler only considers total throughput and energy consumption across all users, it may often be the case that it ends up transmitting to only a single user or to the same small group of users in every slot. This can happen, for instance, if a base station requires less power to send data to a nearby receiver, even when the nearby receiver's channel is in its worst possible condition and a farther away receiver's channel is in its best possible condition. Thus, *fairness* constraints are often imposed to ensure that the transmitter sends packets to all receivers.

A number of different fairness conditions have been examined in the literature. For example, [17] and [97] consider *temporal fairness*, where the scheduler must transmit to each receiver for some minimum fraction of the time over the long run. Under the *proportional fairness* considered by [74] and [164], the scheduler considers the current channel conditions relative to the average channel condition of each receiver. Reference [97] considers a more general *utilitarian fairness*, where the focus is on system performance from the receiver's perspective, rather than on resources consumed by each user. The authors of [22] incorporate fairness directly into the objective function by setting relative throughput target values for each receiver and maximizing the minimum relative long-run average throughput.

# *Chapter 5*

## Energy-Efficient Transmission Scheduling with Strict Underflow Constraints

In this chapter, we examine the problem of energy-efficient transmission scheduling over a wireless channel, subject to underflow constraints. We consider a single source transmitting to one or more receivers/users over a shared wireless channel. Each user has a buffer to store received packets before they are drained at a certain rate. The available data rate of the channel varies with time and from user to user, due to random fading. The transmitter's goal is to minimize total power consumption by exploiting the temporal and spatial variation of the channel, while preventing any user's buffer from emptying.

This problem falls into the general class of opportunistic scheduling problems discussed in Chapter 4. In our model, the strict underflow constraints serve as a notion of both fairness and delay. The notion of fairness is that none of the receivers' buffers are allowed to empty, guaranteeing the required level of service to all users. The underflow constraints also serve as a notion of delay, and can be seen as multiple deadline constraints - certain packets must arrive by the end of the first slot, another group by the end of the second slot, and so forth.

Sections 5.3 and 5.4 of this chapter generalize the works of [51, 52] and [90, 91, 92, 93], respectively, by considering multiple deadlines in the point-to-point communication problem, rather than a single deadline at the end of the horizon. In addition to better representing some delay-sensitive applications, this extension of the model also allows us to consider infinite horizon problems. We discuss these related works in more detail in Chapter 6.

## 5.1 Wireless Media Streaming and Related Work

The primary application we have in mind to motivate this problem is wireless media streaming. For this application, the data are audio/video sequences, and the packets are drained from the receivers' buffers in order to be decoded and played. Enforcing the underflow constraints reduces playout interruptions to the end users. In order to make the presentation concrete, we use the above wireless media streaming terminology throughout the chapter.

Transporting multimedia over wireless networks is a promising application that has seen recent advances [54]. At the same time, a number of resource allocation issues need to be addressed in order to provide high quality and efficient media over wireless. First, streaming is in general bandwidth-demanding. Second, streaming applications tend to have stringent QoS requirements (e.g., they can be delay and jitter intolerant). Third, it is desirable to operate the wireless system in an energy-efficient manner. This is obvious when the source of the media streaming (the sender) is a mobile. When the media comes from a base station that is not power-constrained, it is still desirable to conserve power in order to (i) limit potential interference to other base stations and their associated mobiles, and (ii) maximize the number of receivers the sender can support.

Of the related work in wireless media streaming, [94] has the closest setup to our model. The main differences are that [94] features a loose constraint on underflow (i.e., it is allowed, but at a cost), as opposed to our tight constraint, and the two studies adopt different wireless channel models. In the extension [95], the receiver may slow down its playout rate (at some cost) to avoid underflow. In this setting, the authors investigate the tradeoffs between power consumption and playout quality, and examine joint power/playout rate control policies. In our model, the receiver does not have the option to adjust the playout speeds. Our model also bears resemblance to [102]. The first difference here is that [102] aims to minimize transmission energy subject to a constant end-to-end delay constraint on each video frame. A second difference is that the controller in [102] must assign various source coding parameters such as quantization step size and coding mode, whereas our model assumes a fixed encoding/decoding scheme.

The remainder of this chapter is organized as follows. In the next section, we describe the system model, formulate finite and infinite horizon MDPs, and relate our model to models in inventory theory. In Section 5.3, we consider the case of a single receiver under linear power-rate curves. While this case can be considered a special case of the models of Sections 5.4 and 5.5, we present it first in order to (i) state additional structural properties of the optimal transmission policy to a single user under linear power-rate curves that are not true in general for the cases discussed in Sections 5.4 and 5.5; (ii) highlight some intuitive takeaways that carry over to the generalized models, but are more transparent in the simpler model; and (iii) compare it to related problems in the wireless communications literature. We analyze the structure of the optimal scheduling policy for the finite horizon problem and provide a method to compute the critical numbers that complete the characterization of the optimal policy when some additional technical conditions are met. Section 5.4 generalizes the analysis of Section 5.3 to the case of a single receiver under piecewise-linear convex power-rate curves, and also addresses the infinite horizon problems for the case of a single receiver. In Section 5.5, we analyze the structure of the optimal policy when there are two receivers with linear power-rate curves. We discuss the relaxation of the strict underflow constraints in Section 5.6 and summarize the main results of the chapter in Section 5.7.

## 5.2 Problem Description

In this section, we present an abstraction of the transmission scheduling problem outlined in the previous section and formulate three optimization problems. While most of this chapter focuses on the cases of one and two users, the formulation in this section is for the more general multi-user (multi-receiver) case, so that we can discuss this more general case in Section 8.1.3.

### 5.2.1 System Model and Assumptions

We consider a single source transmitting media sequences to $M$ users/receivers over a shared wireless channel. The sender maintains a separate buffer for each receiver, and is assumed to always have data to transmit to each receiver.[1] We consider a *fluid* packet model that allows packet to be split, with the receiver reassembling fractional packets. Each receiver has a playout buffer at the receiving end, assumed to be infinite. While in reality this cannot be the case, it is nevertheless a reasonable assumption considering the decreasing cost and size of memory, and the fact that our system model allows holding costs to be assessed on packets in the receiver buffers. See Figure 5.1 for a diagram of the system.

We consider time evolution in discrete steps, indexed backwards by $n = N, N-1, \ldots, 1$, with $n$ representing the number of slots remaining in the time horizon. $N$

[1] This assumption is commonly referred to as the infinite backlog assumption.

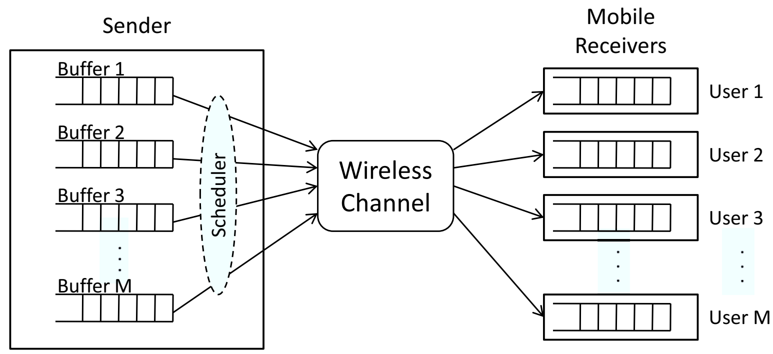

Figure 5.1. System model for Problems (**P5.1**), (**P5.2**), and (**P5.3**).

is the length of the time horizon, and slot $n$ refers to the time interval $[n, n-1)$.

At the beginning of each time slot, the scheduler allocates some amount of power (possibly zero) for transmission to each user. The total power allocated in any one slot must not exceed the fixed power constraint, $P$. Following transmission and reception in each slot, a certain number of packets are removed/purged from each receiver buffer for playing. The transmitter (or scheduler) knows precisely the packet requirements of each receiver (i.e., the number of packets removed from the buffer) in each time slot. This is justified by the assumption that the transmitter knows the encoding and decoding schemes used. We assume that packets transmitted in slot $n$ arrive in time to be used for playing in slot $n$, and that the users' consumption of packets in each slot is constant, denoted by $\mathbf{d} = \left(d^1, d^2, \ldots, d^M\right)$. This latter assumption is less realistic, but may be justified if the receiving buffers are drained at a constant rate at the MAC layer, before packets are decoded by the media players at the application layer. It is also worth noting that the same techniques we use in this chapter to analyze the constant drainage rate case can be used to examine the case of time-varying drainage rates. We discuss the extension to the case of time-varying drainage rates further in Section 5.3.1. We also assume the receiver buffers are empty at the beginning of the time horizon, and that even when the channels are in their worst possible condition, the maximum power constraint $P$ is sufficient to transmit enough packets to satisfy one time slot's packet requirements for every user. We discuss the relaxation of this assumption in Section 5.6.

In general, wireless channel conditions are time-varying. Adopting a block fading model, we assume that the slot duration is within the channel coherence time such that the channel conditions within a single slot are constant. User $m$'s channel condition in slot $n$ is modeled as a random variable, $S_n^m$. We assume that the evolution of a given user's channel condition is independent of all other users' channel conditions and the transmitter's scheduling decisions. We also assume that the transmitter learns all the channel states through a feedback channel at the beginning of each time slot, prior to making the scheduling decisions.

We begin by modeling the evolution of each user's channel condition as a finite-state ergodic homogeneous Markov process, $\{S_n^m\}_{n=N,N-1,\ldots,1}$ with state space $\mathcal{S}^m$.[2] Namely, conditioned on the channel state, $S_n^m$, at time $n$, user $m$'s channel states at *future* times $(n-1, n-2, \ldots)$ are independent of the channel states at *past* times $(n+1, n+2, \ldots)$. Note the somewhat unconventional notation that future times are indexed by lower epoch numbers, as $n$ represents the number of slots remaining in the time horizon. Modeling time backwards facilitates the analysis of the infinite horizon problems, as will be seen for example in Section 5.4.3. It may also be the case that each user's channel condition is independent and identically distributed (IID) from slot to slot. When this is the case, we can often say more about the optimal transmission policy, as will be seen for example in Sections 5.3.2 and 5.4.2.

Associated with each channel condition for a given user is a power-rate function. If user $m$'s channel is in condition $s^m$, then the transmission of $r$ units of data to

[2]Theorems 5.1, 5.3, 5.5, 5.8, 5.9, and 5.10 and their proofs remain valid as stated when each user's channel condition is given by a more general homogeneous Markov process that is not necessarily finite-state and ergodic.

user $m$ incurs a power consumption of $c^m(r, s^m)$. This power-rate function $c^m(\cdot, s^m)$ is commonly assumed to be linear (in the low SNR regime) or convex (in the high SNR regime). In this chapter, we consider power-rate functions that are linear or piecewise-linear convex, the latter of which can be used to approximate more general convex power-rate functions. We assume that sending data consumes a strictly positive amount of power, and therefore take the power-rate functions to be strictly increasing under all channel conditions.

The goal of this study is to characterize the control laws that minimize the transmission power and packet holding costs over a finite or infinite time horizon, subject to tight underflow constraints and a maximum power constraint in each time slot.

### 5.2.2 Problem Formulation

We consider three problems. Problem (**P5.1**) is the finite horizon discounted expected cost problem; Problem (**P5.2**) is the infinite horizon discounted expected cost problem; and Problem (**P5.3**) is the infinite horizon average expected cost problem. The three problems feature the same information state, action space, system dynamics, and cost structure, but different optimization criteria.

The information state at time $n$ is the pair $(\mathbf{X}_n, \mathbf{S}_n)$, where the random vector $\mathbf{X}_n = (X_n^1, X_n^2, \cdots, X_n^M)$ denotes the current receiver buffer queue lengths, and $\mathbf{S}_n = (S_n^1, S_n^2, \cdots, S_n^M)$ denotes the channel conditions in slot $n$ (recall that $n$ is the number of steps remaining until the end of the horizon). The dynamics for the receivers' queues are governed by the simple equation $\mathbf{X}_{n-1} = \mathbf{X}_n + \mathbf{Z}_n - \mathbf{d}$ at all times $n = N, N-1, \ldots, 1$, where $\mathbf{Z}_n$ is a controlled random vector chosen by the scheduler at each time $n$ that represents the number of packets transmitted to each user in the $n^{th}$ slot. At each time $n$, $\mathbf{Z}_n$ must be chosen to meet the peak power constraint:

$$\sum_{m=1}^{M} c^m(Z_n^m, S_n^m) \leq P \ ,$$

and the underflow constraints:

$$X_n^m + Z_n^m \geq d^m \ , \ \forall m \in \{1, 2, \ldots, M\} \ .$$

Clearly, the scheduler cannot transmit a negative number of packets to any user, so it must also be true that $Z_n^m \geq 0$ for all $m$.

We now present the optimization criterion for each problem. In addition to the cost associated with power consumption from transmission, we introduce holding costs on packets stored in each user's playout buffer at the end of a time slot. The holding costs associated with user $m$ in each slot are described by a convex, non-negative, nondecreasing function, $h^m(\cdot)$, of the packets remaining in user $m$'s buffer following playout, with $\lim_{x\to\infty} h^m(x) = \infty$. We assume without loss of generality that $h^m(0) = 0$. Possible holding cost models include a linear model, $h^m(x) = \hat{h}^m \cdot x$ for some positive constant $\hat{h}^m$, or a barrier-type function such as:

$$h^m(x) \ := \ \begin{cases} 0, & \text{if } \ x \leq \mu \\ \kappa \cdot (x - \mu), & \text{if } \ x > \mu \ (\kappa \text{ very large}) \end{cases} \ ,$$

which could represent a finite receiver buffer of length $\mu$.[3]

In Problem (**P5.1**), we wish to find a transmission policy $\boldsymbol{\pi}$ that minimizes $J_{N,\alpha}^{\pi}$, the finite horizon discounted expected cost under policy $\boldsymbol{\pi}$, defined as:

$$J_{N,\alpha}^{\pi} := I\!\!E^{\pi}\left\{\sum_{n=1}^{N}\sum_{m=1}^{M} \alpha^{N-n} \cdot \left\{c^m\big(Z_n^m, S_n^m\big) + h^m\big(X_n^m + Z_n^m - d^m\big)\right\} \mid \mathcal{F}_N\right\} \ ,$$

where $0 \leq \alpha \leq 1$ is the discount factor and $\mathcal{F}_N$ denotes all information available at the beginning of the time horizon. For Problem (**P5.2**), the discount factor must

[3]Taking $\mu$ to be greater than the time horizon $N$ in the finite horizon expected cost problem is equivalent to not assessing any holding costs in Problem (**P5.1**).

satisfy $0 \leq \alpha < 1$, and the infinite horizon discounted expected cost function for minimization is defined as:

$$J^{\pi}_{\infty,\alpha} := \lim_{N\to\infty} J^{\pi}_{N,\alpha}\ ,$$

For Problem (**P5.3**), the average expected cost function for minimization is defined as:

$$J^{\pi}_{\infty,1} := \limsup_{N\to\infty} \frac{1}{N} J^{\pi}_{N,1}\ .$$

In all three cases, we allow the transmission policy $\boldsymbol{\pi}$ to be chosen from the set of all history-dependent randomized and deterministic control laws, $\boldsymbol{\Pi}$ (see, e.g., [68, Definition 2.2.3, pg. 15]).

Combining the constraints and criteria, we present the optimization formulations for Problem (**P5.1**) (or (**P5.2**) or (**P5.3**)):

$$\begin{aligned} & \inf_{\pi\in\Pi} J^{\pi}_{N,\alpha} \left(\text{or } \inf_{\pi\in\Pi} J^{\pi}_{\infty,\alpha} \text{ or } \inf_{\pi\in\Pi} J^{\pi}_{\infty,1}\right) \\ \text{s.t.} \quad & \sum_{m=1}^{M} c^m\left(Z^m_n, S^m_n\right) \leq P,\ w.p.1,\ \forall n \\ & Z^m_n \geq \max\left\{0, d^m - X^m_n\right\},\ w.p.1,\ \forall n,\ \forall m \in \{1,2,\ldots,M\}. \end{aligned}$$

Problem (**P5.1**) may be solved using standard dynamic programming (see, e.g., [68, 20]). The recursive dynamic programming equations are given by:[4]

$$\begin{aligned} V_n(\mathbf{x},\mathbf{s}) &= \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \sum\limits_{m=1}^{M} \left\{c^m\left(z^m, s^m\right) + h^m\left(x^m + z^m - d^m\right)\right\} \\ +\alpha \cdot I\!E\left[V_{n-1}(\mathbf{x}+\mathbf{z}-\mathbf{d},\mathbf{S}_{n-1}) \,\middle|\, \mathbf{S}_n = \mathbf{s}\right] \end{array} \right\} \\ & \quad n = N, N-1, \ldots, 1 \qquad (5.1) \\ V_0(\mathbf{x},\mathbf{s}) &= 0,\ \forall \mathbf{x} \in I\!R^M_+, \forall \mathbf{s} \in \mathcal{S} := \mathcal{S}^1 \times \mathcal{S}^2 \times \ldots \times \mathcal{S}^M, \end{aligned}$$

[4]As will be shown in the proofs of Theorems 5.6 and 5.11, our model satisfies the measurable selection condition 3.3.3 of [68, pg. 28], justifying the use of min rather than inf in the dynamic programming equations.

where $V(\cdot,\cdot)$ is the value function or expected cost-to-go, and the action space is defined as:

$$\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s}) := \left\{ \mathbf{z} \in I\!\!R_{+}^{M} : \begin{array}{l} \mathbf{z} \succeq \max\left\{\mathbf{0}, \mathbf{d}-\mathbf{x}\right\} \text{ and } \\ \sum\limits_{m=1}^{M} c^m\left(z^m, s^m\right) \leq P \end{array} \right\}, \; \forall \mathbf{x} \in I\!\!R_{+}^{M}, \forall \mathbf{s} \in \mathcal{S}, \qquad (5.2)$$

where the maximum in (5.2) is taken element-by-element (i.e., $z^m \geq \max\left\{0, d^m - z^m\right\} \;\; \forall m$). Note that our assumption that the maximum power constraint $P$ is always sufficient to transmit enough packets to satisfy one time slot's packet requirements for every user (i.e., $\sum_{m=1}^{M} c^m\left(d^m, s^m\right) \leq P, \; \forall \mathbf{s} \in \mathcal{S}$) ensures that the action space $\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ is always non-empty.

### 5.2.3 Relation to Inventory Theory

The model outlined in Section 5.2.1 corresponds closely to models used in inventory theory. Borrowing that field's terminology, our abstraction is a multi-period, single-echelon, multi-item, discrete-time inventory model with random (linear or piecewise-linear convex) ordering costs, a budget constraint, and deterministic demands. The items correspond to the streams of data packets, the random ordering costs to the random channel conditions, the budget constraint to the power available in each time slot, and the deterministic demands to the packet requirements for playout.

To the best of our knowledge, this particular problem has not been studied in the context of inventory theory, but similar problems have been examined, and some of the techniques from the inventory theory literature are useful in analyzing our model. References [46], [55], [56], [81], [84], [85], [104], and [103] all consider single-item discrete-time inventory models with linear ordering costs and random prices. The key result for the case of deterministic demand of a single item with no resource constraint is that the optimal policy is a base-stock policy with different target stock levels for each price. Specifically, for each possible ordering price (translates into channel condition in our context), there exists a critical number such that the optimal policy is to fill the inventory (receiver buffer) up to that critical number if the current level is lower than the critical number, and not to order (transmit) anything if the current level is above the critical number.[5] Of the prior work, Kingsman [84, 85] is the only author to consider a resource constraint, and he imposes a maximum on the number of items that may be ordered in each slot. The resource constraint we consider is of a different nature in that we limit the amount of power available in each slot. This is equivalent to a limit on the per slot budget (regardless of the stochastic price realization), rather than a limit on the number of items that can be ordered.

Of the related work on single-item discrete-time inventory models with deterministic linear ordering costs and stochastic demand, [48] and [157] are the most relevant; in those studies, however, the resource constraint also amounts to a limit on the number of items that can be ordered in each slot, and is constant over time. References [16], [150], and [171] consider single-item inventory models with deterministic piecewise-linear convex ordering costs and stochastic demand. The key result in this setup is that the optimal inventory level after ordering is a piecewise-linear nondecreasing function of the current inventory level (i.e., there are a finite number of target stock levels), and the optimal ordering quantity is a piecewise-linear nonincreasing function of the current inventory level. Porteus [116] refers to policies of this form as *finite generalized base-stock policies*, to distinguish them from the superclass of *generalized base-stock policies*, which are optimal when the deterministic ordering

[5] We use the terms target level and critical number interchangeably throughout the thesis.

costs are convex (but not necessarily piecewise-linear), as first studied in [82]. Under a generalized base-stock policy, the optimal inventory level after ordering is a nondecreasing function of the current inventory level, and the optimal ordering quantity is a nonincreasing function of the current inventory level.

References [32], [40], [45], and [79] consider multi-item discrete-time inventory systems under deterministic ordering costs, stochastic demand, and resource constraints. We discuss related results from these studies in more detail in Chapter 7.

In the continuous-time inventory literature, [39], [63], [115], [163], [166], and [173] consider scheduling of the so-called *multiclass make-to-stock queue*, where a single facility produces multiple items subject to stochastic demands. While the production times are also random in these models, some of the structural properties and qualitative features bear a close resemblance to the discrete-time inventory models with deterministic ordering costs and stochastic demand discussed in Chapter 7. For more detailed reviews of the continuous-time models, see the introductions of [32] and [79].

We are not aware of any prior work on (i) single-item inventory models with random piecewise-linear convex ordering costs; (ii) exact computation of the critical numbers in any sort of finite generalized base-stock policy; or (iii) multi-item inventory models with random ordering costs and joint resource constraints. Therefore, not only is this connection between wireless transmission scheduling problems and inventory models novel, but the results we present in this chapter also represent a contribution to the inventory theory literature.

## 5.3 Single Receiver with Linear Power-Rate Curves

In this section, we analyze the finite horizon discounted expected cost problem when there is only a single receiver ($M = 1$), and the power-rate functions under different channel conditions are linear. One such family of power-rate functions is shown in Figure 5.2, where there are three possible channel conditions, and a different linear power-rate function associated with each channel condition. Note that due to the power constraint $P$ in each slot, the effective power-rate function is a two-segment piecewise-linear convex function under all channel conditions. We subsequently simplify our notation and use $c_s$ to denote the power consumption per unit of data transmitted when the channel condition is in state $s$. Because there is just a single receiver, we also drop the dependence of the functions and random variables on $m$. We defer the infinite horizon expected cost problems for this case until Section 5.4.3.

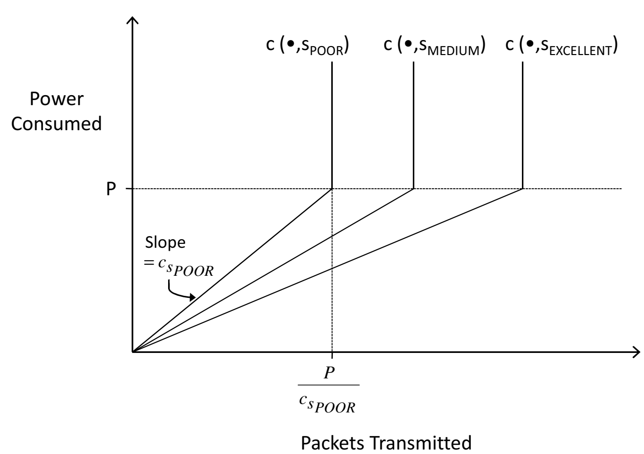


Figure 5.2. A family of linear power-rate functions. Due to the power constraint, the effective power-rate function, shown above for each of the three channel conditions, is a two-segment piecewise-linear convex function. When the channel condition is $s$, the slope of the first segment is $c_s$.

We denote the "best" and "worst" channel conditions by $s_{\text{best}}$ and $s_{\text{worst}}$, respectively, and denote the slopes of the power-rate functions under these respective conditions by $c_{\min}$ and $c_{\max}$. That is,

$$0 < c_{s_{\text{best}}} = c_{\min} := \min_{s\in\mathcal{S}}\{c_s\} \le \max_{s\in\mathcal{S}}\{c_s\} =: c_{\max} = c_{s_{\text{worst}}} \le \frac{P}{d}.$$

With these notations in place, the dynamic program (5.1) for Problem (**P5.1**) becomes:

$$\begin{aligned}
V_n(x,s) &= \min_{\max(0,d-x)\le z\le \frac{P}{c_s}} \left\{ \begin{array}{l} c_s\cdot z + h(x+z-d) \\ +\alpha\cdot \mathbb{E}\big[V_{n-1}(x+z-d,S_{n-1}) \bigm| S_n = s\big] \end{array} \right\} && (5.3)\\
&= \min_{\max(x,d)\le y\le x+\frac{P}{c_s}} \left\{ \begin{array}{l} c_s\cdot (y-x) + h(y-d) \\ +\alpha\cdot \mathbb{E}\big[V_{n-1}(y-d,S_{n-1}) \bigm| S_n = s\big] \end{array} \right\} && (5.4)\\
&= -c_s\cdot x + \min_{\max(x,d)\le y\le x+\frac{P}{c_s}} \Big\{g_n(y,s)\Big\},\ n = N, N-1,\ldots,1\ , \\
V_0(x,s) &= 0,\ \forall x\in \mathbb{R}_+, \forall s\in\mathcal{S},
\end{aligned}$$

where $g_n(y,s) := c_s\cdot y + h(y-d) + \alpha\cdot\mathbb{E}\big[V_{n-1}(y-d,S_{n-1}) \mid S_n = s\big]$. Here, the transition from (5.3) to (5.4) is done by a change of variable in the action space from $Z_n$ to $Y_n$, where $Y_n = X_n + Z_n$. The controlled random variable $Y_n$ represents the queue length of the receiver buffer *after* transmission takes place in the $n^{th}$ slot, but *before* playout takes place (i.e., before $d$ packets are removed from the buffer). The restrictions on the action space, $\max(x,d) \le y \le x + \frac{P}{c_s}$, ensure: (i) a nonnegative number of packets is transmitted; (ii) there are at least $d$ packets in the receiver buffer following transmission, in order to satisfy the underflow constraint; and (iii) the power constraint is satisfied.

### 5.3.1 Structure of Optimal Policy

With the above change of variable in the the action space, the expected cost-to-go at time $n$, $V_n(x,s)$, depends on the current buffer level, $x$, only through the fixed term $-c_s\cdot x$ and the action space; i.e., the function $g_n$ does not depend on

$x$. This separation allows us to leverage the inventory theory techniques of showing "single critical number" or "base-stock" policies, which date as far back as [15]. The following theorem gives the structure of the optimal transmission policy for the finite horizon discounted expected cost problem.

**Theorem 5.1.** *For every $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$, define the critical number*

$$b_n(s) := \min\left\{\hat{y} \in [d, \infty) : \; g_n(\hat{y}, s) = \min_{y \in [d,\infty)} g_n(y, s)\right\} \; .$$

*Then, for Problem (**P5.1**) in the case of a single receiver with linear power-rate curves, the optimal buffer level after transmission with $n$ slots remaining is given by:*

$$y_n^*(x, s) := \begin{cases} x, & \text{if } \; x \geq b_n(s) \\ b_n(s), & \text{if } \; b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ x + \frac{P}{c_s}, & \text{if } \; x < b_n(s) - \frac{P}{c_s} \end{cases} \; , \tag{5.5}$$

*or, equivalently, the optimal number of packets to transmit in slot $n$ is given by:*

$$z_n^*(x, s) := \begin{cases} 0, & \text{if } \; x \geq b_n(s) \\ b_n(s) - x, & \text{if } \; b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ \frac{P}{c_s}, & \text{if } \; x < b_n(s) - \frac{P}{c_s} \end{cases} \; . \tag{5.6}$$

*Furthermore, for a fixed $s$, $b_n(s)$ is nondecreasing in $n$:*

$$N \cdot d \geq b_N(s) \geq b_{N-1}(s) \geq \ldots \geq b_1(s) = d \; . \tag{5.7}$$

*If, in addition, the channel condition is independent and identically distributed from slot to slot, then for a fixed $n$, $b_n(s)$ is nonincreasing in $c_s$; i.e., for arbitrary $s^1, s^2 \in \mathcal{S}$ with $c_{s^1} \leq c_{s^2}$, we have:*

$$n \cdot d \geq b_n(s_{\text{best}}) \geq b_n(s^1) \geq b_n(s^2) \geq b_n(s_{\text{worst}}) = d \; . \tag{5.8}$$

The optimal transmission policy in Theorem 5.1 is a modified base-stock policy. At time $n$, for each possible channel condition realization $s$, the critical number $b_n(s)$ describes the target number of packets to have in the user's buffer after transmission in the $n^{th}$ slot. If that number of packets is already in the buffer, then it is optimal to not transmit any packets; if there are fewer than the target and the available power is enough to transmit the difference, then it is optimal to do so; and if there are fewer than the target and the available power is not enough to transmit the difference, then the sender should use the maximum power to transmit. See Figure 5.3 for diagrams of the optimal policy.

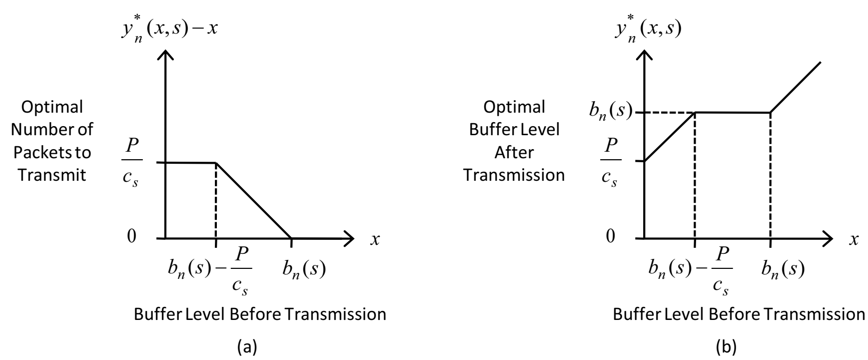


Figure 5.3. Structure of optimal policy for Problem (**P5.1**) in the case of a single receiver with linear power-rate curves. When the state is $(x, s)$ in slot $n$, (a) depicts the optimal transmission quantity, and (b) depicts the resulting number of packets available for playout in slot $n$.

Details of the proof of Theorem 5.1 are included in Appendix A.1. The key realization is that for all $n$ and all $s$, $g_n(\cdot, s) : [d, \infty) \rightarrow I\!R_+$ is a convex function in $y$, with $\lim_{y\rightarrow\infty} g_n(y, s) = \infty$. Thus, for all $n$ and all $s$, $g_n(\cdot, s)$ has a global minimum $b_n(s)$, the target number of packets to have in the buffer following transmission in the $n^{th}$ slot. The key idea to show (5.7) is to fix $s \in \mathcal{S}$, view $g_n(y, s)$ as a function of $y$ and $n$, say $f(y, n)$, and show that the function $f(\cdot, \cdot)$ is submodular. From the proof, one can also see that if we relax the stationary (time-invariant) deterministic demand assumption to a nonstationary (time-varying) deterministic demand sequence, $\{d_N, d_{N-1}, \ldots, d_1\}$ (with $d_n \leq \frac{P}{c_{\max}}$ for all $n$), then the structure of the optimal policy is still as stated in (5.5). If the channel is IID, then the following statement, analogous to (5.8), is true for arbitrary $s^1, s^2 \in \mathcal{S}$ with $c_{s^1} \leq c_{s^2}$:

$$\sum_{i=1}^{n} d_i \geq b_n(s_{\text{best}}) \geq b_n(c_{s^1}) \geq b_n(c_{s^2}) \geq b_n(s_{\text{worst}}) = d_n \ , \ \forall n \in \{1, 2, \ldots, N\} \ . \quad (5.9)$$

However, (5.7), the monotonicity of critical numbers over time for a fixed channel condition, is not true in general under nonstationary deterministic demand. As one counterexample, (5.9) says that under an IID channel, the critical numbers for the worst possible channel condition are equal to the single period demands. Therefore, if the demand sequence is not monotonic, the sequence of critical numbers, $\{b_n\left(s_{\text{worst}}\right)\}_{n=1,2,\ldots,N}$, is not monotonic.

### 5.3.2 Computation of the Critical Numbers

In this section, we consider the special case where the channel condition is independent and identically distributed from slot to slot, the holding cost function is linear (i.e., $h(x) = h \cdot x$ for some $h \geq 0$), and the following technical condition is satisfied: for each possible channel condition $s$,

$$\frac{P}{c_s} = l \cdot d \text{ for some } l \in I\!N; \quad (5.10)$$

i.e., the maximum number of packets that can be transmitted in any slot covers exactly the playout requirements of some integer number of slots. Under these three assumptions, we can completely characterize the optimal transmission policy.

**Theorem 5.2.** *Define the threshold $\gamma_{n,j}$ for $n \in \{1, 2, \ldots, N\}$ and $j \in I\!N$ recursively, as follows:*

**(i)** *If $j = 1$, $\gamma_{n,j} = \infty$;*

**(ii)** *If* $j > n$, $\gamma_{n,j} = 0$;

**(iii)** *If* $2 \leq j \leq n$,

$$\gamma_{n,j} = -h + \alpha \cdot \left( \begin{array}{l} \displaystyle\sum_{s:\ c_s \geq \gamma_{n-1,j-1}} p(s) \cdot \gamma_{n-1,j-1} + \sum_{s:\ c_s < \gamma_{n-1,j-1}} p(s) \cdot c_s \\ \displaystyle + \sum_{s:\ c_s < \gamma_{n-1,j-1+L(s)}} p(s) \cdot \left[\gamma_{n-1,j-1+L(s)} - c_s\right] \end{array} \right), \quad (5.11)$$

*where* $p(s)$ *is the probability of the channel being in state* $s$ *in a time slot, and* $L(s) := \frac{P}{d \cdot c_s}$. *For each* $n \in \{1, 2, \ldots, N\}$ *and* $s \in \mathcal{S}$, *if* $\gamma_{n,j+1} \leq c_s < \gamma_{n,j}$, *define* $b_n(s) := j \cdot d$. *The optimal control strategy for Problem (**P5.1**) is then given by* $\boldsymbol{\pi}^* = \left\{y_N^*, y_{N-1}^*, \ldots, y_1^*\right\}$, *where*

$$y_n^*(x, s) := \begin{cases} x, & \text{if } x \geq b_n(s) \\ b_n(s), & \text{if } b_n(s) - \frac{P}{c_s} \leq x < b_n(s) \\ x + \frac{P}{c_s}, & \text{if } x < b_n(s) - \frac{P}{c_s} \end{cases}. \quad (5.12)$$

Note that with $n$ slots remaining, $0 = \gamma_{n,n+1} \leq \gamma_{n,n} \leq \gamma_{n,n-1} \leq \ \ldots \ \leq \gamma_{n,2} \leq \gamma_{n,1} = \infty$, so $b_n(s)$ is well-defined.

Compared to using standard numerical techniques to *approximately* solve the dynamic program and find a near-optimal policy, the above result not only sheds more insight on the structural properties of the problem and its *exactly*-optimal solution, but also offers a computationally simpler method. In particular, the optimal policy is completely characterized by the thresholds $\{\gamma_{n,j}\}_{n \in \{1,2,\ldots,N\},\ j \in \mathbb{N}}$. Calculating these thresholds recursively, as described in Theorem 5.2, requires $O(N^2 |\mathcal{S}|)$ operations, which is considerably simpler from a computational standpoint than approximately solving the dynamic program [35, 124].

To prove Theorem 5.2, we show by backwards induction that it is worse to transmit either fewer or more packets than the number suggested by the policy $\boldsymbol{\pi}^*$. The detailed proof is omitted, as Theorem 5.2 is a special case of Theorem 5.4; however, we discuss some intuition behind the proof and the thresholds here.

The reason for the technical condition regarding the maximum number of packets that can be transmitted in any slot is as follows. The optimal action at all times (in general, without the technical condition) is either to transmit enough packets to fill the buffer up to a level satisfying the playout requirements of some number of future slots, or to transmit at maximum power. When the technical condition is satisfied, transmitting at maximum power also results in filling the buffer up to a level satisfying the playout requirements of some number of future slots. Thus, under the optimal policy, all realizations result in the buffer level at the end of every time slot being some integer multiple of the demand, $d$. This fact makes it easier to compute the thresholds $\{\gamma_{n,j}\}_{n \in \{1,2,\ldots,N\},\ j \in \mathbb{N}}$.

An intuitive explanation of the recursion (5.11) is as follows. The threshold $\gamma_{n,j}$ may be interpreted as the per packet power cost at which, with $n$ slots remaining in the horizon, the expected cost-to-go of transmitting packets to cover the user's playout requirements for the next $j-1$ slots is the same as the expected cost-to-go of transmitting packets to cover the user's requirements for the next $j$ slots. That is, $\gamma_{n,j}$ should satisfy:

$$\alpha \cdot \mathbb{E}\left[V_{n-1}\Big((j-1)\cdot d, S_{n-1}\Big)\right] + \gamma_{n,j}\cdot d + h \cdot d = \alpha \cdot \mathbb{E}\left[V_{n-1}\Big((j-2)\cdot d, S_{n-1}\Big)\right],$$

which is equivalent to:

$$\begin{aligned}
&\gamma_{n,j} \\
&= -h + \frac{\alpha}{d}\cdot \mathbb{E}\left[V_{n-1}\Big((j-2)\cdot d, S_{n-1}\Big) - V_{n-1}\Big((j-1)\cdot d, S_{n-1}\Big)\right] \qquad (5.13)\\
&= -h + \frac{\alpha}{d}\cdot \sum_{s\in\mathcal{S}} p(s)\cdot \left[V_{n-1}\Big((j-2)\cdot d, s\Big) - V_{n-1}\Big((j-1)\cdot d, s\Big)\right]
\end{aligned}$$

$$= -h + \frac{\alpha}{d} \cdot \left[ \begin{array}{l} \displaystyle\sum_{s:\ b_{n-1}(s) \leq (j-2)\cdot d} p(s) \cdot \left\{ -h \cdot d + \alpha \mathbb{E} \left[ \begin{array}{l} V_{n-2}\Big((j-3)\cdot d, S_{n-2}\Big) \\ -V_{n-2}\Big((j-2)\cdot d, S_{n-2}\Big) \end{array} \right] \right\} \\ + \displaystyle\sum_{s:\ (j-2)\cdot d < b_{n-1}(s) \leq \big(j-2+L(s)\big)\cdot d} p(s) \cdot c_s \cdot d \\ + \displaystyle\sum_{\substack{s:\ b_{n-1}(s) \\ > \big(j-2+L(s)\big)\cdot d}} p(s) \cdot \left\{ \begin{array}{l} -h \cdot d \\ +\alpha \mathbb{E} \left[ \begin{array}{l} V_{n-2}\Big(\big(j-3+L(s)\big)\cdot d, S_{n-2}\Big) \\ -V_{n-2}\Big(\big(j-2+L(s)\big)\cdot d, S_{n-2}\Big) \end{array} \right] \end{array} \right\} \end{array} \right] \tag{5.14}$$

$$= -h + \alpha \cdot \left\{ \begin{array}{l} \displaystyle\sum_{s:\ b_{n-1}(s) \leq (j-2)\cdot d} p(s) \cdot \gamma_{n-1,j-1} \\ + \displaystyle\sum_{s:\ (j-2)\cdot d < b_{n-1}(s) \leq \big(j-2+L(s)\big)\cdot d} p(s) \cdot c_s \\ + \displaystyle\sum_{s:\ b_{n-1}(s) > \big(j-2+L(s)\big)\cdot d} p(s) \cdot \gamma_{n-1,j-1+L(s)} \end{array} \right\} \tag{5.15}$$

$$= -h + \alpha \cdot \left\{ \begin{array}{l} \displaystyle\sum_{s:\ c_s \geq \gamma_{n-1,j-1}} p(s) \cdot \gamma_{n-1,j-1} \\ + \displaystyle\sum_{s:\ \gamma_{n-1,j-1+L(s)} \leq c_s < \gamma_{n-1,j-1}} p(s) \cdot c_s \\ + \displaystyle\sum_{s:\ c_s < \gamma_{n-1,j-1+L(s)}} p(s) \cdot \gamma_{n-1,j-1+L(s)} \end{array} \right\}. \tag{5.16}$$

Here, (5.14) follows from the structure of the optimal control action (5.5). If the channel condition $s$ in the $(n-1)^{st}$ slot is such that $b_{n-1}(s) \leq (j-2)\cdot d$, then no packets are transmitted when the starting buffer level is either $(j-2)\cdot d$ or $(j-1)\cdot d$, and the respective buffer levels at the beginning of slot $n-2$ are $(j-3)\cdot d$ and $(j-2)\cdot d$. The instantaneous costs resulting from the two starting buffer levels differ by $-h\cdot d$. When $(j-2)\cdot d < b_{n-1}(s) \leq \big(j-2+L(s)\big)\cdot d$, the power constraint is not tight starting from $(j-1)\cdot d$, so the buffer level after transmission is the same starting from $(j-2)\cdot d$ or $(j-1)\cdot d$. The instantaneous costs resulting from the two starting buffer levels differ by $c_s \cdot d$, as an extra $d$ packets are transmitted if the starting buffer is $(j-2) \cdot d$. Finally, when $b_{n-1}(s) > \big(j-2+L(s)\big) \cdot d$, the power constraint is tight starting from both $(j-2) \cdot d$ and $(j-1) \cdot d$. Therefore, the instantaneous cost difference is $-h \cdot d$, and the respective buffer levels at the beginning of slot $n-2$ are $(j-3+L(s)) \cdot d$ and $(j-2+L(s)) \cdot d$. Equation (5.15) follows from (5.13), with $n-1, j-1$ substituted for $n, j$, and (5.16) follows from the definition that $b_n(s) = j \cdot d$ if $\gamma_{n,j+1} \leq c_s < \gamma_{n,j}$.

Comparing the threshold $\gamma_{n,j}$ defined in (5.11) to the corresponding threshold in the unrestricted (no power constraint) single user problem [56, 84], the only difference is the third term of the right-hand side of (5.11):

$$\alpha \cdot \sum_{\{s:\ c_s < \gamma_{n-1,j-1+L(s)}\}} p(s) \cdot \left[\gamma_{n-1,j-1+L(s)} - c_s\right],$$

which is absent in the unrestricted case. For all $n \in \{1, 2, \ldots, N\}$ and $j \in I\!N$, this term is nonnegative. Thus, for a fixed $n$ and $j$, the threshold in the restricted case is at least as high as the corresponding threshold in the unrestricted case. It follows that the optimal stock-up level $b_n(s)$ is also at least as high in the restricted case for all $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$. The intuition behind this difference is that the sender should transmit more packets under the same (medium) conditions, because it is not able to take advantage of the best channel conditions to the same extent due to the power constraint.

## 5.4 Single Receiver with Piecewise-Linear Convex Power-Rate Curves

In this section, we analyze Problems (**P5.1**), (**P5.2**), and (**P5.3**) when there is only a single receiver ($M = 1$), and the power-rate functions under different channel conditions are piecewise-linear convex. Note that this is a generalization of the case considered in Section 5.3.

We assume without loss of generality that under each channel condition $s$, the power-rate function has $K+1$ segments, and thus the power consumed in transmitting $z$ packets under channel condition $s$ can be represented as follows:

$$c(z,s) = z \cdot \tilde{c}_0(s) + \sum_{k=0}^{K-1} \Big\{ \big( \tilde{c}_{k+1}(s) - \tilde{c}_k(s) \big) \cdot \max\big\{ z - \tilde{z}_k(s), 0 \big\} \Big\} \ , \text{ where}$$
$$0 < \tilde{c}_0(s) \leq \tilde{c}_1(s) \leq \cdots \leq \tilde{c}_K(s) \ , \text{ and}$$
$$0 = \tilde{z}_{-1}(s) < \tilde{z}_0(s) < \tilde{z}_1(s) < \cdots < \tilde{z}_{K-1}(s) < \tilde{z}_K(s) = \infty \ .$$

The terms $\{\tilde{c}_k(s)\}_{k \in \{0,1,\ldots,K\}}$ represent the slopes of the segments of $c(\cdot, s)$, and the terms $\{\tilde{z}_k(s)\}_{k \in \{0,1,\ldots,K-1\}}$ represent the points at which the slopes of $c(\cdot, s)$ change. An example of a family of such power-rate functions is shown in Figure 5.4. For each channel condition $s \in \mathcal{S}$, we define the maximum number of packets that can be transmitted without exceeding the per slot power constraint $P$ as:

$$\tilde{z}_{\max}(s) := \{z : c(z,s) = P\} \ .$$

Note that $\tilde{z}_{\max}(s)$ is well-defined due to the strictly increasing nature of $c(\cdot, s)$. Recall that we assume $\tilde{z}_{\max}(s) \geq d, \ \forall s \in \mathcal{S}$. We also assume without loss of generality that $\tilde{z}_{\max}(s) > \tilde{z}_{K-1}(s), \ \forall s \in \mathcal{S}$.

In this case, the dynamic program (5.1) for Problem (**P5.1**) becomes:

$$\begin{aligned} V_n(x,s) &= \min_{\left\{\max(0,d-x) \leq z \leq \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ + \alpha \cdot \mathbb{E}\big[ V_{n-1}(x+z-d, S_{n-1}) \bigm| S_n = s \big] \end{array} \right\} \\ &= \min_{\left\{\max(0,d-x) \leq z \leq \tilde{z}_{\max}(s)\right\}} \Big\{ c(z,s) + \tilde{g}_n(x+z,s) \Big\}, \\ & \qquad\qquad n = N, N-1, \ldots, 1 \qquad (5.17) \\ V_0(x,s) &= 0, \ \forall x \in \mathbb{R}_+, \forall s \in \mathcal{S} \ , \end{aligned}$$

where $\tilde{g}_n(y,s) := h(y-d) + \alpha \cdot \mathbb{E}\left[ V_{n-1}(y-d, S_{n-1}) | S_n = s \right]$.

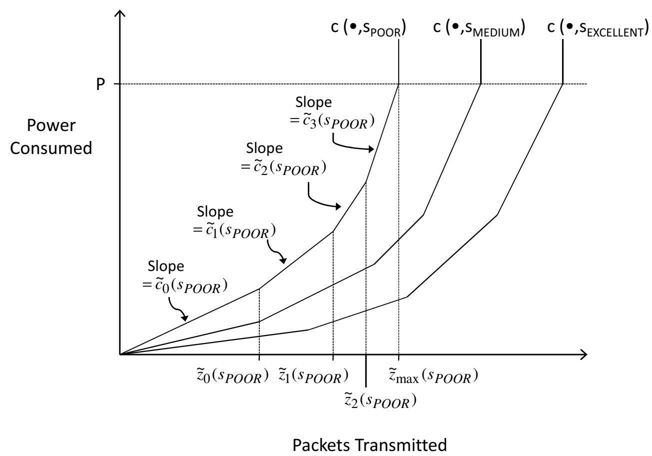


Figure 5.4. A family of piecewise-linear convex power-rate functions. Like Figure 5.2, we incorporate the power constraint into each curve to show the effective power-rate curve. As an example, the power-rate function $c(\cdot, s_{POOR})$ is completely characterized by the sequence of slopes $\{\tilde{c}_k(s_{POOR})\}_{k\in\{0,1,2,3\}}$ and the sequence of points where the slopes change $\{\tilde{z}_k(s_{POOR})\}_{k\in\{0,1,2\}}$. The maximum number of packets that can be transmitted in a slot when the channel condition is $s_{POOR}$ is $\tilde{z}_{\max}(s_{POOR})$.

### 5.4.1 Structure of Optimal Policy for the Finite Horizon Discounted Expected Cost Problem

We showed in Theorem 5.1 that the the optimal transmission policy to a single receiver in the case of linear power-rate curves is a modified base-stock policy characterized by a single critical level for each channel condition. In this section, we generalize this result to the case of piecewise-linear power-rate curves, and show that the optimal receiver buffer level after transmission (respectively, optimal number of packets to transmit) is no longer a three-segment piecewise-linear nondecreasing (respectively, nonincreasing) function of the starting buffer level as in Figure 5.3, but a more general piecewise-linear nondecreasing (respectively, nonincreasing) function.

**Theorem 5.3.** *In Problem (**P5.1**) with a single receiver under piecewise-linear convex power-rate curves, for every $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$, there exists a nonincreasing sequence of critical numbers $\left\{b_{n,k}(s)\right\}_{k\in\{0,1,\ldots,K\}}$ such that the optimal num-*

*ber of packets to transmit with $n$ slots remaining is given by:*

$$z_n^*(x,s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } \; b_{n,k}(s) - \tilde{z}_{k-1}(s) < x \le b_{n,k-1}(s) - \tilde{z}_{k-1}(s) \;, \\ & \quad k \in \{0, 1, \ldots, K\} \\ b_{n,k}(s) - x, & \text{if } \; b_{n,k}(s) - \tilde{z}_{k}(s) < x \le b_{n,k}(s) - \tilde{z}_{k-1}(s) \;, \\ & \quad k \in \{0, 1, \ldots, K-1\} \\ b_{n,K}(s) - x, & \text{if } \; b_{n,K}(s) - \tilde{z}_{\max}(s) < x \le b_{n,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } \; 0 \le x \le b_{n,K}(s) - \tilde{z}_{\max}(s) \end{cases} \tag{5.18}$$

*where $b_{n,-1}(s) := \infty, \ \forall s \in \mathcal{S}$. The optimal receiver buffer level after transmission is given by $y_n^*(x,s) = x + z_n^*(x,s)$.*

The optimal transmission policy in Theorem 5.3 is a finite generalized base-stock policy. It can be interpreted as follows. Under each channel condition $s$, there is a target level or critical number associated with each segment of the associated piecewise-linear convex power-rate curve shown in Figure 5.4. If the starting buffer level is below the critical number associated with the first segment, $b_{n,0}(s)$, the scheduler should try to bring the buffer level as close as possible to the target, $b_{n,0}(s)$. If the maximum number of packets sent at this per packet power cost, $\tilde{z}_0(s)$, does not suffice to reach the critical number $b_{n,0}(s)$, then those $\tilde{z}_0(s)$ packets are scheduled, and the next segment of the power-rate curve is considered. This second segment has a slope of $\tilde{c}_1(s)$ and an associated critical number $b_{n,1}(s)$, which is no higher than $b_{n,0}(s)$, the first critical number. If the starting buffer level plus the $\tilde{z}_0(s)$ already-scheduled packets brings the buffer level above $b_{n,1}(s)$, then no more packets are scheduled for transmission. Otherwise, it is optimal to transmit so as to bring the buffer level as close as possible to $b_{n,1}(s)$, by transmitting up to $\tilde{z}_1(s) - \tilde{z}_0(s)$ additional packets at a cost of $\tilde{c}_1(s)$ power units per packet. This process continues with the sequential consideration of each segment of the power-rate curve. At each successive iteration, the target level is lower and the starting buffer level, updated to include already-scheduled packets, is higher. The process continues until the buffer level reaches or exceeds a critical number, or the full power $P$ is consumed. Note that this sequential consideration is not actually done online, but only meant to provide an intuitive explanation of the optimal policy. See Figure 5.5 for diagrams of the structure of the optimal finite generalized base-stock policy, and Appendix A.2 for a detailed proof of Theorem 5.3.

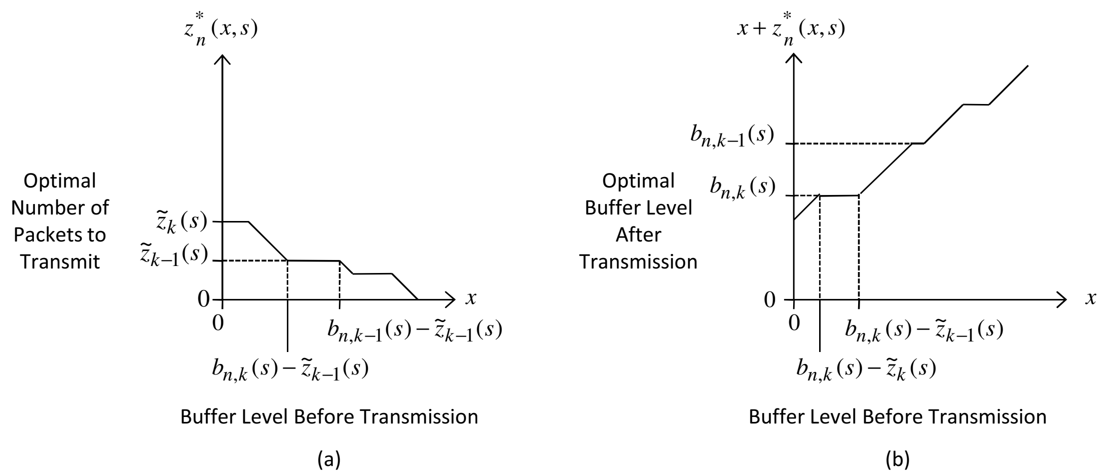


Figure 5.5. Structure of optimal policy for Problem (**P5.1**) in the case of a single receiver with piecewise-linear convex power-rate curves. When the state is $(x,s)$ in slot $n$, (a) depicts the optimal transmission quantity, and (b) depicts the resulting number of packets available for playout in slot $n$.

### 5.4.2 Computation of Critical Numbers

While finite generalized base-stock policies have been considered in the inventory literature for almost three decades, we are not aware of any previous studies that explicitly compute the critical numbers for any model where such a policy is optimal. In this section, we compute the critical numbers under each channel condition when technical conditions similar to those of Section 5.3.2 are satisfied. We consider the special case when the channel condition is independent and identically distributed from slot to slot; the holding cost function is linear (i.e., $h(x) = h \cdot x$); and the following technical condition on the power-rate functions is satisfied for each possible channel condition $s \in \mathcal{S}$: $\tilde{z}_{\max}(s) = \tilde{l}_{\max} \cdot d$ for some $\tilde{l}_{\max} \in \mathbb{N}$, and for every $k \in \{0, 1, \ldots, K-1\}$, $\tilde{z}_k(s) = \tilde{l}_k \cdot d$ for some $\tilde{l}_k \in \mathbb{N}$; i.e., the slopes of the effective power-rate functions only change at integer multiples of the drainage rate $d$. Under these conditions, we can completely characterize the optimal transmission policy.

As in Theorem 5.2, we recursively define a set of thresholds, and use them to determine the critical numbers, $\{b_{n,k}(s)\}_{k \in \{-1,0,\ldots,K\}}$, for each channel condition, at each time.

**Theorem 5.4.** *Define the thresholds $\tilde{\gamma}_{n,j}$ for $n \in \{1, 2, \ldots, N\}$ and $j \in \mathbb{N}$ recursively, as follows:*

**(i)** *If $j = 1$, $\tilde{\gamma}_{n,j} = \infty$;*

**(ii)** *If $j > n$, $\tilde{\gamma}_{n,j} = 0$;*

**(iii)** *If $2 \leq j \leq n$,*

$$\tilde{\gamma}_{n,j} = -h + \alpha \cdot \left( \begin{array}{l} \displaystyle\sum_{s:\ \tilde{c}_0(s) \geq \tilde{\gamma}_{n-1,j-1}} p(s) \cdot \tilde{\gamma}_{n-1,j-1} \\ \displaystyle + \sum_{k=0}^{K-1} \left\{ \begin{array}{l} \displaystyle\sum_{\substack{s:\ \tilde{\gamma}_{n-1,j-1+\tilde{L}_k(s)} \\ \leq \tilde{c}_k(s) \\ < \tilde{\gamma}_{n-1,j-1+\tilde{L}_{k-1}(s)}}} p(s) \cdot \tilde{c}_k(s) \\ \displaystyle + \sum_{\substack{s:\ \tilde{c}_k(s) \\ < \tilde{\gamma}_{n-1,j-1+\tilde{L}_k(s)} \\ \leq \tilde{c}_{k+1}(s)}} p(s) \cdot \tilde{\gamma}_{n-1,j-1+\tilde{L}_k(s)} \end{array} \right\} \\ \displaystyle + \sum_{s:\ \tilde{\gamma}_{n-1,j-1+\tilde{L}_{\max}(s)} \leq \tilde{c}_K(s) < \tilde{\gamma}_{n-1,j-1+\tilde{L}_{K-1}(s)}} p(s) \cdot \tilde{c}_K(s) \\ \displaystyle + \sum_{s:\ \tilde{c}_K(s) < \tilde{\gamma}_{n-1,j-1+\tilde{L}_{\max}(s)}} p(s) \cdot \tilde{\gamma}_{n-1,j-1+\tilde{L}_{\max}(s)} \end{array} \right), \quad (5.19)$$

*where $p(s)$ is the probability of the channel being in state $s$ in a time slot, $\tilde{L}_k(s) := \frac{\tilde{z}_k(s)}{d}$ for all $s \in \mathcal{S}$ and $k \in \{0, 1, \ldots, K-1\}$, and $\tilde{L}_{\max}(s) := \frac{\tilde{z}_{\max}(s)}{d}$ for all $s \in$*

*$\mathcal{S}$. For each $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$, define $b_{n,-1}(s) := \infty$ and for all $k \in \{0, 1, \ldots, K\}$, if $\tilde{\gamma}_{n,j+1} \leq \tilde{c}_k(s) < \tilde{\gamma}_{n,j}$, define $b_{n,k}(s) := j \cdot d$. The optimal control strategy for Problem (**P5.1**) is then given by $\boldsymbol{\pi}^* = \left\{z_N^*, z_{N-1}^*, \ldots, z_1^*\right\}$, where for all $n \in \{N, N-1, \ldots, 1\}$, $z_n^*(x, s)$ is given by* (5.18).

It is straightforward to check that Theorem 5.4 is in fact a generalization of Theorem 5.2. To see this, set $K = 0$ so that the summation from $k = 0$ to $k = K - 1$ on the right-hand side of (5.19) drops out. Then $\tilde{\gamma}_{n,j}$ in (5.19) is the same as $\gamma_{n,j}$ in (5.11), $\tilde{c}_0(s)$ corresponds to $c_s$ in (5.11), $b_{n,0}(s)$ corresponds to $b_n(s)$, $\tilde{z}_{\max}(s)$ corresponds to $\frac{P}{c_s}$, $\tilde{L}_{\max}(s)$ corresponds to $L(s)$, and $\tilde{L}_{K-1}(s) = 0$. The resulting optimal transmission policies are also the same.

In Theorem 5.4, the threshold $\tilde{\gamma}_{n,j}$ may again be interpreted as the per packet power cost at which, with $n$ slots remaining in the horizon, the expected cost-to-go of transmitting packets to cover the user's playout requirements for the next $j - 1$ slots is the same as the expected cost-to-go of transmitting packets to cover the user's requirements for the next $j$ slots. The intuition behind the recursion (5.19) is similar to the detailed explanation given in Section 5.3.2. Namely, we can start with equation (5.13) and expand out the right-hand side based on the known structure of the optimal policy, until, after a fair bit of algebra, the result is (5.19). A detailed proof of Theorem 5.4 is included in Appendix A.3.

### 5.4.3 Structure of the Optimal Policy for the Infinite Horizon Discounted Expected Cost Problems

In this section, we show that the optimal policy for the infinite horizon discounted expected cost problem is the natural extension of the optimal policy for the finite horizon discounted expected cost problem; namely, it is a finite generalized base-stock policy characterized by *time-invariant* sequences of critical numbers for each channel condition. These time-invariant sequences of critical numbers for the infinite horizon discounted expected cost problem are equal to the limit of the finite horizon sequences of critical numbers as the time horizon $N$ goes to infinity.

**Theorem 5.5.**

**(a)** *For a fixed* $x \in I\!R_+$ *and* $s \in \mathcal{S}$, $V_n(x,s)$ *is nondecreasing in* $n$. *Moreoever,* $\lim_{n\to\infty} V_n(x,s)$ *exists and is finite,* $\forall x \in I\!R_+, \forall s \in \mathcal{S}$.

**(b)** *Define* $V_\infty(x,s) := \lim_{n\to\infty} V_n(x,s)$. *Then* $V_\infty(x,s)$ *is convex in* $x$ *for any fixed* $s \in \mathcal{S}$.

**(c)** *Define* $\tilde{g}_\infty(y,s) := h(y-d) + \alpha \cdot I\!E\left[V_\infty(y-d,S') \mid S=s\right]$, *where* $S'$ *is the channel condition in the subsequent slot. Then* $\tilde{g}_n(y,s)$ *converges monotonically to* $\tilde{g}_\infty(y,s), \forall y \in [d,\infty), \forall s \in \mathcal{S}$; $\tilde{g}_\infty(y,s)$ *is convex in* $y$ *for any fixed* $s \in \mathcal{S}$; *and* $\lim_{y\to\infty} \tilde{g}_\infty(y,s) = \infty, \forall s \in \mathcal{S}$.

**(d)** *Define* $b_{\infty,-1}(s) := \infty$ *and*

$$b_{\infty,k}(s) := \max\Big\{d, \inf\big\{b \mid \tilde{g}'^{+}_{\infty}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} \ , \ \forall k \in \{0,1,\ldots,K\} \ ,$$

*where* $\tilde{g}'^{+}_{\infty}(b,s)$ *represents the right derivative:*

$$\tilde{g}'^{+}_{\infty}(b,s) := \lim_{y\downarrow b} \frac{\tilde{g}_\infty(y,s) - \tilde{g}_\infty(b,s)}{y-b} \ .$$

*Then* $b_{\infty,k}(s) = \lim_{n\to\infty} b_{n,k}(s)$ *for all* $k \in \{-1,0,1,\ldots,K\}$.

**(e)** $V_\infty(x,s)$ *satisfies the* $\alpha$*-discounted cost optimality equation (*$\alpha$*-DCOE):*

$$\begin{aligned} V_\infty(x,s) &= \min_{\left\{\max(0,d-x)\leq z\leq \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ +\alpha \cdot I\!E\big[V_\infty(x+z-d,S') \bigm| S=s\big] \end{array} \right\} \\ &= \min_{\left\{\max(0,d-x)\leq z\leq \tilde{z}_{\max}(s)\right\}} \Big\{c(z,s) + \tilde{g}_\infty(x+z,s)\Big\}, \forall x, \forall s, \end{aligned} \qquad (5.20)$$

*and the minimum on the right hand side of (5.20) is achieved by:*

$$z^*_\infty(x,s) := \begin{cases} \tilde{z}_{k-1}(s), & \textit{if } \; b_{\infty,k}(s) - \tilde{z}_{k-1}(s) < x \le b_{\infty,k-1}(s) - \tilde{z}_{k-1}(s) \;, \\ & \; k \in \{0,1,\ldots,K\} \\ b_{\infty,k}(s) - x, & \textit{if } \; b_{\infty,k}(s) - \tilde{z}_{k}(s) < x \le b_{\infty,k}(s) - \tilde{z}_{k-1}(s) \;, \\ & \; k \in \{0,1,\ldots,K-1\} \\ b_{\infty,K}(s) - x, & \textit{if } \; b_{\infty,K}(s) - \tilde{z}_{\max}(s) < x \le b_{\infty,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \textit{if } \; 0 \le x \le b_{\infty,K}(s) - \tilde{z}_{\max}(s) \end{cases}$$

**(f)** *The optimal stationary policy for Problem (**P5.2**) in the case of a single receiver with piecewise-linear convex power-rate curves is given by* $\boldsymbol{\pi^*_\infty} = (z^*_\infty, z^*_\infty, \ldots)$.

A detailed proof, which follows the logic conveyed in the statement of the theorem, is included in Appendix B.1. As a special case of Theorem 5.5, the optimal policy in Problem (**P5.2**) for the case discussed in Section 5.3 of a single receiver with linear power-rate curves is given by $\boldsymbol{\pi^*_\infty} = (z^*_\infty, z^*_\infty, \ldots)$, where:

$$z^*_\infty(x,s) := \begin{cases} 0, & \text{if } \; x \ge b_\infty(s) \\ b_\infty(s) - x, & \text{if } \; b_\infty(s) - \frac{P}{c_s} \le x < b_\infty(s) \\ \frac{P}{c_s}, & \text{if } \; x < b_\infty(s) - \frac{P}{c_s} \end{cases} ,$$

and $b_\infty(s) := \lim_{n \to \infty} b_n(s)$.

### 5.4.4 Structure of the Optimal Policy for the Infinite Horizon Average Expected Cost Problems

In this section we use the *vanishing discount approach* to show that the finite generalized base-stock structure is also optimal for the infinite horizon average expected cost problem, **(P5.3)**. We show that an optimal policy for the infinite horizon average expected cost problem exists and can be represented as the limit as the discount factor increases to one of optimal policies identified in Section 5.4.3 for the infinite horizon discounted expected cost problem.

In Section 5.4.3, we suppressed the dependence of the value functions and optimal policies on the discount factor, $\alpha$. Here, we make this dependence explicit by including the discount factor in the subscript labeling of the value functions and optimal policies for the infinite horizon discounted expected cost problem. For example, the value function defined in (b) of Theorem 5.5 is now denoted by $V_{\infty,\alpha}(x,s)$.

**Theorem 5.6.** *For all $\alpha \in [0,1)$, define:*

$$\begin{aligned} m_{\infty,\alpha} &:= \inf_{\substack{x \in \mathbb{R}_+ \\ s \in \mathcal{S}}} V_{\infty,\alpha}(x,s), \\ \rho^* &:= \lim_{\alpha \nearrow 1} (1-\alpha) \cdot m_{\infty,\alpha}, \text{ and} \\ w_{\infty,\alpha}(x,s) &:= V_{\infty,\alpha}(x,s) - m_{\infty,\alpha}, \ \forall x \in \mathbb{R}_+, \ \forall s \in \mathcal{S}. \end{aligned}$$

*Then:*

**(a)** *There exists a continuous function $w_{\infty,1}(\cdot,\cdot)$ and a selector $z^*_{\infty,1}(\cdot,\cdot)$ that satisfy the ACOE:*

$$\begin{aligned} \rho^* + w_{\infty,1}(x,s) &= \min_{\left\{\max(0,d-x) \leq z \leq \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ + \mathbb{E}\left[w_{\infty,1}(x+z-d,S') \mid S=s\right] \end{array} \right\} \\ &= c\Big(z^*_{\infty,1}(x,s),s\Big) + h\Big(x + z^*_{\infty,1}(x,s) - d\Big) \\ &\quad + \mathbb{E}\left[w_{\infty,1}\Big(x + z^*_{\infty,1}(x,s) - d, S'\Big) \Big| S=s\right], \ \forall x \in \mathbb{R}_+, \ \forall s \in \mathcal{S}. \end{aligned}$$

**(b)** *The stationary policy $\boldsymbol{\pi^*_{\infty,1}} = (z^*_{\infty,1}, z^*_{\infty,1}, \ldots)$ is optimal for Problem (**P5.3**) in the case of a single receiver with piecewise-linear convex power-rate curves.*

**(c)** *The resulting optimal average cost beginning from any initial state $(x,s) \in \mathbb{R}_+ \times \mathcal{S}$ is $\rho^*$.*

**(d)** *For every increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\ldots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\ldots}$ approaching 1 such that:*

$$w_{\infty,1}(x,s) \;\; = \lim_{i\to\infty} w_{\infty,\alpha(l_i)}(x,s), \;\; \forall x \in I\!R_+, \;\; \forall s \in \mathcal{S}.$$

*Therefore, for every $s \in \mathcal{S}$, $w_{\infty,1}(x,s)$ is convex in $x$.*

**(e)** *For every $(x,s) \in I\!R_+ \times \mathcal{S}$ and increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\ldots}$ approaching 1, there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\ldots}$ approaching 1 and a sequence $\{x(i)\}_{i=1,2,\ldots}$ approaching $x$ such that:*

$$z^*_{\infty,1}(x,s) \;\; = \lim_{i\to\infty} z^*_{\infty,\alpha(l_i)}(x(i),s) \; .$$

**(f)** *A stationary finite generalized base-stock policy is average cost optimal in the case of piecewise-linear convex power-rate curves, and a stationary modified base-stock policy is average cost optimal in the case of linear power-rate curves.*

Thus, the structure of the optimal policy is the same for all three problems, (**P5.1**), (**P5.2**), and (**P5.3**). The proof of Theorem 5.6 is discussed in Appendix C.

### 5.4.5 General Convex Power-Rate Curves

As mentioned in Section 5.2.1, in general, the power-rate curve under each possible channel condition is convex. It can be shown that under convex power-rate curves at each time, the optimal number of packets to send is a nonincreasing function of the starting buffer level. However, without any further structure on the power-rate curves, it is not computationally tractable to compute such optimal policies, known as generalized base-stock policies (a superclass of the *finite* generalized base-stock policies discussed above). This is why we have chosen to analyze piecewise-linear convex power-rate curves, which can be used to approximate general convex power-rate curves. More specifically, our analysis suggests approximating the general convex power-rate curves by piecewise-linear convex power-rate curves where the slopes change at integer multiples of the demand $d$, in order to be able to apply Theorem 5.4 to compute the critical numbers in an extremely efficient manner. Doing so represents an approximation at the modeling stage followed by an exact solution, as compared to modeling the power-rate curves as more general convex functions and having to approximate the solution. Finally, we note that increasing the number of segments used to model the piecewise-linear convex functions leads to a better approximation, but comes at the cost of some extra complexity in implementing the optimal policy, as the scheduler needs to store at least one critical number for each segment of each power-rate curve.

## 5.5 Two Receivers with Linear Power-Rate Curves

In this section, we analyze the finite and infinite horizon discounted expected cost problems when there are two receivers ($M = 2$), and the power-rate functions under different channel conditions are linear for each user. Each user $m$'s channel condition evolves as a homogeneous Markov process, $\{S_n^m\}_{n=N,N-1,\ldots,1}$. As discussed earlier, the time-varying channel conditions of the two users are independent of each other, and the transmission scheduler can exploit this spatial diversity. Like Section 5.3, we denote the power consumption per unit of data transmitted to receiver $m$ under channel condition $s^m$ by $c_s^m$. The row vector of these per unit power consumptions is given by $\mathbf{c}_\mathbf{s}^{\mathrm{T}}$, so that the total power consumption in slot $n$ is given by $\sum_{m=1}^{2} c^m(Z_n^m, S_n^m) = \mathbf{c}_\mathbf{s}^{\mathrm{T}}\mathbf{Z}_n$. We denote the total holding costs $\sum_{m=1}^{2} h^m(X_n^m + Z_n^m - d^m)$ by $h(\mathbf{X}_n + \mathbf{Z}_n - \mathbf{d})$.

With these notations, the dynamic program (5.1) for Problem (**P5.1**) becomes:

$$\begin{aligned} V_n(\mathbf{x},\mathbf{s}) &= \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}+\mathbf{z}-\mathbf{d}) \\ +\alpha\cdot I\!\!E\big[V_{n-1}(\mathbf{x}+\mathbf{z}-\mathbf{d},\mathbf{S}_{n-1}) \bigm| \mathbf{S}_n=\mathbf{s}\big] \end{array} \right\} && (5.21) \\ &= \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] + h(\mathbf{y}-\mathbf{d}) \\ +\alpha\cdot I\!\!E\big[V_{n-1}(\mathbf{y}-\mathbf{d},\mathbf{S}_{n-1}) \bigm| \mathbf{S}_n=\mathbf{s}\big] \end{array} \right\} && (5.22) \\ &= -\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{x} + \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \Big\{G_n(\mathbf{y},\mathbf{s})\Big\},\ n=N,N-1,\ldots,1\ , \\ V_0(\mathbf{x},\mathbf{s}) &= 0,\ \forall \mathbf{x}\in I\!\!R_+^2, \forall \mathbf{s}\in\mathcal{S}:=\mathcal{S}^1\times\mathcal{S}^2, \end{aligned}$$

where

$$G_n(\mathbf{y},\mathbf{s}) := \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{y} + h(\mathbf{y}-\mathbf{d}) + \alpha\cdot I\!\!E\big[V_{n-1}(\mathbf{y}-\mathbf{d},\mathbf{S}_{n-1}) \bigm| \mathbf{S}_n=\mathbf{s}\big],$$
$$\forall \mathbf{y}\in[d^1,\infty)\times[d^2,\infty), \forall \mathbf{s}\in\mathcal{S},\ \text{and}$$
$$\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s}) := \Big\{\mathbf{y}\in I\!\!R_+^2 : \mathbf{y}\succeq \mathbf{d}\vee\mathbf{x} \text{ and } \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}]\le P\Big\},\ \forall \mathbf{x}\in I\!\!R_+^2, \forall \mathbf{s}\in\mathcal{S}. \quad (5.23)$$

The transition from (5.21) to (5.22) follows again from a change of variable in the action space from $\mathbf{Z}_n$ to $\mathbf{Y}_n$, where $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}_n$. The controlled random vector $\mathbf{Y}_n$ represents the queue lengths of the receiver buffers *after* transmission takes place in the $n^{th}$ slot, but *before* playout takes place (i.e., before $d^m$ packets are removed from user $m$'s buffer). The restrictions on the action space, $\mathbf{y}\succeq\mathbf{d}\vee\mathbf{x}$ and $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}]\le P$, ensure: (i) a nonnegative number of packets is transmitted to each user; (ii) there are at least $d^m$ packets in user $m$'s receiver buffer following transmission, in order to satisfy the underflow constraint; and (iii) the power constraint is satisfied.

Without the per slot peak power constraint, this $M$-dimensional problem would be separable, and could be solved by solving $M$ instances of the one-dimensional problem of Section 5.3; however, the joint power constraint couples the queues.[6] As

[6] This problem therefore falls into the class of *weakly coupled stochastic dynamic programs* [1, 65].

a result, the optimal transmission quantity to one receiver depends on the other receivers' queue length, as the following example shows.

**Example 5.7.** Assume receiver 1's channel is currently in a "poor" condition, receiver 2's channel is currently in a "medium" condition, and receiver 2's buffer contains enough packets to satisfy the demand for the next few slots. We consider two different scenarios for receiver 1's buffer level to show how the optimal transmission quantity to receiver 2 depends on receiver 1's buffer level. In Scenario 1, receiver 1's buffer already contains many packets. In this scenario, it may be beneficial for the scheduler to wait for receiver 2 to have a better channel condition, because it will be able to take full advantage of an "excellent" condition when it comes. In Scenario 2, receiver 1's queue only contains enough packets for playout in the current slot. It may be optimal to transmit some packets to receiver 2 in the current slot in this scenario. To see this, note that even if receiver 2 experiences the best possible channel condition in the next slot, the scheduler will need to allocate some power to receiver 1 in order to prevent receiver 1's buffer from emptying. Therefore, the scheduler anticipates not being able to take full advantage of receiver 2's "excellent" condition in the next slot, and may compensate by sending some packets in the current slot under the "medium" condition.

### 5.5.1 Structure of Optimal Policy for the Finite Horizon Discounted Expected Cost Problem

Before proceeding to the structure of the optimal transmission policy, we state some key properties of the value functions in the following theorem.

**Theorem 5.8.** *With two receivers and linear power-rate curves, the following statements are true for $n = 1, 2, \ldots, N$, and for all $\boldsymbol{s} \in \mathcal{S}$:*

**(i)** $V_{n-1}(\boldsymbol{x}, \boldsymbol{s})$ *is convex in* $\boldsymbol{x}$.

**(ii)** *$V_{n-1}(\boldsymbol{x},\boldsymbol{s})$ is supermodular in $\boldsymbol{x}$; i.e., for all $\bar{\boldsymbol{x}},\tilde{\boldsymbol{x}} \in I\!R^2_+$,*

$$V_{n-1}(\bar{\boldsymbol{x}},\boldsymbol{s}) + V_{n-1}(\tilde{\boldsymbol{x}},\boldsymbol{s}) \leq V_{n-1}(\bar{\boldsymbol{x}} \wedge \tilde{\boldsymbol{x}},\boldsymbol{s}) + V_{n-1}(\bar{\boldsymbol{x}} \vee \tilde{\boldsymbol{x}},\boldsymbol{s}) \ .$$

**(iii)** *$G_n(\boldsymbol{y},\boldsymbol{s})$ is convex in $\boldsymbol{y}$.*

**(iv)** *$G_n(\boldsymbol{y},\boldsymbol{s})$ is supermodular in $\boldsymbol{y}$; i.e., for all $\bar{\boldsymbol{y}},\tilde{\boldsymbol{y}} \in [d^1,\infty) \times [d^2,\infty)$,*

$$G_n(\bar{\boldsymbol{y}},\boldsymbol{s}) + G_n(\tilde{\boldsymbol{y}},\boldsymbol{s}) \leq G_n(\bar{\boldsymbol{y}} \wedge \tilde{\boldsymbol{y}},\boldsymbol{s}) + G_n(\bar{\boldsymbol{y}} \vee \tilde{\boldsymbol{y}},\boldsymbol{s}) \ .$$

**(v)** *$y_n^1 < \hat{y}_n^1$ implies:*

$$\inf\left\{ \operatorname*{argmin}_{y_n^2 \in [d^2,\infty)} \left\{ G_n\left(y_n^1, y_n^2, s^1, s^2\right) \right\} \right\} \geq \inf\left\{ \operatorname*{argmin}_{y_n^2 \in [d^2,\infty)} \left\{ G_n\left(\hat{y}_n^1, y_n^2, s^1, s^2\right) \right\} \right\}$$

*and $y_n^2 < \hat{y}_n^2$ implies:*

$$\inf\left\{ \operatorname*{argmin}_{y_n^1 \in [d^1,\infty)} \left\{ G_n\left(y_n^1, y_n^2, s^1, s^2\right) \right\} \right\} \geq \inf\left\{ \operatorname*{argmin}_{y_n^1 \in [d^1,\infty)} \left\{ G_n\left(y_n^1, \hat{y}_n^2, s^1, s^2\right) \right\} \right\} .$$

A detailed proof is included in Appendix A.4. Because $-\mathbf{c}_\mathbf{s}^{\mathrm{T}}\mathbf{x}$ is supermodular in $\mathbf{x}$, the key part of the induction step in the proof of (ii) is to show that $\min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \{G_{n-1}(\mathbf{y},\mathbf{s})\}$ is also supermodular in $\mathbf{x}$. Denoting $\operatorname{argmin}_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \{G_{n-1}(\mathbf{y},\mathbf{s})\}$ by $\mathbf{y}^*(\mathbf{x},\mathbf{s})$, we do this constructively by showing that for all $\bar{\mathbf{x}},\tilde{\mathbf{x}} \in I\!R^2_+$:

$$\begin{aligned}
& \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}},\mathbf{s})} \{G_{n-1}(\mathbf{y},\mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}},\mathbf{s})} \{G_{n-1}(\mathbf{y},\mathbf{s})\} \\
\leq \ & G_{n-1}(\bar{\mathbf{y}},\mathbf{s}) + G_{n-1}(\tilde{\mathbf{y}},\mathbf{s}) \\
\leq \ & G_{n-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}},\mathbf{s}),\mathbf{s}\Big) + G_{n-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}},\mathbf{s}),\mathbf{s}\Big) \qquad (5.24) \\
= \ & \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}},\mathbf{s})} \{G_{l-1}(\mathbf{y},\mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}},\mathbf{s})} \{G_{l-1}(\mathbf{y},\mathbf{s})\} ,
\end{aligned}$$

for a specific choice of $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}},\mathbf{s})$ and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}},\mathbf{s})$. The difficulty is cleverly constructing $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$, depending on the relative locations of $\bar{\mathbf{x}}$, $\tilde{\mathbf{x}}$, $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})$, and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}})$, so as to ensure (5.24) is true.

It follows from Theorem 5.8 that the structure of the optimal transmission policy for the finite horizon discounted expected cost problem is given by the following theorem.

**Theorem 5.9.** *For every $n \in \{1, 2, \dots, N\}$ and $\boldsymbol{s} \in \mathcal{S}^1 \times \mathcal{S}^2$, define the non-empty set of global minimizers of $G_n(\cdot, \boldsymbol{s})$:*

$$\mathcal{B}_n(\boldsymbol{s}) := \left\{ \hat{\boldsymbol{y}} \in [d^1, \infty) \times [d^2, \infty) : \ G_n(\hat{\boldsymbol{y}}, \boldsymbol{s}) = \min_{\boldsymbol{y} \in [d^1,\infty) \times [d^2,\infty)} G_n(\boldsymbol{y}, \boldsymbol{s}) \right\} \ .$$

*Define also*

$$b_n^1\left(\boldsymbol{s}\right) := \min\Big\{ y^1 \in [d^1, \infty) : (y^1, y^2) \in \mathcal{B}_n(\boldsymbol{s}) \text{ for some } y^2 \in [d^2, \infty) \Big\} \ ,$$

*and*

$$b_n^2\left(\boldsymbol{s}\right) := \min\Big\{ y^2 \in [d^2, \infty) : \big(b_n^1\left(\boldsymbol{s}\right), y^2\big) \in \mathcal{B}_n(\boldsymbol{s}) \Big\} \ .$$

*Then the vector $\boldsymbol{b}_n(\boldsymbol{s}) = \Big(b_n^1\left(\boldsymbol{s}\right), b_n^2\left(\boldsymbol{s}\right)\Big) \in \mathcal{B}_n(\boldsymbol{s})$ is a global minimizer of $G_n(\cdot, \boldsymbol{s})$. Define also the functions:*

$$f_n^1(x^2, \boldsymbol{s}) := \inf \left\{ \operatorname*{argmin}_{y^1 \in [d^1,\infty)} \left\{ G_n\left(y^1, x^2, s^1, s^2\right) \right\} \right\}, \text{ for } x^2 \in [d^2, \infty), \text{ and}$$
$$f_n^2(x^1, \boldsymbol{s}) := \inf \left\{ \operatorname*{argmin}_{y^2 \in [d^2,\infty)} \left\{ G_n\left(x^1, y^2, s^1, s^2\right) \right\} \right\}, \text{ for } x^1 \in [d^1, \infty).$$

*Note that by construction, $f_n^1\big(b_n^2(\boldsymbol{s}), \boldsymbol{s}\big) = b_n^1(\boldsymbol{s})$ and $f_n^2\big(b_n^1(\boldsymbol{s}), \boldsymbol{s}\big) = b_n^2(\boldsymbol{s})$. Partition $\mathbb{R}_+^2$ into the following seven regions:*

$$\mathcal{R}_I(n, \boldsymbol{s}) := \Big\{ \boldsymbol{x} \in \mathbb{R}_+^2 : \boldsymbol{x} \succeq \big(f_n^1(x^2, \boldsymbol{s}), f_n^2(x^1, \boldsymbol{s})\big) \text{ and } \boldsymbol{x} \neq \boldsymbol{b}_n(\boldsymbol{s}) \Big\}$$
$$\mathcal{R}_{II}(n, \boldsymbol{s}) := \Big\{ \boldsymbol{x} \in \mathbb{R}_+^2 : \boldsymbol{x} \preceq \boldsymbol{b}_n(\boldsymbol{s}) \text{ and } \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}} \left[\boldsymbol{b}_n(\boldsymbol{s}) - \boldsymbol{x}\right] \leq P \Big\}$$
$$\mathcal{R}_{III-A}(n, \boldsymbol{s}) := \Big\{ \boldsymbol{x} \in \mathbb{R}_+^2 : x^2 > b_n^2(\boldsymbol{s}) \text{ and } f_n^1(x^2, \boldsymbol{s}) - \frac{P}{c_{s^1}} \leq x^1 < f_n^1(x^2, \boldsymbol{s}) \Big\}$$
$$\mathcal{R}_{III-B}(n, \boldsymbol{s}) := \Big\{ \boldsymbol{x} \in \mathbb{R}_+^2 : x^1 > b_n^1(\boldsymbol{s}) \text{ and } f_n^2(x^1, \boldsymbol{s}) - \frac{P}{c_{s^2}} \leq x^2 < f_n^2(x^1, \boldsymbol{s}) \Big\}$$

$$\mathcal{R}_{IV-A}(n, \boldsymbol{s}) := \left\{ \boldsymbol{x} \in I\!R_+^2 : x^2 > b_n^2(\boldsymbol{s}) \text{ and } x^1 < f_n^1(x^2, \boldsymbol{s}) - \frac{P}{c_{s^1}} \right\}$$

$$\mathcal{R}_{IV-B}(n, \boldsymbol{s}) := \left\{ \boldsymbol{x} \in I\!R_+^2 : \boldsymbol{x} \preceq \boldsymbol{b}_n(\boldsymbol{s}) \text{ and } \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}} \left[ \boldsymbol{b}_n(\boldsymbol{s}) - \boldsymbol{x} \right] > P \right\}$$

$$\mathcal{R}_{IV-C}(n, \boldsymbol{s}) := \left\{ \boldsymbol{x} \in I\!R_+^2 : x^1 > b_n^1(\boldsymbol{s}) \text{ and } x^2 < f_n^2(x^1, \boldsymbol{s}) - \frac{P}{c_{s^2}} \right\} ,$$

*and define* $\mathcal{R}_{IV}(n, \boldsymbol{s}) := \mathcal{R}_{IV-A}(n, \boldsymbol{s}) \cup \mathcal{R}_{IV-B}(n, \boldsymbol{s}) \cup \mathcal{R}_{IV-C}(n, \boldsymbol{s})$.

*Then for Problem (**P5.1**) in the case of two receivers with linear power-rate curves, for all* $\boldsymbol{x} \notin \mathcal{R}_{IV}(n, \boldsymbol{s})$*, an optimal control action with* $n$ *slots remaining is given by:*

$$\boldsymbol{y}_n^*(\boldsymbol{x}, \boldsymbol{s}) := \begin{cases} \boldsymbol{x}, & \text{if } \ \boldsymbol{x} \in \mathcal{R}_I(n, \boldsymbol{s}) \\ \boldsymbol{b}_n(\boldsymbol{s}), & \text{if } \ \boldsymbol{x} \in \mathcal{R}_{II}(n, \boldsymbol{s}) \\ \left( f_n^1(x^2, \boldsymbol{s}), x^2 \right), & \text{if } \ \boldsymbol{x} \in \mathcal{R}_{III-A}(n, \boldsymbol{s}) \\ \left( x^1, f_n^2(x^1, \boldsymbol{s}) \right), & \text{if } \ \boldsymbol{x} \in \mathcal{R}_{III-B}(n, \boldsymbol{s}) \end{cases} . \tag{5.25}$$

*For all* $\boldsymbol{x} \in \mathcal{R}_{IV}(n, \boldsymbol{s})$*, there exists an optimal control action with* $n$ *slots remaining,* $\boldsymbol{y}_n^*(\boldsymbol{x}, \boldsymbol{s})$*, which satisfies:*

$$\boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}} \left[ \boldsymbol{y}_n^*(\boldsymbol{x}, \boldsymbol{s}) - \boldsymbol{x} \right] = P \ . \tag{5.26}$$

A detailed proof is included in Appendix A.5. Equation (5.26) says that it is optimal for the transmitter to allocate the full power budget for transmission when the vector of receiver buffer levels at the beginning of slot $n$ falls in region $\mathcal{R}_{IV}(n, \mathbf{s})$. We cannot say anything in general about the optimal allocation (split) of the full power budget between the two receivers when the starting buffer levels lie in region $\mathcal{R}_{IV}(n, \mathbf{s})$. Figure 5.6 shows the partition of $I\!R_+^2$ into the seven regions, and a diagram of the structure of the optimal transmission policy. Note that the figure shows the seven regions of the optimal policy for a *fixed realization* of the pair of channel conditions. Under different pairs of channel realizations, the seven regions have the same general form, but the targets $\mathbf{b}_n(\mathbf{s})$ are shifted and the boundary functions $f_n^1(x^2, \mathbf{s})$ and $f_n^2(x^1, \mathbf{s})$ are different.

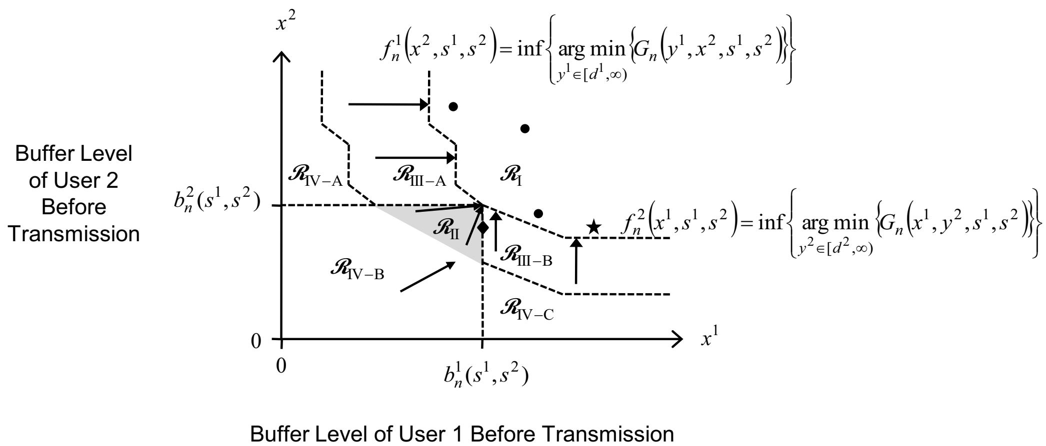


Figure 5.6. Structure of optimal policy for Problem (**P5.1**) in the case of two receivers with linear power-rate curves. The state in slot $n$ is $(\mathbf{x}, \mathbf{s})$. The seven regions described in Theorem 5.9 are labeled. The tails of the arrows represent the vectors of the receiver buffer levels at the beginning of slot $n$, and the heads of the arrows represent the vectors of the receiver buffer levels after transmission but before playout in slot $n$ under the optimal transmission policy. In region $\mathcal{R}_I(n, \mathbf{s})$, a single dot represents that it is optimal to not transmit any packets to either user. The ★ and ♦ represent possible starting buffer levels for Scenarios 1 and 2, respectively, in Example 5.7.

In some sense, the structure of the optimal policy outlined in Theorem 5.9 can be interpreted as an extension of the modified base-stock policy for the case of a single receiver outlined in Theorem 5.1. Namely, under each channel condition at each time, there is a critical number for each receiver $\left(b_n^m(\mathbf{s})\right)$ such that it is optimal to bring both receivers' buffer levels up to those critical numbers if it is possible to do so $\left(\text{region } \mathcal{R}_{II}(n, \mathbf{s})\right)$, and it is optimal to not transmit any packets if both receivers' buffer levels start beyond their critical numbers $\left(\text{region } \mathcal{R}_I(n, \mathbf{s})\right)$. However, this extended notion of the modified base-stock policy only captures the optimal behavior in two of the seven regions, and does not account for the coupling behavior between users that arises through the joint power constraint. For instance, possible starting buffer levels for Scenario 1 and Scenario 2 in Example 5.7 are illustrated in Figure

5.6 by the ★ and ⧫, respectively. Even though the buffer level of receiver 2 before transmission is the same under both scenarios, the optimal transmission quantity to receiver 2 is different under the two scenarios due to the different starting buffer levels of receiver 1.

### 5.5.2 Structure of the Optimal Policy for the Infinite Horizon Discounted Expected Cost Problems

In this section, we show that the structure of the optimal *stationary (or time-invariant)* policy for the infinite horizon discounted expected cost problem is the same as the structure of the optimal policy for the finite horizon discounted expected cost problem. Moreover, the boundaries of the seven regions of the finite horizon optimal policy shown in Figure 5.6 converge to the boundaries of the seven regions of the infinite horizon discounted expected cost optimal policy as the time horizon $N$ goes to infinity.

**Theorem 5.10.** *Define:*

**(i)** $V_\infty(\boldsymbol{x}, \boldsymbol{s}) := \lim\limits_{n\to\infty} V_n(\boldsymbol{x}, \boldsymbol{s})$, *for all* $\boldsymbol{x} \in I\!R^2_+$ *and* $\boldsymbol{s} \in \mathcal{S}$ *(this limit exists).*

**(ii)** $G_\infty(\boldsymbol{y}, \boldsymbol{s}) := \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}\boldsymbol{y} + h(\boldsymbol{y} - \boldsymbol{d}) + \alpha \cdot I\!E\big[V_\infty(\boldsymbol{y} - \boldsymbol{d}, \boldsymbol{S}') \mid \boldsymbol{S} = \boldsymbol{s}\big]$, *for all* $\boldsymbol{y} \in [d^1, \infty) \times [d^2, \infty)$ *and* $\boldsymbol{s} \in \mathcal{S}$.

**(iii)** $\mathcal{B}_\infty(\boldsymbol{s}) := \left\{ \hat{\boldsymbol{y}} \in [d^1, \infty) \times [d^2, \infty) : \ G_\infty(\hat{\boldsymbol{y}}, \boldsymbol{s}) = \min\limits_{\boldsymbol{y} \in [d^1,\infty)\times[d^2,\infty)} G_\infty(\boldsymbol{y}, \boldsymbol{s}) \right\}$.

**(iv)** $b^1_\infty(\boldsymbol{s}) := \min\Big\{ y^1 \in [d^1, \infty) : (y^1, y^2) \in \mathcal{B}_\infty(\boldsymbol{s})$ *for some* $y^2 \in [d^2, \infty)\Big\}$.

**(v)** $b^2_\infty(\boldsymbol{s}) := \min\Big\{ y^2 \in [d^2, \infty) : \big(b^1_\infty(\boldsymbol{s}), y^2\big) \in \mathcal{B}_\infty(\boldsymbol{s})\Big\}$.

**(vi)** $\boldsymbol{b}_\infty(\boldsymbol{s}) := \Big(b^1_\infty(\boldsymbol{s}), b^2_\infty(\boldsymbol{s})\Big)$.

**(vii)** *The functions*

$$f^1_\infty(x^2, \boldsymbol{s}) := \inf\left\{\underset{y^1\in[d^1,\infty)}{\operatorname{argmin}}\left\{G_\infty\left(y^1, x^2, s^1, s^2\right)\right\}\right\}, \text{ for } x^2 \in [d^2, \infty), \text{ and}$$

$$f^2_\infty(x^1, \boldsymbol{s}) := \inf\left\{\underset{y^2\in[d^2,\infty)}{\operatorname{argmin}}\left\{G_\infty\left(x^1, y^2, s^1, s^2\right)\right\}\right\}, \text{ for } x^1 \in [d^1, \infty).$$

**(viii)** *The seven regions $\mathcal{R}_I(\infty, \boldsymbol{s}) - \mathcal{R}_{IV-C}(\infty, \boldsymbol{s})$, defined in the same way as in Theorem 5.9, with $n$ replaced by $\infty$.*

*Then*

**(a)** *$V_\infty(\boldsymbol{x}, \boldsymbol{s})$ satisfies the $\alpha$-discounted optimality equation ($\alpha$-DCOE):*

$$V_\infty(\boldsymbol{x}, \boldsymbol{s}) = \min_{\boldsymbol{y}\in\tilde{A}^d(\boldsymbol{x},\boldsymbol{s})}\left\{\begin{array}{l} \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}[\boldsymbol{y}-\boldsymbol{x}] + \boldsymbol{h}\left(\boldsymbol{y}-\boldsymbol{d}\right) \\ +\alpha\cdot I\!E\left[V_\infty(\boldsymbol{y}-\boldsymbol{d}, \boldsymbol{S}') \;\middle|\; \boldsymbol{S}=\boldsymbol{s}\right] \end{array}\right\},$$

$$\forall \boldsymbol{x}\in I\!R^2_+, \forall \boldsymbol{s}\in\mathcal{S}\ . \tag{5.27}$$

**(b)** *An optimal stationary policy for Problem (**P5.2**) in the case of two receivers with linear power-rate curves is given by $\boldsymbol{\pi}^*_{\boldsymbol{\infty}} = (\boldsymbol{y}^*_\infty, \boldsymbol{y}^*_\infty, \ldots)$, where*

$$\boldsymbol{y}^*_\infty(\boldsymbol{x}, \boldsymbol{s}) := \begin{cases} \boldsymbol{x}, & \text{if } \ \boldsymbol{x}\in\mathcal{R}_I(\infty, \boldsymbol{s}) \\ \boldsymbol{b}_\infty(\boldsymbol{s}), & \text{if } \ \boldsymbol{x}\in\mathcal{R}_{II}(\infty, \boldsymbol{s}) \\ \left(f^1_\infty(x^2, \boldsymbol{s}), x^2\right), & \text{if } \ \boldsymbol{x}\in\mathcal{R}_{III-A}(\infty, \boldsymbol{s}) \\ \left(x^1, f^2_\infty(x^1, \boldsymbol{s})\right), & \text{if } \ \boldsymbol{x}\in\mathcal{R}_{III-B}(\infty, \boldsymbol{s}) \end{cases},$$

*and for all $\boldsymbol{x}\in\mathcal{R}_{IV}(\infty, \boldsymbol{s})$, there exists an optimal control action, $\boldsymbol{y}^*_\infty(\boldsymbol{x}, \boldsymbol{s})$, which satisfies:*

$$\boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}\left[\boldsymbol{y}^*_\infty(\boldsymbol{x}, \boldsymbol{s}) - \boldsymbol{x}\right] = P\ .$$

**(c)** $\lim_{n\to\infty} \boldsymbol{b}_n(\boldsymbol{s}) = \boldsymbol{b}_\infty(\boldsymbol{s})$ *for all $\boldsymbol{s}\in\mathcal{S}$.*

**(d)** $\lim_{n\to\infty} f_n^1(x^2,\boldsymbol{s}) = f_\infty^1(x^2,\boldsymbol{s})$ *for all* $x^2 \in [d^2,\infty)$ *and* $\boldsymbol{s} \in \mathcal{S}$.

**(e)** $\lim_{n\to\infty} f_n^2(x^1,\boldsymbol{s}) = f_\infty^2(x^1,\boldsymbol{s})$ *for all* $x^1 \in [d^1,\infty)$ *and* $\boldsymbol{s} \in \mathcal{S}$.

A detailed proof of Theorem 5.10 is included in Appendix B.2.

### 5.5.3 Structure of the Optimal Policy for the Infinite Horizon Average Expected Cost Problems

In this section, we again use the vanishing discount approach to show that the structure of the optimal policy for the finite horizon expected cost and infinite horizon discounted expected cost problems extends to the infinite horizon average expected cost problem. As in Section 5.4.4, we make explicit the dependence of the value functions and optimal policies from the corresponding infinite horizon discounted expected cost problem on the discount factor, $\alpha$.

**Theorem 5.11.** *For all* $\alpha \in [0,1)$, *define:*

$$\begin{aligned} m_{\infty,\alpha} &:= \inf_{\substack{\boldsymbol{x}\in I\!R_+^2 \\ \boldsymbol{s}\in\mathcal{S}}} V_{\infty,\alpha}(\boldsymbol{x},\boldsymbol{s}), && (5.28)\\ \rho^* &:= \lim_{\alpha\nearrow 1}(1-\alpha)\cdot m_{\infty,\alpha}, \text{ and} && (5.29)\\ w_{\infty,\alpha}(\boldsymbol{x},\boldsymbol{s}) &:= V_{\infty,\alpha}(\boldsymbol{x},\boldsymbol{s}) - m_{\infty,\alpha},\ \forall \boldsymbol{x}\in I\!R_+^2,\ \forall \boldsymbol{s}\in\mathcal{S}. && (5.30) \end{aligned}$$

*Then:*

**(a)** *There exists a continuous function* $w_{\infty,1}(\cdot,\cdot)$ *and a selector* $\boldsymbol{y}^*_{\infty,1}(\cdot,\cdot)$ *that satisfy the ACOE:*

$$\begin{aligned} \rho^* + w_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) &= \min_{\boldsymbol{y}\in\tilde{\mathcal{A}}^d(\boldsymbol{x},\boldsymbol{s})} \left\{ \begin{array}{l} \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}[\boldsymbol{y}-\boldsymbol{x}] + \boldsymbol{h}(\boldsymbol{y}-\boldsymbol{d}) \\ + I\!E\left[w_{\infty,1}(\boldsymbol{y}-\boldsymbol{d},\boldsymbol{S}') \mid \boldsymbol{S}=\boldsymbol{s}\right] \end{array} \right\} \qquad (5.31)\\ &= \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}\Big[\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{x}\Big] + \boldsymbol{h}\Big(\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{d}\Big) \\ &\quad + I\!E\left[w_{\infty,1}\Big(\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{d},\boldsymbol{S}'\Big)\Big|\boldsymbol{S}=\boldsymbol{s}\right],\ \forall \boldsymbol{x}\in I\!R_+^2,\ \forall \boldsymbol{s}\in\mathcal{S}. \end{aligned}$$

**(b)** *The stationary policy* $\boldsymbol{\pi^*_{\infty,1}} = (\boldsymbol{y}^*_{\infty,1}, \boldsymbol{y}^*_{\infty,1}, \ldots)$ *is optimal for Problem (**P5.3**) in the case of two receivers with linear power-rate curves.*

**(c)** *The resulting optimal average cost beginning from any initial state* $(\boldsymbol{x}, \boldsymbol{s}) \in I\!R^2_+ \times \mathcal{S}$ *is* $\rho^*$.

**(d)** *For every increasing sequence of discount factors* $\{\alpha(l)\}_{l=1,2,\ldots}$ *approaching 1, there exists a subsequence* $\{\alpha(l_i)\}_{i=1,2,\ldots}$ *approaching 1 such that:*

$$w_{\infty,1}(\boldsymbol{x}, \boldsymbol{s}) \;\; = \lim_{i \to \infty} w_{\infty,\alpha(l_i)}(\boldsymbol{x}, \boldsymbol{s}), \;\; \forall \boldsymbol{x} \in I\!R^2_+, \;\; \forall \boldsymbol{s} \in \mathcal{S}.$$

*Therefore, for every* $\boldsymbol{s} \in \mathcal{S}$, $w_{\infty,1}(\boldsymbol{x}, \boldsymbol{s})$ *is convex and supermodular in* $\boldsymbol{x}$.

**(e)** *For every* $(\boldsymbol{x}, \boldsymbol{s}) \in I\!R^2_+ \times \mathcal{S}$ *and increasing sequence of discount factors* $\{\alpha(l)\}_{l=1,2,\ldots}$ *approaching 1, there exists a subsequence* $\{\alpha(l_i)\}_{i=1,2,\ldots}$ *approaching 1 and a sequence* $\{\boldsymbol{x}(i)\}_{i=1,2,\ldots}$ *approaching* $\boldsymbol{x}$ *such that:*

$$\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x}, \boldsymbol{s}) \;\; = \lim_{i \to \infty} \boldsymbol{y}^*_{\infty,\alpha(l_i)}(\boldsymbol{x}(i), \boldsymbol{s}) \; .$$

**(f)** *There exists an optimal stationary policy with the same structure as statement (b) in Theorem 5.10.*

A detailed proof of Theorem 5.11 is included in Appendix C.

## 5.6 Relaxation of the Strict Underflow Constraints

In some applications, it may not be the case that the peak power per slot is always sufficient to transmit one slot's worth of packets to each receiver, even under the worst channel conditions. In this case, a more appropriate model is to relax the strict underflow constraints, and allow underflow at a cost. One way to model this situation is to allow the receivers' queues to be negative, with a negative buffer level representing the number of packets that the playout process is behind. Then, in addition to the holding costs assessed on positive buffer levels, shortage costs are assessed on negative buffer levels. With some minor alterations to the proofs, it is straightforward to show that as long as the shortage cost function is a convex function of the negative buffer level, the structural results of Theorems 5.1, 5.3 and 5.9 are essentially unchanged by the relaxation of the strict underflow constraints to loose underflow constraints with penalties on underflow. This is not too surprising as the strict underflow constraint case we consider can be thought of as the limiting case as the penalties on underflow go to infinity.[7]

## 5.7 Summary

In this chapter, we considered the problem of transmitting data to one or more receivers over a shared wireless channel in a manner that minimizes power consumption and prevents the receivers' buffers from emptying. We showed that under the finite horizon discounted expected cost, infinite horizon discounted expected cost, and infinite horizon average expected cost criteria, the optimal transmission policy to a single receiver under linear power-rate curves has a modified base-stock structure. When the power-rate curves are generalized to piecewise-linear power-rate curves, the optimal transmission policy to a single receiver has a finite generalized base-stock structure. For the special case when holding costs are linear, the stochastic process representing the channel condition evolution over time is IID, and the maximum number of packets that can be transmitted at any given marginal power cost in a slot is an integer multiple of the drainage rate of the receiver's buffer, we

[7] Tracking the number of packets that the playout process is behind in this manner corresponds to the *complete backlogging* assumption in inventory theory. An alternate model is to say that a packet is of no use once it misses its deadline, penalize missed packets, and keep the receiver queue length at zero. This model corresponds to the *lost sales* assumption in inventory theory.

presented an efficient method to compute the critical numbers that fully characterize the modified base-stock and finite generalized base-stock policies.

We also analyzed the structure of the optimal transmission policy for the case of two receivers. In some sense, the structure of the optimal policy was shown to be an extension of the modified base-stock policy; however, the peak power constraint couples the optimal scheduling of the two data streams. In the next two chapters, we compare the problems considered in this chapter to related problems from the wireless communications and inventory theory literatures, respectively.

# *Chapter 6*

# The Role of Deadline Constraints in Opportunistic Scheduling

In this chapter, we compare the finite horizon problem from the previous chapter, Problem **(P5.1)**, to two related problems from the wireless communications literature. The overarching goal in all three problems is to do energy-efficient transmission scheduling, subject to deadline constraints. The primary purpose of our comparison is to elucidate the role of the deadline constraints.

## 6.1 Problem Formulations

In all three problems, a single source transmits data to a single user/receiver over a time-varying wireless channel. As in the previous chapter, we consider a discrete time horizon of length $N$, the scheduler learns the channel condition perfectly at the beginning of each slot, and the transmission of $z$ data packets under channel condition $s$ incurs an energy cost of $c(z, s)$. In this chapter, we take the channel condition to be independent and identically distributed (IID) from slot to slot and the discount factor to be equal to 1.

The primary objective in deriving a good transmission policy is once again to minimize energy consumption while meeting the deadline constraint(s) and possibly a power constraint in each slot. Thus, all three problems we discuss in this chapter can be formulated as Markov Decision Processes (MDPs) with the following common form:

$$
\begin{aligned}
& \min_{\boldsymbol{\pi} \in \boldsymbol{\Pi}} \; \mathbb{E}^{\boldsymbol{\pi}} \left\{ \sum_{n=1}^{N} c(Z_n, S_n) \mid \mathcal{F}_N \right\} \\
\text{s.t.} \quad & \text{Per Slot Power Constraints and} \\
& \text{Deadline Constraint(s)} \; ,
\end{aligned} \tag{6.1}
$$

where, as in Chapter 5, $\mathcal{F}_N$ denotes all information available at the beginning of the time horizon, $Z_n = \pi_n \left(Z_N, Z_{N-1}, \ldots, Z_{n+1}, S_N, S_{N-1}, \ldots, S_n\right)$ is the number of packets the scheduler decides to transmit in slot $n$, and $\boldsymbol{\Pi}$ denotes the set of all randomized and deterministic control laws. Next, we specify the precise variant of (6.1) for each the three problems.

### 6.1.1 Strict Underflow Constraints, Linear Power-Rate Curves, and a Power Constraint in Each Slot

The first problem we consider is Problem (**P5.1**) in the case of a single receiver with linear power-rate curves, no holding costs, and a discount rate of $\alpha = 1$. As mentioned in the previous chapter, the strict underflow constraints can also be interpreted as multiple deadline constraints: the source must transmit at least $d$ packets by the end of the first slot, $2d$ packets by the end of the second slot, and so forth. So at each time $n$, the underflow constraint:

$$X_n + Z_n \geq \max\{0, d - X_n\}$$

is equivalent to:

$$\sum_{t=n}^{N} Z_t \geq (N - n + 1) \cdot d \; .$$

Therefore, for this problem, the general formulation (6.1) becomes:

$$\begin{aligned}
&\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \; I\!E^{\boldsymbol{\pi}}\left\{\sum_{n=1}^{N} c_{S_n}\cdot Z_n \mid \mathcal{F}_N\right\} \\
\text{s.t.} \quad & c_{S_n}\cdot Z_n \le P,\; w.p.1,\; \forall n \in \left\{N, N-1, \ldots, 1\right\} \\
\text{and} \quad & \sum_{t=n}^{N} Z_t \ge (N-n+1)\cdot d,\; w.p.1,\; \forall n \in \left\{N, N-1, \ldots, 1\right\} .
\end{aligned}$$

We refer to this problem as Problem (**P6.1**).

### 6.1.2 Single Deadline Constraint, Linear Power-Rate Curves, and a Power Constraint in Each Slot

The second problem we consider features linear power-rate curves, a power constraint in each slot, and a *single* deadline constraint. The single constraint is to send $d_{\text{total}}$ packets across the channel by the end of the $N$ slot horizon. As with Problem (**P6.1**), the maximum transmission power in any given slot is denoted by $P$, and for each possible channel condition $s$, there exists a constant $c_s$ such that $c(z,s) = c_s \cdot z$. In order to ensure that it is always possible to satisfy the deadline constraint, we assume that $N \cdot \left(\frac{P}{c_{s_{\text{worst}}}}\right) \ge d_{\text{total}}$, or, equivalently, $c_{s_{\text{worst}}} \le \frac{N\cdot P}{d_{\text{total}}}$, where $c_{s_{\text{worst}}}$ is the energy cost per packet transmitted under the worst possible channel condition.[1] Thus, even if the channel is in the worst possible condition for the entire duration of the time horizon, it is still possible to send $d_{\text{total}}$ packets by transmitting at full power in every slot. The general formulation (6.1) becomes:

$$\begin{aligned}
&\min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \; I\!E^{\boldsymbol{\pi}}\left\{\sum_{n=1}^{N} c_{S_n}\cdot Z_n \mid \mathcal{F}_N\right\} \\
\text{s.t.} \quad & c_{S_n}\cdot Z_n \le P,\; w.p.1,\; \forall n \in \left\{N, N-1, \ldots, 1\right\} \\
\text{and} \quad & \sum_{n=1}^{N} Z_n \ge d_{\text{total}},\; w.p.1 .
\end{aligned}$$

We refer to this problem as Problem (**P6.2**). It was introduced and analyzed by Fu, Modiano, and Tsitsiklis in [51, Section III-D] and [52, Section III-D].

[1] This is an unstated assumption in [51, Section III-D] and [52, Section III-D].

### 6.1.3 Single Deadline Constraint and Convex Monomial Power-Rate Curves

The third problem we consider features the same single deadline constraint as Problem (**P6.2**); however, there is no per slot power constraint imposed, and the energy cost from transmission is a convex monomial function of the number of packets sent. Namely, for every channel condition $s$, there exists a constant $k_s$ such that $c(z,s) = \frac{z^\mu}{k_s}$, where $\mu > 1$ is the fixed monomial order of the cost function. As mentioned in Section 4.2.1, such a power-rate curve may be more appropriate in the high SNR regime. The general formulation (6.1) becomes:

$$\begin{aligned} \min_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \; & I\!\!E^{\boldsymbol{\pi}}\left\{\sum_{n=1}^{N}\frac{(Z_n)^\mu}{k_{S_n}} \mid \mathcal{F}_N\right\} \\ \text{s.t.} \quad & \sum_{n=1}^{N} Z_n \geq d_{\text{total}}, \;\; w.p.1 \; . \end{aligned}$$

We refer to this problem as Problem (**P6.3**). It was introduced and analyzed by Lee and Jindal in [92].

## 6.2 Structures of the Optimal Policies

In this section, we present the structures of the optimal policies for each of the three problems as straightforwardly as possible, without changing drastically the original presentations. All three problems can be solved using standard dynamic programming (see, e.g., [20]), and the structures of the optimal policies follow from properties of the value functions or expected costs-to-go.

Specializing the dynamic program (5.4) to the case of no holding costs, a discount rate of 1, and an IID channel, the dynamic program for Problem (**P6.1**) is:

$$
\begin{aligned}
V_n(x,s) &= \min_{\max(0,d-x)\le z\le \frac{P}{c_s}} \left\{c_s\cdot z + \mathbb{E}\big[V_{n-1}(x+z-d,S_{n-1})\big]\right\} && (6.2)\\
&= -c_s\cdot x + \min_{\max(x,d)\le y\le x+\frac{P}{c_s}} \left\{c_s\cdot y + \mathbb{E}\big[V_{n-1}(y-d,S_{n-1})\big]\right\} && (6.3)\\
& \qquad\qquad n = N, N-1, \ldots, 1\\
V_0(x,s) &= 0,\ \forall x, \forall s.
\end{aligned}
$$

For Problem (**P6.2**), Fu, Modiano, and Tsitsiklis [52] take the information state at time $n$ to be the pair $(Q_n, S_n)$, where $Q_n$ represents the number of packets remaining to be transmitted at time $n$, and $S_n$ denotes the channel condition in slot $n$. The dynamics of packets remaining to be transmitted are $Q_{n-1} = Q_n - Z_n$, as $Z_n$ packets are transmitted during slot $n$. The dynamic programming equations for this problem are given by:

$$
\begin{aligned}
V_n(q,s) &= \min_{0\le z\le \min\left(q,\frac{P}{c_s}\right)} \left\{c_s\cdot z + \mathbb{E}\big[V_{n-1}(q-z,S_{n-1})\big]\right\} && (6.4)\\
&= c_s\cdot q + \min_{\max(0,q-\frac{P}{c_s})\le u\le q} \left\{-c_s\cdot u + \mathbb{E}\big[V_{n-1}(u,S_{n-1})\big]\right\}, && (6.5)\\
& \qquad\qquad n = N, N-1, \ldots, 1\\
V_0(q,s) &= \begin{cases} 0, & \text{if } q=0\\ \infty, & \text{if } q>0 \end{cases},\ \forall s\ .
\end{aligned}
$$

Here, the transition from (6.4) to (6.5) is done by a change of variable in the action space from $Z_n$ to $U_n$, where $U_n = Q_n - Z_n$. The controlled random variable $U_n$ represents the number of packets remaining to be transmitted *after* transmission takes place in the $n^{th}$ slot. The restrictions on the action space, $\max(0, q - \frac{P}{c_s}) \le u \le q$, ensure: (i) a nonnegative number of packets is transmitted; (ii) no more than $d_{\text{total}}$ packets are transmitted over the course of the horizon; and (iii) the power constraint is satisfied.

Note that the dynamic programming equations (6.3) and (6.5) have the following common form:

$$V_n(x,s) \;=\; f(x,s) + \min_{w_1(x,s)\le a\le w_2(x,s)} \left\{h_1(a) + \mathbb{E}\big[V_{n-1}(h_2(a),s)\big]\right\}, \qquad n = N, N-1, \ldots, 1\ , \tag{6.6}$$

where $(x, s)$ is the current state and $a$ represents the action. The key realizations for both problems are (i) $h_1(a)$ is convex in $a$ and $h_2(a)$ is an affine function of $a$; and (ii) for any fixed $s$, $f(x, s)$, $w_1(x, s)$, and $V_{n-1}(x, s)$ are all convex in $x$, and $w_2(x, s)$ is concave in $x$. These functional properties can be shown inductively. By Lemma A.1 in Appendix A.1, we have:

$$V_n(x,s) = f(x,s) + F(w_1(x,s)) + G(w_2(x,s)),$$

which is convex in $x$ for a fixed $s$, because $F(w_1(x, s))$ is the composition of a convex nondecreasing function with a convex function and $G(w_2(x, s))$ is the composition of a convex nonincreasing function with a concave function (see, e.g., [23, Section 3.2] for the relevant results on convexity-preserving operations). Furthermore, by Lemma A.1, if, for a fixed $s$, $\beta_n(s)$ is a global minimizer of $h_1(a)+\mathbb{E}\big[V_{n-1}(h_2(a), s)\big]$ over all $a$, then the optimal action has the form:

$$a_n^*(x,s) := \begin{cases} w_1(x,s), & \text{if } \beta_n(s) < w_1(x,s) \\ \beta_n(s), & \text{if } w_1(x,s) \le \beta_n(s) \le w_2(x,s) \\ w_2(x,s), & \text{if } w_2(x,s) < \beta_n(s) \end{cases} . \tag{6.7}$$

Recall from Section 5.3.1 that the optimal transmission policy in (6.7) is a *modified base-stock policy*. For Problem (**P6.1**), the structure of this policy is shown in Figure 5.3.

Applying this line of analysis to the dynamic program (6.5) for Problem (**P6.2**), we see that when the channel condition is in state $s$ at time $n$, and there are $q$ packets remaining to be transmitted by the deadline, the optimal action is given by:

$$u_n^*(q,s) := \begin{cases} \max\left(0, q - \frac{P}{c_s}\right), & \text{if } \beta_n(s) < \max\left(0, q - \frac{P}{c_s}\right) \\ \beta_n(s), & \text{if } \max\left(0, q - \frac{P}{c_s}\right) \leq \beta_n(s) \leq q \\ q, & \text{if } q < \beta_n(s) \end{cases}, \tag{6.8}$$

for some sequence of *critical numbers* $\{\beta_n(s)\}_{s \in \mathcal{S}}$. Changing variables back to the original action variable $Z_n$ and noting that $\beta_n(s) \geq 0$ for all $n$ and $s$, (6.8) is equivalent to:

$$z_n^*(q,s) := \begin{cases} \frac{P}{c_s}, & \text{if } \beta_n(s) + \frac{P}{c_s} < q \\ q - \beta_n(s), & \text{if } \beta_n(s) \leq q \leq \beta_n(s) + \frac{P}{c_s} \\ 0, & \text{if } q < \beta_n(s) \end{cases}, \tag{6.9}$$

See Fig. 6.1 for diagrams of this optimal policy. The critical numbers $\{\beta_n(s)\}_{s \in \mathcal{S}}$ can be calculated recursively when for every $s \in \mathcal{S}$,

$$c_s = \frac{c_{s_{\text{worst}}}}{\hat{l}} \text{ for some } \hat{l} \in I\!N. \tag{6.10}$$

For further details on the calculation of these critical numbers, see [52].

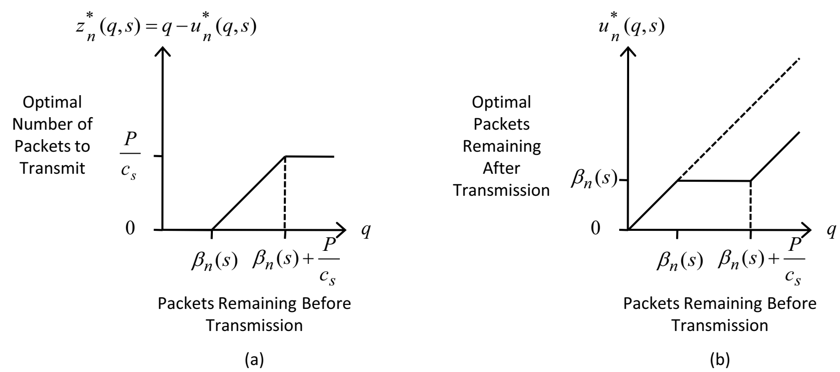


Figure 6.1. Structure of optimal policy for Problem (**P6.2**). The state in slot $n$ is $(q, s)$, the number of packets remaining to be transmitted before transmission and the current channel condition. (a) depicts $z^*$, the optimal transmission quantity, and (b) depicts $u^*$, the optimal number of packets remaining to be transmitted after transmission in slot $n$.

Like Problem (**P6.2**), Lee and Jindal [92] take the information state for Problem (**P6.3**) to be the pair $(Q_n, S_n)$, where $Q_n$ represents the number of packets remaining to be transmitted at time $n$, and $S_n$ denotes the channel condition in slot $n$. The dynamics of packets remaining to be transmitted are once again $Q_{n-1} = Q_n - Z_n$. The dynamic programming equations for Problem (**P6.3**) are given by:

$$\begin{aligned} V_n(q,s) &= \min_{z \geq 0} \left\{ \frac{z^\mu}{k_s} + \mathbb{E}\big[V_{n-1}(q-z, S_{n-1})\big] \right\}, n = N, N-1, \ldots, 1 \quad (6.11) \\ V_0(q,s) &= \begin{cases} 0, & \text{if } q = 0 \\ \infty, & \text{if } q > 0 \end{cases} . \end{aligned}$$

The key idea of Lee and Jindal is to show inductively that $\mathbb{E}\big[V_{n-1}(q-z, S_{n-1})\big] = \xi_{n-1,\mu} \cdot (q-z)^\mu$ for some constant $\xi_{n-1,\mu}$ that depends on the time $n-1$ and the known monomial order $\mu$. Therefore,

$$z_n^*(q,s) = \operatorname*{argmin}_{z \geq 0} \left\{ \frac{z^\mu}{k_s} + \xi_{n-1,\mu} \cdot (q-z)^\mu \right\} . \tag{6.12}$$

Differentiating the inner term of the right-hand side of (6.12) with respect to $z$ and setting it equal to zero yields $z_n^*(q,s) = \lambda_{n,\mu}(s) \cdot q$, for some $\lambda_{n,\mu}(s) \in [0,1]$. Namely, with $n$ slots remaining in the time horizon, the optimal control action is to send a fraction, $\lambda_{n,\mu}(s)$, of the remaining packets to be sent. Here, the fraction to send depends on the time remaining in the horizon, $n$; the current condition of the channel, $s$; and the parameter representing the monomial order of the cost function, $\mu$. The fractions $\lambda_{n,\mu}(s)$ can be computed recursively. Note that plugging the optimal $z_n^*(q,s)$ back into (6.11) yields:

$$\mathbb{E}\left[V_n(q,S)\right] = \mathbb{E}\left[\frac{(\lambda_{n,\mu}(S) \cdot q)^\mu}{k_S} + \xi_{n-1,\mu} \cdot (q - \lambda_{n,\mu}(S) \cdot q)^\mu\right] = \xi_{n,\mu} \cdot q^\mu$$

for some constant $\xi_{n,\mu}$, completing the induction step on the form of $\mathbb{E}\left[V_n(q,S)\right]$.

Finally, Lee and Jindal also show that for each fixed channel state $s$, the fraction $\lambda_{n,\mu}(s)$ is decreasing in $n$. In other words, the scheduler is more *selective* or *opportunistic* when the deadline is far away, as it sends a lower fraction of the remaining packets than it would under the same state closer to the deadline. This makes intuitive sense as it has more opportunities to wait for a very good channel realization when the deadline is farther away.

## 6.3 Comparison of the Problems

In this section, we provide further intuition behind the role of deadlines by comparing the above problems. First, we show that Problems (**P6.1**) and (**P6.2**) are equivalent when a certain technical condition holds. Next, we examine how the extra deadline constraints in Problem (**P6.1**) affect the optimal scheduling policy, as compared with Problem (**P6.2**). We finish with some conclusions on the role of deadline constraints.

### 6.3.1 A Sufficient Condition for the Equivalence of Problems (P6.1) and (P6.2)

In this section, we transform the dynamic programs (6.5) and (6.3) to find a condition under which Problems (**P6.1**) and (**P6.2**) are equivalent.

In Problem (**P6.1**), we change the state space from packets in the receiver's buffer to total packets transmitted since the beginning of the horizon ($T_n = X_n + (N - n) \cdot d$), and we change the action space from packets in the receiver's buffer following transmission to total packets sent (since the beginning of the horizon) after transmission in the $n^{th}$ slot ($A_n = Y_n + (N - n) \cdot d$). A straightforward interchange argument shows it is not optimal to send more than $N \cdot d$ packet during the horizon, so with the above changes of variables, the dynamic program (6.3) becomes:

$$
\begin{aligned}
V_n(t,s) &= -c_s \cdot t \\
&\quad + \min_{\max(t,(N-n+1)\cdot d) \leq a \leq \min\left(t+\frac{P}{c_s}, N\cdot d\right)} \left\{ c_s \cdot a + \mathbb{E}\left[V_{n-1}(t, S_{n-1})\right] \right\}, \\
&\qquad\qquad n = N, N-1, \ldots, 1, \\
V_0(t,s) &= 0, \ \forall t, \forall s \ . \qquad (6.13)
\end{aligned}
$$

In Problem (**P6.2**), there is just a single deadline constraint; however, because the terminal cost is set to $\infty$ if all the data is not transmitted by the deadline, the scheduler must transmit enough data in each slot so that it can still complete the job if the channel is in the worst possible condition in all subsequent slots. Thus, the scheduler can leave no more than $\frac{P}{c_{s_{\text{worst}}}}$ packets for the final slot, no more than $2 \cdot \frac{P}{c_{s_{\text{worst}}}}$ packets for the last two slots, and so forth. So there are in fact implicit constraints on how much data can remain to be transmitted at the end of each slot. If we make these implicit constraints explicit, then the dynamic program (6.5) becomes:

$$
\begin{aligned}
V_n(q,s) &= c_s \cdot q \\
&\quad + \min_{\max(0,q-\frac{P}{c_s}) \leq u \leq \min\left(q,(n-1)\cdot\frac{P}{c_{s_{\text{worst}}}}\right)} \left\{ -c_s \cdot u + \mathbb{E}\left[V_{n-1}(u, S_{n-1})\right] \right\}, \\
&\qquad\qquad n = N, N-1, \ldots, 1, \\
V_0(q,s) &= 0, \ \forall q, \forall s \ .
\end{aligned}
$$

Next, we change the state space from total packets remaining to be transmitted to total packets transmitted since the beginning of the horizon ($T_n = d_{\text{total}} - Q_n$), and we change the action space from total packets remaining to be transmitted after transmission in the $n^{th}$ slot to total packets sent (since the beginning of the horizon) after transmission in the $n^{th}$ slot ($A_n = d_{\text{total}} - U_n$). The resulting dynamic program is:

$$\begin{aligned} V_n(t,s) &= -c_s \cdot t \\ &\quad + \min_{\max\left(t, d_{\text{total}}-(n-1)\cdot \frac{P}{c_{s_{\text{worst}}}}\right) \le a \le \min\left(t+\frac{P}{c_s}, d_{\text{total}}\right)} \left\{ c_s \cdot a + \mathbb{E}\left[V_{n-1}(t, S_{n-1})\right]\right\}, \\ &\qquad\qquad n = N, N-1, \ldots, 1, \\ V_0(t,s) &= 0, \ \forall t, \forall s \,. \end{aligned} \tag{6.14}$$

The dynamic programs (6.13) and (6.14) associated with Problems (**P6.1**) and (**P6.2**), respectively, become identical when the following two conditions are satisfied:

**(C1)** $d_{\text{total}} = N \cdot d$ (i.e., the total number of packets to send over the horizon of $N$ slots is the same for both problems).

**(C2)** $\frac{P}{c_{s_{\text{worst}}}} = d$ (i.e., the maximum number of packets that can be transmitted under the worst channel condition is equal to the number of packets removed from the receiver's buffer at the end of each slot in Problem (**P6.1**)).

Furthermore, if condition (**C2**) is not satisfied, then $\frac{P}{c_{s_{\text{worst}}}} > d$, because we require that $\frac{P}{c_{s_{\text{worst}}}} \geq d$ for Problem (**P6.1**) to be well-defined. Thus, when condition (**C1**) is satisfied, but condition (**C2**) is not satisfied, the action space at time $n$ and state $(t, s)$ in (6.14) contains the action space at the same time and state in (6.13). This is because the explicit deadline constraints resulting from the strict underflow constraints in Problem (**P6.1**) are more restrictive than the implicit deadline constraints in Problem (**P6.2**).

### 6.3.2 Inverse Water-Filling Interpretations

In this section, we interpret Problems (**P6.1**) and (**P6.2**) within the context of the inverse water-filling procedure introduced in Section 4.1. The aim is to show

how the extra deadline constraints in Problem (**P6.1**) affect the optimal scheduling policy.

We start with Problem (**P6.2**), which features a single deadline constraint. If, at the beginning of the horizon, the scheduler happens to know the realizations of all future channel conditions, $s_N, s_{N-1}, \ldots, s_1$, then Problem (**P6.2**) reduces to the following convex optimization problem:

$$\begin{aligned} \min_{(z_N, z_{N-1}, \ldots, z_1) \in \mathbb{R}_+^N} \quad & \textstyle\sum_{n=1}^N c_{s_n} \cdot z_n \\ \text{s.t.} \quad & \textstyle\sum_{n=1}^N z_n \geq d_{\text{total}} \\ \text{and} \quad & z_n \leq \frac{P}{c_{s_n}} \ , \ \forall n \in \{1, 2, \ldots, N\} \, . \end{aligned} \tag{6.15}$$

It should be clear that (6.15) is essentially the same problem as (4.2), and the solution can be found by scheduling data transmission during the slot with the best condition until all the data is sent or the power limit is reached, and then scheduling data transmission during the slot with the second best condition until all the data is sent or the power limit is reached, and so forth. See Figure 6.2 for a diagram of this solution.

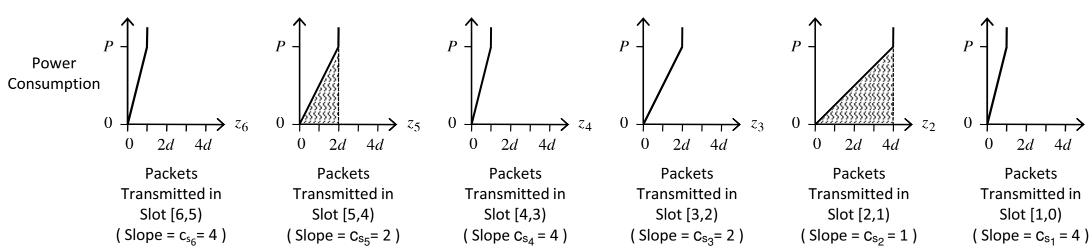


Figure 6.2. Pictorial representation of the solution to Problem (**P6.2**) in the somewhat unrealistic case that all future channel conditions are known at the beginning of the horizon. Packets are scheduled in slots in ascending order of $c_{s_n}$, until all the data is transmitted or the power constraint for the slot is reached. In the example shown, the time horizon to send the data is $N = 6$, the total number of data packets to be sent is $6d$, and the power constraint in each slot is $P = 4d$. One optimal policy is to transmit $4d$ packets in slot 2, which has the best channel condition, and the remaining $2d$ packets in slot 5, which has the second best channel condition. This policy results in a total cost of $2P$.

If we are focused on finding the optimal amount to transmit in the current slot, we can also aggregate the power-rate functions of all future slots, by reordering them according to the strength of the channel, as shown in Figure 6.3. The aggregate power-rate curve shown is defined by:

$$
\begin{aligned}
\hat{c}_{N-1}(\hat{z}, s_{N-1}, s_{N-2}, \ldots, s_1) := \quad & \min_{(z_{N-1},\ldots,z_1)\in \mathbb{R}_+^N} && \textstyle\sum_{n=1}^{N-1} c_{s_n} \cdot z_n \\
& \text{s.t.} && \textstyle\sum_{n=1}^{N-1} z_n = \hat{z} \\
& \text{and} && z_n \leq \frac{P}{c_{s_n}} \ , \ \forall n \in \{1, 2, \ldots, N-1\} \, ,
\end{aligned}
$$

where $\hat{z}$ is the aggregate number of packets to be transmitted in slots $N-1, N-2, \ldots, 1$. The optimal number of packets to transmit in the current slot is then deter-

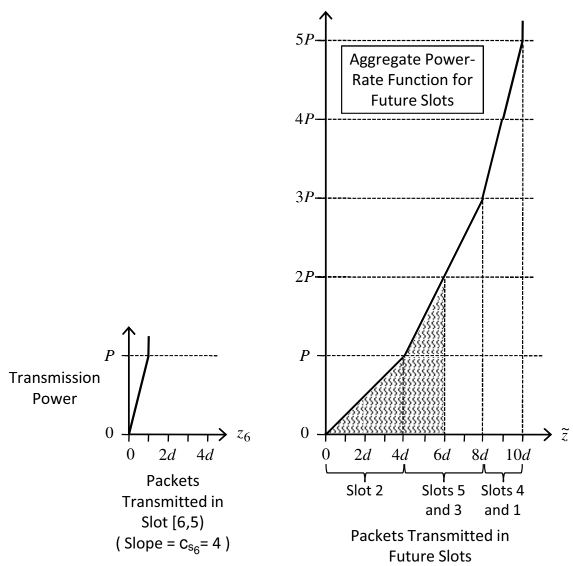


Figure 6.3. The aggregate power-rate function for future slots. Aggregating the power-rate functions of the final 5 slots from Figure 6.2 allows us to determine the optimal number of packets to transmit in the current slot by comparing the current slope to the slopes of the aggregate curve. In this case, the slope of the current curve is greater than the slope of the aggregate curve at all points up to $d_{\text{total}} = 6d$, so it is optimal to not transmit any packets in the current slot.

mined as follows. Define $\gamma_N := \min\left\{\hat{z}_0 : \hat{\psi}_{N-1}(\hat{z}) \geq c_{s_N} \ , \ \forall \hat{z} > \hat{z}_0\right\}$, where $\hat{\psi}_{N-1}(\cdot)$ is the slope from above of the aggregate power-rate curve, $\hat{c}_{N-1}(\cdot, s_{N-1}, s_{N-2}, \ldots, s_1)$, shown in Figure 6.3. Then the optimal number of packets to transmit in slot $N$ is given by:

$$z_N^* = \min \left\{ \frac{P}{c_{s_N}}, \max \left( d_{\text{total}} - \gamma_N, 0 \right) \right\} . \tag{6.16}$$

This policy says that if the current per packet energy cost from transmission is greater than the slope of the aggregate curve at all points up to $d_{\text{total}}$, then it is optimal to not transmit any packets in the current slot. Otherwise, the optimal number of packets to transmit in the current slot $N$ is the minimum of the maximum number of packets that can be transmitted under the current channel condition, and the number of packets that would otherwise be transmitted in worse channel conditions in future slots.

Now, as Fu, Modiano, and Tsitsiklis explain in [51, Section III-D] and [52, Section III-D], in the more realistic case that the channel condition in slot $n$ is not learned until the beginning of the $n^{th}$ slot, a very similar aggregate method can be used as long as the number of possible channel conditions is finite, and the condition (6.10) is satisfied. In this situation, however, the slopes of the piecewise-linear aggregate power-rate function for future slots are not defined in terms of the actual channel conditions of future slots (which are not available), but rather by a series of thresholds that only depend on the statistics of future channel conditions. Condition (6.10) ensures that the slopes of this aggregate *expected* power-rate curve only change at integer multiples of $\frac{P}{c_{s_{\text{worst}}}}$. The form of the optimal policy at time $N$ is the same as (6.16), with $d_{\text{total}}$ being the number of packets remaining to transmit at time $N$. Because the slopes of the aggregate *expected* power-rate curve only change at integer multiples of $\frac{P}{c_{s_{\text{worst}}}}$, we have $\gamma_N \in \left\{0, \frac{P}{c_{s_{\text{worst}}}}, 2 \cdot \frac{P}{c_{s_{\text{worst}}}}, \ldots, (N-1) \cdot \frac{P}{c_{s_{\text{worst}}}}\right\}$.

We now return to the wireless streaming model considered in (**P6.1**), with $d$ packets removed from the receiver's buffer at the end of every slot. Let us once again begin by considering the unrealistic case that the scheduler knows all future channel conditions at the beginning of the horizon. The optimal solution can be found by using the same basic inverse water-filling type principle of transmitting as much as possible in the slot with the best channel condition, and then the second best, and so forth; however, due to the additional underflow constraints, one needs to solve $N$ sequential problems of this form. The first problem is the trivial problem of sending $d$ packets in the first slot, $[N, N-1)$. The second problem is to send $2d$ packets in the first two slots. If the power limit in the first slot has not been reached after allocating the initial $d$ packets there, then the scheduler may choose to send the second batch of $d$ packets in either the first or second slot, according to their respective channel conditions. For each sequential problem, whatever packets have been allocated in the previous problem must be "carried over" to the subsequent problem, where there is one additional time slot available and the next $d$ packets are allocated. The solution to the $N^{th}$ problem represents the optimal allocation. See Figure 6.4 for a diagram of this solution. Comparing Figure 6.4 to Figure 6.2, we see that when $N \cdot d = d_{\text{total}}$ and the known sequence of channel conditions is the same for both problems, the additional underflow constraints cause more data to be scheduled in earlier time slots with worse channel conditions.

When all future channel conditions are known ahead of time, as in Figure 6.4, we can also use the same aggregation technique from above to represent Problems 2 through $N$ as comparisons between the current channel condition and the aggregate of the future channel conditions. Furthermore, when the future channel conditions are not known ahead of time and the condition (5.10) is satisfied, we can once again define the aggregate expected power-rate function for future slots in terms of a series of thresholds that only depend on the statistics of future channel conditions. Due to the underflow constraints, however, these thresholds are computed differently than

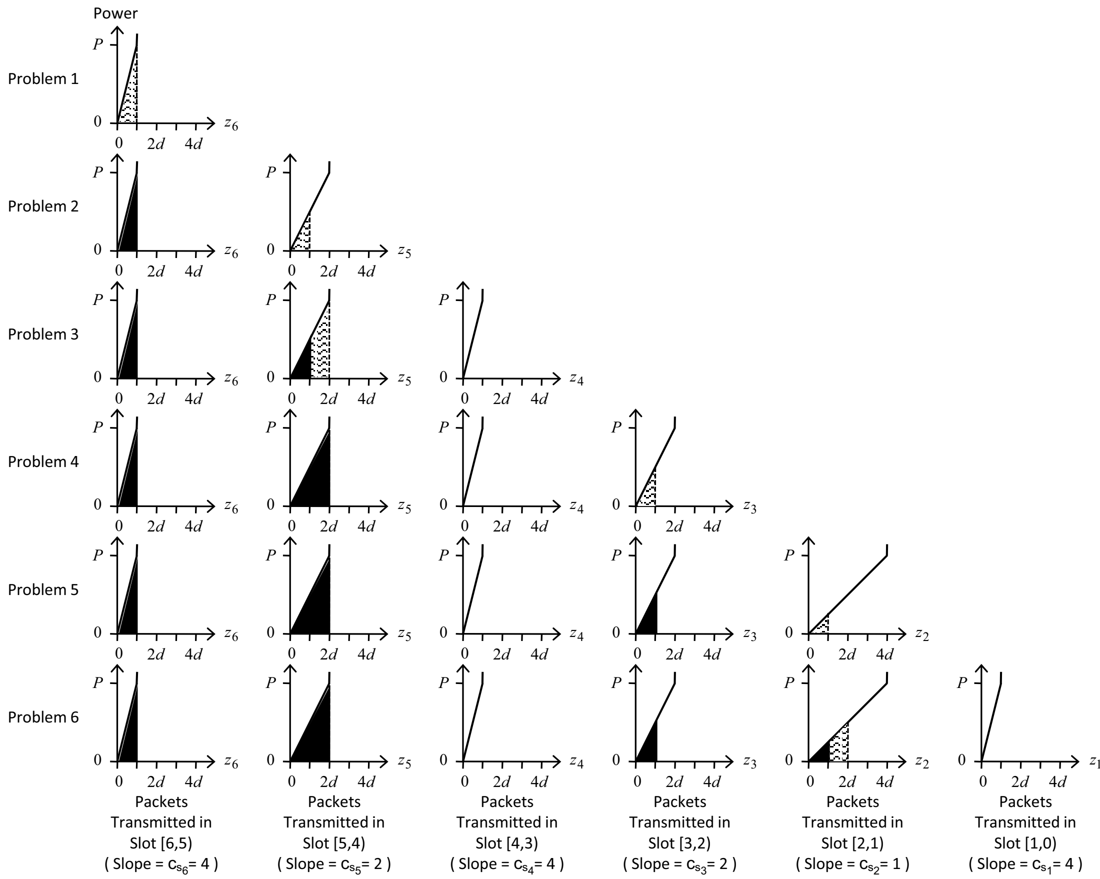


Figure 6.4. Pictorial representation of the solution to Problem (**P6.1**) in the somewhat unrealistic case that all future channel conditions are known at the beginning of the horizon. In the example shown, the time horizon is $N = 6$, $d$ packets are removed from the receiver's buffer at the end of every slot, and the power constraint in each slot is $P = 4d$. To satisfy the underflow constraints, 6 sequential problems are considered, with an additional $d$ packets allocated in each problem. Packets allocated in one problem are "carried over" to all subsequent problems, and shown in solid black filling. The optimal policy, given by the solution to Problem 6, is to transmit $d$ packets in slots 6 and 3, and $2d$ packets in slots 5 and 2. This policy results in a total cost of $3P$.

those in Problem (**P6.2**). The net result for this more realistic case is the same as the case when all future channel conditions are known - the additional underflow constraints make it optimal to send more data in earlier time slots with worse channel conditions.

## 6.4 Summary Takeaways on the Role of Deadline Constraints

As mentioned earlier, the main idea of opportunistic scheduling is to reduce energy consumption by sending more data when the channel is in a "good" state, and less data when the channel is in a "bad" state. However, deadline constraints may force the sender to transmit data when the channel is in a relatively poor state. One strategy when faced with such deadline constraints would be to deal with them as they come, by always sending just enough packets when the channel is "bad" to ensure the deadline can be met, and holding out for the best channel conditions to send a lot of data. Yet, a key conclusion from the analysis of the three problems we presented in this chapter is that it is better to anticipate the need to comply with these constraints in future slots by sending more packets (than one would without the deadlines) under "medium" channel conditions in earlier slots. In some sense, doing so is a way to manage the risk of being stuck sending a large amount of data over a poor channel to meet an imminent deadline constraint. We also saw that the extent to which the scheduler should plan for the deadline by sending data under such "medium" channel conditions depends on the time remaining until the deadline(s), and on how many deadlines it must meet. Namely, the closer the deadlines and the more deadlines it faces, the less opportunistic the scheduler can afford to be. So perhaps the essence of opportunistic scheduling with deadline constraints is that the scheduler should be opportunistic, but not too opportunistic.

## 6.5 Other Energy-Minimizing Transmission Scheduling Studies Featuring Strict Deadline Constraints

In addition to the three problems discussed above and the more general variant of Problem (**P6.1**) discussed in Chapter 5, there have been a few other studies of energy-minimizing transmission scheduling that feature a time-varying wireless channel and strict deadline constraints. In [90], [91], and [93], Lee and Jindal consider the same setup as Problem (**P6.3**), except that the convex power-rate curves are of the form $c(z,s) = \frac{2^z-1}{\alpha_s}$ or $\frac{e^z-1}{\alpha_s}$, which are based on the Gaussian noise channel capacity. The earlier models of Zafer and Modiano in [169] and [170] also include essentially the same setup as Problem (**P6.3**), with the exception that the underlying time scale is continuous rather than discrete. Using continuous-time stochastic control theory, they also reach the key conclusion that the optimal number of packets to transmit under convex monomial power-rate curves is the product of the number of packets remaining to be sent and an "urgency" fraction that depends on the current channel condition and the time remaining until the end of the horizon. Chen, Mitra, and Neely [33, 34] and Uysal-Biyikoglu and El Gamal [161] consider packets arriving at different times, analyze offline scheduling problems, and use the properties of the optimal offline scheduler to develop heuristics for online (or causal) scheduling problems. In [155], Tarello et al. extend Problem (**P6.2**) (without the power constraint) to the case of multiple identical receivers, and assume that the source can only transmit to one user in each slot. An overview of the models considered in each of these studies is provided in Table 6.1. Additionally, Luna et al. [102] consider an energy minimization problem subject to end-to-end delay constraints, where the scheduler

| Study | Num. of Receivers | Data | Deadline Constraints | Scheduler's Information | Power-Rate Curves |
|---|---|---|---|---|---|
| Fu, Modiano, and Tsitsiklis [51, 52] | 1 | Infinite backlog | Single deadline | Non-causal; causal | Convex; linear with power constraint |
| Shuman, Liu, and Wu [140, 144, 145] | Multiple (focus on 1, 2) | Infinite backlog | Multiple deadlines (underflow constraints) | Causal | Linear and piecewise-linear convex with power constraint |
| Lee and Jindal [92] | 1 | Infinite backlog | Single deadline | Causal | Convex monomial |
| Lee and Jindal [90, 91, 93] | 1 | Infinite backlog | Single deadline | Causal | Convex (Gaussian noise channel capacity) |
| Zafer and Modiano [169, 170] | 1 | Infinite backlog; random packet arrivals | Single deadline; multiple variable deadlines | Causal | Convex; convex monomial |
| Chen, Mitra, and Neely [33, 34] | 1 | Packet arrivals | Individual packet deadlines | Non-causal; causal | Convex |
| Uysal-Biyikoglu and El Gamal [161] | Multiple (focus on 2) | Packet arrivals | Single deadline | Non-causal; causal | Convex |
| Tarello et al. [155] | Multiple | Infinite backlog | Single deadline | Non-causal; causal | Linear; convex |

Table 6.1. Energy-minimizing transmission scheduling models that feature a time-varying wireless channel and strict deadline constraints. We use the term "infinite backlog" to include problems where there are a finite number of packets to be sent, all of which are queued at the beginning of the time horizon. "Non-causal" refers to the offline scheduling situation where the transmission scheduler has knowledge of future channel states and packet arrival times.

must select various source coding parameters in addition to the transmission powers. Finally, there is a sizeable literature on energy-efficient transmission scheduling studies such as [162] that feature a *time-invariant* or *static* channel and strict deadline constraints.

# *Chapter 7*

# Stochastic Versus Deterministic Prices in Inventory Theory

The large majority of models in the classical inventory literature consider *deterministic, time-invariant* prices and *stochastic* demands, the reverse of the model we introduced in Chapter 5. In this chapter, we consider a variant of the two-item resource-constrained inventory problem with deterministic, time-invariant prices and stochastic demands originally studied by Evans in [45], and revisited in [32], [40], and [79]. Our purpose in discussing this model is to compare the qualitative properties of the optimal policy to those of Problem (**P5.1**). Specifically, the question at hand is whether models with stochastic prices deserve their own analysis or if the qualitative behavior follows in a straightforward manner from analysis of models with deterministic prices. The main thesis of the chapter is that inventory models with stochastic prices do indeed merit their own line of analysis as structural phenomena that cannot appear in the corresponding models with deterministic prices are liable to appear in the stochastic price inventory models.

## 7.1 Problem Formulation

There are two main differences between the problem considered in this chapter and Evans' problem in [45]:

(1) In the case of stochastic demand, it is less realistic to impose strict constraints on the fulfillment of demand. The models most often considered in inventory theory are (i) the complete backlogging model, where unsatisfied demand is backlogged in order to be satisfied in future periods, with some penalty for the delay; and (ii) the lost sales model, where unsatisfied demand incurs a one-time penalty and is not backlogged. We consider complete backlogging of unfilled demand rather than the lost sales model considered in [45]. Evans points out that the qualitative results are the same under both models.

(2) We allow the random vector of demands, $\mathbf{D}$, to have any distribution. Evans requires that it have continuous support, in order to ensure differentiability and strict convexity of the value functions. In the words of Milgrom and Shannon [105], "the *only* role these assumptions play is as servants to a method." By relaxing the continuous support assumption, we lose these properties, but we will show that the fundamental structure of the optimal policy is unaffected.

We consider a two-item inventory model where the total ordering cost in each period cannot exceed a joint budget, $P$. The ordering costs for each item are linear, with the deterministic, time-invariant vector of ordering prices given by $\mathbf{c}$. The vector of inventories in period $n$ is given by $\mathbf{X}_n$, and the vector of controlled order quantities is denoted by $\mathbf{Z}_n$. The demands for each item are stochastic, and represented by the random vector $\mathbf{D}_n$ in period $n$. We assume the vector of demands is IID across time. Unmet demands are completely backlogged until future slots (i.e., $\mathbf{X}$ can take on negative values), so the system dynamics are given by:

$$\mathbf{X}_{n-1} = \mathbf{X}_n + \mathbf{Z}_n - \mathbf{D}_n \ .$$

The total shortage and holding costs at the end of each period are given by:

$$l(\mathbf{x}) := l^1(x^1) + l^2(x^2) \ ,$$

where $l^j(0) = 0$ and $l^j(\cdot)$ is convex, nondecreasing above 0, and nonincreasing below 0, for item $j = 1, 2$. We consider the finite horizon discounted expected cost problem with horizon $N$. Thus, the stochastic dynamic optimization problem, which we refer to as Problem (**P7.1**), is given by:

$$\begin{array}{lll} & \inf\limits_{\pi \in \Pi} & \mathbb{E}^{\pi}\left\{ \sum\limits_{n=1}^{N} \alpha^{N-n} \cdot \left\{ \mathbf{c}^{\mathrm{T}} \mathbf{Z}_n + l(\mathbf{X}_n + \mathbf{Z}_n - \mathbf{D}_n) \right\} \mid \mathcal{F}_N \right\} \\ \text{s.t.} & & \mathbf{c}^{\mathrm{T}} \mathbf{Z}_n \le P, \ w.p.1, \ \forall n \\ \text{and} & & \mathbf{Z}_n \succeq \mathbf{0}, \ w.p.1, \ \forall n \ , \end{array}$$

where $0 \le \alpha \le 1$ is the discount factor and $\mathcal{F}_N$ denotes all information available at the beginning of the time horizon.

Using the normal change of variable $\mathbf{Y}_n = \mathbf{X}_n + \mathbf{Z}_n$, the dynamic program for Problem (**P7.1**) is given by:

$$\begin{aligned} V_n(\mathbf{x}) &= \min_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \left\{ \mathbf{c}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}) + \mathbb{E}\big[l(\mathbf{y} - \mathbf{D})\big] + \alpha \cdot \mathbb{E}\big[V_{n-1}(\mathbf{y} - \mathbf{D})\big] \right\} \\ &= -\mathbf{c}^{\mathrm{T}}\mathbf{x} + \min_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \left\{ \mathbf{c}^{\mathrm{T}}\mathbf{y} + \mathbb{E}\big[l(\mathbf{y} - \mathbf{D})\big] + \alpha \cdot \mathbb{E}\big[V_{n-1}(\mathbf{y} - \mathbf{D})\big] \right\} \\ &= -\mathbf{c}^{\mathrm{T}}\mathbf{x} + \min_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \left\{ \hat{G}_n(\mathbf{y}) \right\}, \ n = N, N-1, \ldots, 1, \\ V_0(\mathbf{x}) &= 0, \ \forall \mathbf{x}, \end{aligned}$$

where $\hat{G}_n(\mathbf{y}) := \mathbf{c}^{\mathrm{T}}\mathbf{y} + \mathbb{E}\big[l(\mathbf{y} - \mathbf{D})\big] + \alpha \cdot \mathbb{E}\big[V_{n-1}(\mathbf{y} - \mathbf{D})\big]$, and the action space is defined as:

$$\hat{\mathcal{A}}(\mathbf{x}) := \left\{ \mathbf{y} \in \mathbb{R}^2 : \mathbf{x} \preceq \mathbf{y} \text{ and } \mathbf{c}^{\mathrm{T}}(\mathbf{y} - \mathbf{x}) \le P \right\} .$$

## 7.2 The Direct Value Order

Before proceeding to the optimal policy structure for Problem (**P7.1**), we discuss a partial order introduced by Antoniadou in [8, 11] called the *direct value order.*

**Definition 7.1** (Antoniadou, 1996)**.** Let $\mathbf{c} \in I\!R_{++}$ and let $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in I\!R^2$. Then the direct $(\mathbf{c}, i)$ value order, $\leq_{dv(\mathbf{c},i)}$ for $i = 1, 2$, is defined by:

$$\bar{\mathbf{x}} \leq_{dv(\mathbf{c},i)} \tilde{\mathbf{x}} \quad \text{if and only if} \quad \mathbf{c}^{\mathrm{T}}\bar{\mathbf{x}} \leq \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{x}} \text{ and } \bar{x}^i \leq \tilde{x}^i \ .$$

Recall that a function $f : I\!R^2 \to I\!R$ is supermodular (submodular) with respect to a given partial order if for all $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in I\!R^2$:

$$f(\bar{\mathbf{x}}) + f(\tilde{\mathbf{x}}) \leq (\geq)\ f(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) + f(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}) \ ,$$

where $\wedge$ and $\vee$ denote the meet and join with respect to the given partial order, as defined in Section 1.4. The meet and join of two points with respect to the usual Euclidean, $dv(\mathbf{c}, 1)$, and $dv(\mathbf{c}, 2)$ partial orders are shown in Figure 7.1.

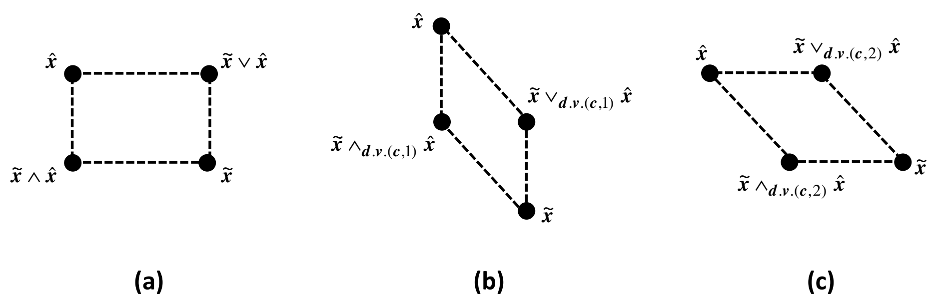


Figure 7.1. The direct value order. (a) shows the standard Euclidean order; (b) shows the $dv(\mathbf{c}, 1)$ order; and (c) shows the $dv(\mathbf{c}, 2)$ order.

If $f : I\!R^2 \to I\!R$ is twice continuously differentiable on $I\!R^2$, then $f$ is supermodular (submodular) with respect to the usual Euclidean order if and only if [158, p. 310]:

$$\frac{\partial^2 f}{\partial x^1 \partial x^2} \geq (\leq)\ 0 \ .$$

The following proposition is analogous for the direct $(\mathbf{c}, i)$ value order.

**Proposition 7.2.** *If $f : I\!R^2 \to I\!R$ is twice continuously differentiable on $I\!R^2$, then $f$ is supermodular (submodular) with respect to the direct $(\boldsymbol{c}, 1)$ value order if and only if:*

$$\frac{\partial^2 f}{\partial x^2 \partial x^2} \leq (\geq)\ \frac{c^2}{c^1} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^2}\ , \tag{7.1}$$

*and is supermodular (submodular) with respect to the direct $(\boldsymbol{c}, 2)$ value order if and only if:*

$$\frac{\partial^2 f}{\partial x^1 \partial x^1} \leq (\geq)\ \frac{c^1}{c^2} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^2}\ . \tag{7.2}$$

A detailed proof of Proposition 7.2 is included in Appendix D.

In [45], Evans shows inductively that the functions $\{\hat{G}_n(\cdot)\}_{n\in\{N,N-1,\ldots,1\}}$ defined in the previous section satisfy (7.1) and (7.2) with a strict $>$ inequality. He refers to these conditions as "dominance of the second partials over the mixed partials." Without the assumption that the random demand has continuous support, the functions $\{\hat{G}_n(\cdot)\}_{n\in\{N,N-1,\ldots,1\}}$ are not necessarily twice differentiable. The approach in the next section is to instead show that the functions $\{\hat{G}_n(\cdot)\}_{n\in\{N,N-1,\ldots,1\}}$ are submodular with respect to the $dv(\mathbf{c}, 1)$ and $dv(\mathbf{c}, 2)$ partial orders, and that the same structural features of the optimal policy follow.

In [32], Chen defines a function $f : I\!R^2 \to I\!R$ to be $\boldsymbol{\mu}$-*difference monotone* if for any $t > 0$, the function

$$h(x^1, x^2) := f(x^1 + \mu^1 \cdot t, x^2) - f(x^1, x^2 + \mu^2 \cdot t) \qquad \uparrow x^1,\ \downarrow x^2\ ,$$

where $\uparrow x^1$ means nondecreasing in $x^1$ and $\downarrow x^2$ means nonincreasing in $x^2$. Note that a function is $\boldsymbol{\mu}$-difference monotone if and only if it is submodular with respect to both the $dv(\mathbf{c}, 1)$ and $dv(\mathbf{c}, 2)$ partial orders, where $c^i = \frac{1}{\mu^i}$ for $i = 1, 2$.

To the best of our knowledge, the analysis of Problem (**P7.1**) in [32] and this chapter represents the first time the direct value order has been used to inductively show functional properties in a dynamic optimization problem, as most of its applications to date have been in comparative statics [8, 9, 10, 11, 106, 107, 119].

Finally, we remark that that the triangular action spaces of Problems (**P5.1**) and (**P7.1**) are lattices with respect to both the both the $dv(\mathbf{c}, 1)$ and $dv(\mathbf{c}, 2)$ partial orders.

## 7.3 Structure of Optimal Policy

As in Section 5.5.1, we prove properties of the value functions, and then deduce the structure of the optimal policy from these properties. Note that Problem (**P7.1**) is essentially the same problem as the one considered by Chen in [32]. However, we present our versions of theorems describing the properties of the value functions and structure of the optimal policy (which were derived independently before learning of [32] and have different proofs) in order to maintain comparability with the corresponding statements about Problem (**P5.1**).

**Theorem 7.3.** *The following statements are true for $n = 1, 2, \ldots, N$:*

**(i)** $V_{n-1}(\boldsymbol{x})$ *is convex in* $\boldsymbol{x}$.

**(ii)** $V_{n-1}(\boldsymbol{x})$ *is supermodular in* $\boldsymbol{x}$*; i.e., for all* $\bar{\boldsymbol{x}}, \tilde{\boldsymbol{x}} \in I\!R^2$,

$$V_{n-1}(\bar{\boldsymbol{x}}) + V_{n-1}(\tilde{\boldsymbol{x}}) \le V_{n-1}(\bar{\boldsymbol{x}} \wedge \tilde{\boldsymbol{x}}) + V_{n-1}(\bar{\boldsymbol{x}} \vee \tilde{\boldsymbol{x}}) \ .$$

**(iii)** $V_{n-1}(\boldsymbol{x})$ *is* $dv(\boldsymbol{c}, 1)$*-submodular in* $\boldsymbol{x}$*, and* $dv(\boldsymbol{c}, 2)$*-submodular in* $\boldsymbol{x}$*; i.e., for all* $\bar{\boldsymbol{x}}, \tilde{\boldsymbol{x}} \in I\!R^2$,

$$V_{n-1}(\bar{\boldsymbol{x}}) + V_{n-1}(\tilde{\boldsymbol{x}}) \ge V_{n-1}(\bar{\boldsymbol{x}} \wedge_{dv(\boldsymbol{c},i)} \tilde{\boldsymbol{x}}) + V_{n-1}(\bar{\boldsymbol{x}} \vee_{dv(\boldsymbol{c},i)} \tilde{\boldsymbol{x}}) \ , \ i = 1, 2 \ .$$

**(iv)** $\hat{G}_n(\boldsymbol{y})$ *is convex in* $\boldsymbol{y}$.

**(v)** $\hat{G}_n(\boldsymbol{y})$ *is supermodular in* $\boldsymbol{y}$*; i.e., for all* $\bar{\boldsymbol{y}}, \tilde{\boldsymbol{y}} \in I\!R^2$,

$$\hat{G}_n(\bar{\boldsymbol{y}}) + \hat{G}_n(\tilde{\boldsymbol{y}}) \leq \hat{G}_n(\bar{\boldsymbol{y}} \wedge \tilde{\boldsymbol{y}}) + \hat{G}_n(\bar{\boldsymbol{y}} \vee \tilde{\boldsymbol{y}}) \ .$$

**(vi)** $y_n^1 < \hat{y}_n^1$ *implies:*

$$\inf\left\{\operatorname*{argmin}_{y_n^2 \in I\!R}\left\{\hat{G}_n\left(y_n^1, y_n^2\right)\right\}\right\} \geq \inf\left\{\operatorname*{argmin}_{y_n^2 \in I\!R}\left\{\hat{G}_n\left(\hat{y}_n^1, y_n^2\right)\right\}\right\}$$

*and* $y_n^2 < \hat{y}_n^2$ *implies:*

$$\inf\left\{\operatorname*{argmin}_{y_n^1 \in I\!R}\left\{\hat{G}_n\left(y_n^1, y_n^2\right)\right\}\right\} \geq \inf\left\{\operatorname*{argmin}_{y_n^1 \in I\!R}\left\{\hat{G}_n\left(y_n^1, \hat{y}_n^2\right)\right\}\right\} .$$

**(vii)** $\hat{G}_n(\boldsymbol{y})$ *is* $dv(\boldsymbol{c}, 1)$*-submodular in* $\boldsymbol{y}$*, and* $dv(\boldsymbol{c}, 2)$*-submodular in* $\boldsymbol{y}$*; i.e., for all* $\bar{\boldsymbol{y}}, \tilde{\boldsymbol{y}} \in I\!R^2$,

$$\hat{G}_n(\bar{\boldsymbol{y}}) + \hat{G}_n(\tilde{\boldsymbol{y}}) \geq \hat{G}_n(\bar{\boldsymbol{y}} \wedge_{dv(\boldsymbol{c},i)} \tilde{\boldsymbol{y}}) + \hat{G}_n(\bar{\boldsymbol{y}} \vee_{dv(\boldsymbol{c},i)} \tilde{\boldsymbol{y}}) \ , \ i = 1, 2 \ .$$

A detailed proof is included in Appendix E.1. Similar to (ii), because $-\mathbf{c}^{\mathrm{T}}\mathbf{x}$ is $dv(\mathbf{c}, i)$-submodular in $\mathbf{x}$, the key part of the induction step in the proof of (iii) is to show that $\min_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \left\{\hat{G}_{n-1}(\mathbf{y})\right\}$ is also $dv(\mathbf{c}, i)$-submodular in $\mathbf{x}$. Denoting $\operatorname{argmin}_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \left\{\hat{G}_{n-1}(\mathbf{y})\right\}$ by $\mathbf{y}^*(\mathbf{x})$, we do this constructively by showing that for all $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in I\!R^2$:

$$\begin{aligned} &\min_{\mathbf{y} \in \hat{\mathcal{A}}(\bar{\mathbf{x}})} \left\{\hat{G}_{n-1}(\mathbf{y})\right\} + \min_{\mathbf{y} \in \hat{\mathcal{A}}(\tilde{\mathbf{x}})} \left\{\hat{G}_{n-1}(\mathbf{y})\right\} \\ =\ & \hat{G}_{n-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}})\Big) + \hat{G}_{n-1}\Big(\mathbf{y}^*(\tilde{\mathbf{x}})\Big) \\ \geq\ & \hat{G}_{n-1}(\hat{\mathbf{y}}) + \hat{G}_{n-1}(\check{\mathbf{y}}) \\ \geq\ & \min_{\mathbf{y} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},i)} \tilde{\mathbf{x}}\right)} \left\{\hat{G}_{n-1}(\mathbf{y})\right\} + \min_{\mathbf{y} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},i)} \tilde{\mathbf{x}}\right)} \left\{\hat{G}_{n-1}(\mathbf{y})\right\}, \end{aligned} \tag{7.3}$$

for a specific choice of $\hat{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},i)} \tilde{\mathbf{x}}\right)$ and $\check{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},i)} \tilde{\mathbf{x}}\right)$. The difficulty is cleverly constructing $\hat{\mathbf{y}}$ and $\check{\mathbf{y}}$, depending on the relative locations of $\bar{\mathbf{x}}$, $\tilde{\mathbf{x}}$, $\mathbf{y}^*(\bar{\mathbf{x}})$, and $\mathbf{y}^*(\tilde{\mathbf{x}})$, so as to ensure (7.3) is true.

It follows from Theorem 7.3 that the structure of the optimal transmission policy for Problem (**P7.1**) is given by the following theorem.

**Theorem 7.4.** *For every $n \in \{1, 2, \ldots, N\}$, define the non-empty set of global minimizers of $\hat{G}_n(\cdot, \boldsymbol{s})$:*

$$\hat{\mathcal{B}}_n := \left\{ \hat{\boldsymbol{y}} \in \mathbb{R}^2 : \; \hat{G}_n(\hat{\boldsymbol{y}}) = \min_{\boldsymbol{y} \in \mathbb{R}^2} \hat{G}_n(\boldsymbol{y}) \right\} \; .$$

*Define also*

$$\hat{b}_n^1 := \min \left\{ y^1 \in \mathbb{R} : (y^1, y^2) \in \hat{\mathcal{B}}_n \text{ for some } y^2 \in \mathbb{R} \right\} \; ,$$

*and*

$$\hat{b}_n^2 := \min \left\{ y^2 \in \mathbb{R} : \left( \hat{b}_n^1, y^2 \right) \in \hat{\mathcal{B}}_n \right\} \; .$$

*Then the vector $\hat{\boldsymbol{b}}_n = \left( \hat{b}_n^1, \hat{b}_n^2 \right) \in \hat{\mathcal{B}}_n$ is a global minimizer of $\hat{G}_n(\cdot)$. Define also the functions:*

$$\hat{f}_n^1(x^2) := \inf \left\{ \underset{y^1 \in \mathbb{R}}{\operatorname{argmin}} \left\{ \hat{G}_n \left( y^1, x^2 \right) \right\} \right\}, \text{ for } x^2 \in \mathbb{R}, \text{ and}$$

$$\hat{f}_n^2(x^1) := \inf \left\{ \underset{y^2 \in \mathbb{R}}{\operatorname{argmin}} \left\{ \hat{G}_n \left( x^1, y^2 \right) \right\} \right\}, \text{ for } x^1 \in \mathbb{R}.$$

*Note that by construction, $\hat{f}_n^1(\hat{b}_n^2) = \hat{b}_n^1$ and $\hat{f}_n^2(\hat{b}_n^1) = \hat{b}_n^2$. Partition $\mathbb{R}^2$ into the following seven regions:*

$$\hat{\mathcal{R}}_I(n) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \succeq \left( \hat{f}_n^1(x^2), \hat{f}_n^2(x^1) \right) \text{ and } \boldsymbol{x} \neq \hat{\boldsymbol{b}}_n \right\}$$

$$\hat{\mathcal{R}}_{II}(n) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : \boldsymbol{x} \preceq \hat{\boldsymbol{b}}_n \text{ and } \boldsymbol{c}^{\mathrm{T}} \left[ \hat{\boldsymbol{b}}_n - \boldsymbol{x} \right] \leq P \right\}$$

$$\hat{\mathcal{R}}_{III-A}(n) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x^2 > \hat{b}_n^2 \text{ and } \hat{f}_n^1(x^2) - \frac{P}{c^1} \leq x^1 < \hat{f}_n^1(x^2) \right\}$$

$$\hat{\mathcal{R}}_{III-B}(n) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x^1 > \hat{b}_n^1 \text{ and } \hat{f}_n^2(x^1) - \frac{P}{c^2} \leq x^2 < \hat{f}_n^2(x^1) \right\}$$

$$\hat{\mathcal{R}}_{IV-A}(n) := \left\{ \boldsymbol{x} \in \mathbb{R}^2 : x^2 > \hat{b}_n^2 \text{ and } x^1 < \hat{f}_n^1(x^2) - \frac{P}{c^1} \right\}$$

$$\hat{\mathcal{R}}_{IV-B}(n) := \left\{ \boldsymbol{x} \in I\!R^2 : \boldsymbol{x} \preceq \hat{\boldsymbol{b}}_n \text{ and } \boldsymbol{c}^{\mathrm{T}} \left[ \hat{\boldsymbol{b}}_n - \boldsymbol{x} \right] > P \right\}$$
$$\hat{\mathcal{R}}_{IV-C}(n) := \left\{ \boldsymbol{x} \in I\!R^2 : x^1 > \hat{b}_n^1 \text{ and } x^2 < \hat{f}_n^2(x^1) - \frac{P}{c^2} \right\} .$$

*Then for Problem (**P7.1**), for all $\boldsymbol{x} \notin \hat{\mathcal{R}}_{IV-B}(n)$, an optimal control action with $n$ slots remaining is given by:*

$$\boldsymbol{y}_n^*(\boldsymbol{x}) := \begin{cases} \boldsymbol{x}, & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_I(n) \\ \hat{\boldsymbol{b}}_n, & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_{II}(n) \\ \left(\hat{f}_n^1(x^2), x^2\right), & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_{III-A}(n) \\ \left(x^1, \hat{f}_n^2(x^1)\right), & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_{III-B}(n) \\ \left(x^1 + \frac{P}{c^1}, x^2\right), & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_{IV-A}(n) \\ \left(x^1, x^2 + \frac{P}{c^2}\right), & \text{if } \boldsymbol{x} \in \hat{\mathcal{R}}_{IV-B}(n) \end{cases} . \tag{7.4}$$

*For all $\boldsymbol{x} \in \hat{\mathcal{R}}_{IV-B}(n)$, there exists an optimal control action with $n$ slots remaining, $\boldsymbol{y}_n^*(\boldsymbol{x})$, which satisfies:*

$$\boldsymbol{c}^{\mathrm{T}} \left[ \boldsymbol{y}_n^*(\boldsymbol{x}) - \boldsymbol{x} \right] = P \text{ and } \boldsymbol{y}_n^*(\boldsymbol{x}) \preceq \hat{\boldsymbol{b}}_n . \tag{7.5}$$

A detailed proof of Theorem 7.4 is included in Appendix E.2.

## 7.4 Comparison of Problems (P5.1) and (P7.1)

As first glance, the structures of the optimal policies for Problems (**P5.1**) and (**P7.1**), described in Theorems 5.9 and 7.4, respectively, may seem extremely similar. However, there are two fundamental differences that distinguish these two problems.

First, the function $\hat{G}_n(\cdot)$ in Problem (**P7.1**) has the additional property of submodularity with respect to the direct value orders (statement (vii) in Theorem 7.3). This functional property leads to two additional structural results on the optimal control action: (i) when the initial vector of inventories (corresponds to the vector of receivers' buffer levels in Problem (**P5.1**)) is in region $\hat{\mathcal{R}}_{IV-B}(n)$, there exists an optimal control action such that $\mathbf{y}_n^*(\mathbf{x}) \preceq \mathbf{b}_n$;[1] and (ii) when the initial vector of inventories is in region $\hat{\mathcal{R}}_{IV-A}(n)$ (respectively, $\hat{\mathcal{R}}_{IV-C}(n)$), there exists an optimal control action that includes not ordering any of item 2 (respectively, item 1), corresponding to not transmitting any packets to user 2 (respectively, user 1) in Problem (**P5.1**). Due to the time-varying ordering prices (channel conditions), this property does not hold for the function $G_n(\cdot, \mathbf{s})$ in Problem (**P5.1**), and these two additional statements on the structure of the optimal policy are not true in general for Problem (**P5.1**), as shown by the following example.

**Example 7.5.** Consider the following instance of Problem (**P5.1**). A single sender transmits to two statistically identical receivers, whose channel conditions are IID over time and independent of each other. The power-rate curves are linear, and the possible per packet power costs are 1.750 (best possible channel condition), 2.000, 2.001, and 2.100 (worst possible channel condition). The associated probabilities of each user experiencing these channel conditions are 0.4, 0.4, 0.1, and 0.1, respectively. The total power constraint in each slot is $P = 4.2$, and 1 packet is removed from each receiver's buffer at the end of each time slot (i.e., $\mathbf{d} = (1, 1)$). We consider a finite horizon problem with the discount rate $\alpha = 1$, and no holding costs. We are interested in the optimal control action with $T = 3$ time slots remaining, and the current channel conditions are such that it costs 2.000 units of power to transmit a packet to user 1, and 2.001 units of power to transmit a packet to user 2.

Exactly solving the dynamic program shows that the unique global minimizer of the function $G_3(\cdot, \cdot, \mathbf{s}_3)$ is the vector $(\frac{101}{75}, \frac{101}{75})$. However, if the vector of starting

[1] In [32], Chen elaborates on the optimal allocation of the budget between the two items in Region $\hat{\mathcal{R}}_{IV-B}(n)$ by defining a curve splitting the region into the area where item 1 should be ordered and the area where item 2 should be ordered. He refers to this policy as a *hedging point policy.*

receiver buffer levels at time $T = 3$ is $\mathbf{x}_3 = (0.2, 0.2)$, the unique optimal scheduling decision in the slot is to transmit 0.8 packets to user 2, and use the remaining power for transmission to user 1, which results in 1.2996 packets being sent to user 1. A diagram of this optimal control action is shown in Figure 7.2. The interesting thing to note here is that despite being power-constrained (the vector of starting buffer levels is in Region $\mathcal{R}_{IV-B}$), the unique optimal scheduling decision calls for filling user 1's buffer beyond its critical number $b_3^1(\mathbf{s}_3) = \frac{101}{75}$. That is, the optimal scheduling decision brings the buffer levels from Region $\mathcal{R}_{IV-B}$ to Region $\mathcal{R}_{III-B}$ rather than Region $\mathcal{R}_{II}$.

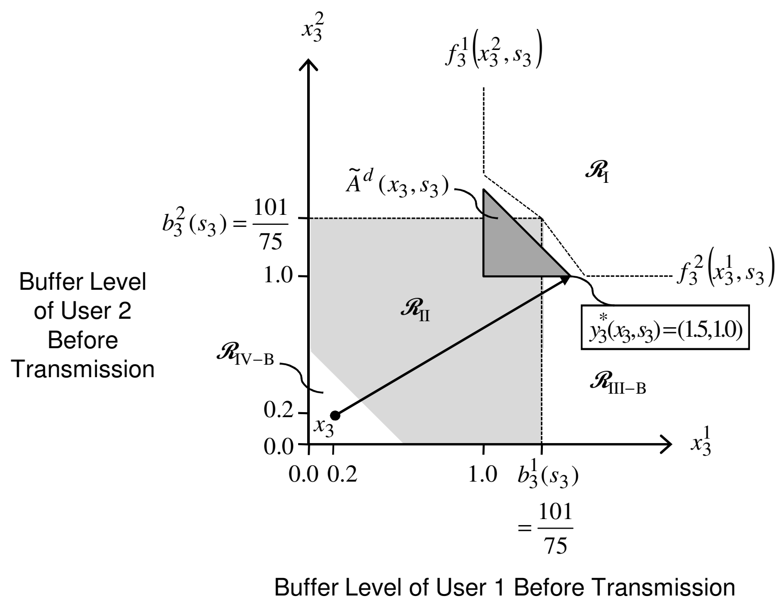


Figure 7.2. Optimal scheduling decision with 3 slots remaining in Example 7.5. The action space is represented by the triangle $\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}_3, \mathbf{s}_3)$. The critical vector $\mathbf{b}_3(\mathbf{s}_3)$ is not reachable from the starting buffer levels $\mathbf{x}_3 = (0.2, 0.2)$. The unique optimal control action is to choose $\mathbf{y}_3(\mathbf{x}_3, \mathbf{s}_3)$ (the buffer levels after transmission but before playout) to be (1.5, 1.0). The interesting feature of the example is that even though $\mathbf{x}_3 \preceq \mathbf{b}_3(\mathbf{s}_3)$, we have $\mathbf{y}_3^*(\mathbf{x}_3, \mathbf{s}_3) \npreceq \mathbf{b}_3(\mathbf{s}_3)$.

The second fundamental difference is also a consequence of the time-varying ordering prices in Problem (**P5.1**). In the infinite horizon version of Problem (**P7.1**), the critical vector $\hat{\mathbf{b}}$ is time-invariant. Combined with the above property that it is optimal to not order inventory so as to move out of regions $\mathcal{R}_{II}$ and $\mathcal{R}_{IV-B}$, the time-invariant critical vector means that the region $\mathcal{R}_{II} \cup \mathcal{R}_{IV-B}$ (i.e., the lower-left square below the critical vector) is a "stability" region.[2] Eventually, the vector of inventories enters this region under the optimal ordering policy, and once it does, it never leaves. This behavior both simplifies the analysis and opens the door for new mathematical techniques, such as analyzing *shortfall* to compute the critical numbers and determine the optimal allocation between items in the budget-constrained region [79, 157]. In Problems (**P5.2**) and (**P5.3**), even though the boundaries of the seven regions for each possible channel condition are time-invariant, no such stability region exists, because the critical numbers vary over time due to the time-varying channel conditions. Therefore, the same vector of inventories may be in region $\mathcal{R}_{II}(\mathbf{s})$ in one time slot and say $\mathcal{R}_{III-A}(\mathbf{s}')$ in the next time slot. This makes it significantly more difficult to determine optimal and near-optimal policies.

[2] A key assumption needed to ensure the stability region is that $\mathbf{c}^{\mathrm{T}}\mathbb{E}[\mathbf{D}] < P$; that is, the budget in a period suffices to purchase inventory to fulfill the aggregate average demand. Without this assumption, the infinite horizon average cost is infinite as the shortage costs accumulate [79].

# *Chapter 8*

## Conclusion

In this thesis, we studied three stochastic scheduling problems motivated by wireless communications applications. The common theme was conserving energy, first by limiting the radio's idle time and then by adjusting transmission powers. We now summarize the main results, discuss future directions, and present some final thoughts.

The first problem we considered was sleep scheduling for a wireless sensor node. We formulated infinite horizon average expected cost and finite horizon expected cost problems to model the fundamental tradeoff between delay and energy consumption. In the infinite horizon problem, we showed that it is optimal to stay awake to transmit a packet if the queue is non-empty. At the boundary state when the queue is empty, the optimal control action is given by a threshold decision rule, which is a function of the following problem parameters: the energy consumption cost of staying awake in a slot, the packet holding cost, the probability of a packet arrival in each time slot, and the length of a sleep period. For the finite horizon problem, we showed there is a shutdown period at the end of the horizon where it is optimal to sleep regardless of the number of packets in the queue. Before this shutdown period, it is optimal to stay awake if packets are present in the queue. At times far away from the end of the horizon, the optimal control action when the queue is empty converges to the optimal control action from the infinite horizon problem at the same state (as determined by the threshold decision rule). The interesting question was whether the optimal control action at this boundary state is monotonic over time. We showed through a counterexample that it is not monotonic, but conjectured that it has one of three simple structural forms.

The second problem we introduced was dynamic clock calibration via temperature measurement. The main idea of the novel approach to clock calibration was that a little extra energy consumption in taking temperature measurements during the sleep phases may lead to significant energy savings in the communication mode. The objective was to dynamically schedule a limited number of temperature measurements in a manner most useful to improving the accuracy of the ultra-low power timer. We formulated two different optimization problems, with continuous and discrete underlying time scale, respectively, and showed how both could be reduced to finite state, finite action Markov decision processes. The discrete-time formulation required extra conditions on the statistics of the ambient temperature process and possible frequencies of the clock, but resulted in a computationally simpler dynamic program.

The third problem we introduced was energy-efficient transmission scheduling with strict underflow constraints. The main idea was that the scheduler should exploit the temporal and spatial variations of the wireless channel by sending more data to a user with a good channel condition and less data to a user with a bad channel condition. In the case of a single user with linear power-rate curves, we showed that an easily-implementable modified base-stock policy is optimal under the finite horizon, infinite horizon discounted, and infinite horizon average expected cost criteria. For a single user with piecewise-linear convex power-rate curves, we showed that a finite generalized base-stock policy is optimal under all three expected cost criteria. We also presented the sequences of critical numbers that complete the characterization of the optimal control laws in each of these cases when some additional technical conditions are satisfied. In the case of two receivers, we showed how the joint power constraint couples the optimal scheduling of the two data streams. We further compared the optimal structure in the two receiver case to an analogous two-item inventory model with deterministic prices in Chapter 7, and saw that the time-varying channel conditions lead to structural phenomena that do not appear in the classical inventory models.

## 8.1 Future Directions

In this section, we discuss possible extensions to the the three main problems, which were introduced in Chapters 2, 3, and 5, respectively.

### 8.1.1 Optimal Sleep Scheduling for a Wireless Sensor Network Node

As discussed in Section 2.4.5, it may be possible to relax a number of the assumptions (e.g., general nondecreasing holding costs in place of linear holding costs, general batch arrivals in place of Bernoulli arrivals) and add fixed switching costs to the model, while still retaining the optimality of a threshold policy.

An interesting alternative formulation of the problem is to frame it as a constrained optimization problem. Specifically, rather than associate costs with packet delay and energy consumption, we could directly minimize packet delay subject to a constraint that the node must be asleep for a certain portion of the time horizon. This formulation may be more "user-friendly," as we replace the energy costs (which need to be measured and may be different in different hardware platforms) with an energy conservation constraint that has a clear physical interpretation which is consistent across hardware platforms. We have not yet considered this model, but believe analysis on this front may be tractable.

Finally, one might consider optimal sleep scheduling for multiple nodes in a wireless sensor network. This extension is not at all straightforward, as incorporating additional quality of service objectives such as connectivity or coverage drastically changes the nature of the problem.

### 8.1.2 Dynamic Clock Calibration via Temperature Measurement

Future work includes improving the numerical implementation of the solution so that it scales to higher dimensional instances of the problem. In order to do so, it may be necessary to consider suboptimal heuristics. We would also like to study the tradeoff between the number of temperature measurements allowed and expected energy savings. Specifically, by varying the limit on the number of temperature measurements allowed and solving one instance of the current problem for each limit, we could compare the marginal benefit of each additional measurement to the marginal energy cost of waking the processor up to take that measurement.

### 8.1.3 Energy-Efficient Transmission Scheduling with Strict Underflow Constraints

Our ongoing work includes examining the extension to the most general case of $M$ receivers. It is unlikely that the structure of the optimal policy in this case has a simple, intuitive, and implementable form. Therefore, our approach is to find lower bounds on the value function and a feasible policy whose expected cost is as close as possible to these bounds. One simple lower bound to the value function can be found by relaxing the per slot peak power constraint of $P$ units of total power allocated to all users, and allowing up to $P$ units of power to be allocated to *each* receiver in a single slot (for a total of up to $M \cdot P$). The advantage of this technique is that it is easy to compute the lower bound, as the $M$-dimensional problem separates into $M$ instances of the 1-dimensional problem we know how to solve from Section 5.3. However, the resulting bound is likely to be loose. A second lower bounding method we are investigating is the information relaxation method of Brown, Smith, and Sun [25]. The main idea there is to assume the scheduler has access to future channel states (corresponding to the *non-causal* or *offline* model often considered in the literature), but penalize the scheduler for using this information. A clever choice of the penalty function often leads to tight lower bounds on the value function. A third method is the Lagrangian relaxation method discussed in [1] and [65]. For our problem, this method is equivalent to relaxing the per slot peak power constraint to an average power constraint (i.e., the scheduler may allocate more than $P$ units of power in some slots, but the average power consumed per slot over the duration of the horizon cannot exceed $P$). Like the first method we mentioned, the resulting relaxed problem under this method can be separated into $M$ instances of a 1-dimensional problem, this time with an average power constraint of $\frac{P}{M}$ instead of a strict power constraint of $P$ for each receiver. A fourth lower bounding method is the linear programming approach to approximate dynamic programming discussed in [1], [38], and [132]. The idea there is to formulate the dynamic program as a linear program, and approximate the value functions as linear combinations of a set of basis functions. For a more in-depth comparison of the Lagrangian relaxation and approximate linear programming approaches, see [1]. Once lower bounds to the value function are determined from any of these methods, feasible policies can be generated based on our structural results or via one-step greedy optimization with the lower bounds substituted into the right-hand side of the dynamic programming equation.

These same numerical techniques are most likely also the best way to approximate the boundaries of the seven regions of the two receiver optimal policy, and determine a near-optimal split of the power $P$ between the two receivers when the vector of starting receiver buffer levels is in the power-constrained region $\mathcal{R}_{IV}(n, \mathbf{s})$.

## 8.2 Final Thoughts

We conclude with a brief commentary on a couple topics arising out of this thesis.

### 8.2.1 Relation Between Inventory Theory and Wireless Communications

In Chapter 5, we presented a novel connection between inventory theory and wireless communications. Namely, the uncertain fluctuating wireless channel conditions can be thought of as uncertain fluctuating prices, and the idea of opportunistically scheduling transmissions when the channel condition is good is analogous to purchasing inventory at a low price. The techniques of inventory theory certainly aided our analysis in Chapter 5, and we believe these techniques could prove to be useful in other wireless communications problems. However, the literature on inventory theory models with stochastic prices is relatively thin compared to the more classical inventory models with deterministic prices and stochastic demands. We saw in Chapter 7 that while some results for the stochastic price models follow in an expected manner from the more classical setup (for example, the modified base-stock policy discussed in Section 5.3), some stochastic price models may result in fundamentally different structural phenomena, and therefore merit their own line of analysis. So in this regard, we hope that introducing a new motivating application can continue to lead to new theoretical developments, as is often the case.

### 8.2.2 The Combination of Structural Results and Numerical Approximation Techniques

The two most common reasons to search for structural results on the optimal policy in a Markov decision process are (in no particular order): (i) to improve one's intuitive understanding of the problem; and (ii) to enable efficient computation of the optimal policy. Structural results may lead to efficient computation in a number of different ways. First, the structure may allow the optimal policy to be completely specified in closed form (as in Theorem 2.4). Second, the structure may lead to a numerical solution that is far less complex from a computational standpoint than solving the full dynamic program (as in Theorems 5.2 and 5.4). Third, one can accelerate standard methods such as value iteration and policy iteration by restricting the class of policies considered to those satisfying the known structure.

In multi-item/multi-queue stochastic control problems, there is often a significant jump in structural complexity from 1 to 2 items, and another significant jump from 2 to $M$ items. In the absence of structural results on the optimal policy in the higher dimensional problems, numerical approximation techniques such as those described in Section 8.1.3 are often used to find bounds on the value functions. Such bounds provide a benchmark against which to test suboptimal heuristics.

As seen in Section 8.1.3, one idea underlying some bounding techniques is to relax the higher dimensional problem so it decouples into multiple instances of a lower dimensional subproblem. Therefore, a third reason to search for structural results on lower dimensional problems is to indirectly improve the quality of approximate numerical solutions to related higher dimensional problems. Deriving new approximation/bounding techniques remains an extremely active area of research, so this may become an increasingly important reason to search for structural results.

# *Appendix A*

## Finite Horizon Proofs for Problem (P5.1)

### A.1 Proof of Theorem 5.1

Before proceeding to the proof of Theorem 5.1, we present a lemma due to Karush [83], which is presented in [117, pp. 237–238].

**Lemma A.1** (Karush, 1959)**.** *Suppose that $f : \mathbb{R} \to \mathbb{R}$ and that $f$ is convex on $\mathbb{R}$. For $v \leq w$, define $\tilde{f}(v,w) := \min_{z\in[v,w]} f(z)$. Then it follows that:*

**(a)** *$\tilde{f}$ can be expressed as $\tilde{f}(v,w) = F_1(v) + F_2(w)$, where $F_1$ is convex nondecreasing and $F_2$ is convex nonincreasing on $\mathbb{R}$.*

**(b)** *Suppose that $S$ is a minimizer of $f$ over $\mathbb{R}$. Then $\tilde{f}$ can be expressed as:*

$$\tilde{f}(v,w) = \begin{cases} f(v), & \text{if } S \leq v \\ f(S), & \text{if } v \leq S \leq w \\ f(w), & \text{if } w \leq S \end{cases} .$$

*Proof of Theorem 5.1:* We present the proof in three parts.

***Part I - Modified Base-Stock Structure:*** Recall the dynamic programming equation (5.4):

$$V_n(x,s) = -c_s \cdot x + \min_{\max(x,d) \le y \le x + \frac{P}{c_s}} \{g_n(y,s)\} \ , \ n = N, N-1, \ldots, 1 \ ,$$

where $g_n(y,s) := c_s \cdot y + h(y-d) + \alpha \cdot \mathbb{E}\big[V_{n-1}(y-d, S_{n-1}) \mid S_n = s\big]$. We now show by induction on $n$ that the following statements are true for every $n \in \{1, 2, \ldots, N\}$ and all $s \in \mathcal{S}$:

(i) $g_n(y,s)$ is convex in $y$ on $[d, \infty)$.

(ii) $\lim_{y \to \infty} g_n(y,s) = \infty$.

(iii) $V_n(x,s)$ is convex in $x$ on $\mathbb{R}_+$.

Base Case: $n = 1$

Let $s_1 \in \mathcal{S}$ be arbitrary. We have $g_1(y, s_1) = c_{s_1} \cdot y + h(y-d)$, which clearly satisfies (i) and (ii). $y_1^*(x, s_1) = \max(x,d)$ and thus $V_1(x, s_1) = c_{s_1} \cdot (d-x)^+ + h\Big((x-d)^+\Big)$, which is convex in $x$. We conclude (i)-(iii) are true at time $n = 1$, for all $s \in \mathcal{S}$.

Induction Step: We now assume (i)-(iii) are true for $n = m-1$ and all $s \in \mathcal{S}$, and show they hold for $n = m$ and an arbitrary $s_m \in \mathcal{S}$. Let $s_{m-1} \in \mathcal{S}$ also be arbitrary. $V_{m-1}(y-d, s_{m-1})$ is convex in $y$, so $g_m(y, s_m)$ is convex in $y$ as it is the sum of an affine function, $c_{s_m} \cdot y$, a convex function, $h(y-d)$, and a nonnegative weighted sum/integral of convex functions, $\alpha \cdot \mathbb{E}\big[V_{m-1}(y-d, S_{m-1}) \mid S_m = s_m\big]$ (see, e.g., [23, Section 3.2] for the relevant results on convexity-preserving operations). To show (ii) for $n = m$, we have $\lim\limits_{y \to \infty} g_m(y, s_m) \ge \lim\limits_{y \to \infty} c_{s_m} \cdot y = \infty$, where the inequality follows from $V_{m-1}(x, s_{m-1}) \ge 0, \forall x \in \mathbb{R}_+, \forall s_{m-1} \in \mathcal{S}$ and $h(y-d) \ge 0$. Moving on to (iii), we have:

$$V_m(x, s_m) \quad = \quad -c_{s_m} \cdot x + \min_{\max(x,d) \le y \le x + \frac{P}{c_{s_m}}} \{g_m(y, s_m)\}$$

$$= -c_{s_m} \cdot x + F_1(\max(x,d)) + F_2(x + \frac{P}{c_{s_m}}),$$

where, by Lemma A.1, $F_1$ is convex nondecreasing and $F_2$ is convex nonincreasing. $F_1(\max(x,d))$ is also convex in $x$, as it is the composition of a convex increasing function with a convex function, and $V_m(x, s_m)$ is therefore convex in $x$. This concludes the induction step, and we conclude (i)-(iii) are true for all $n \in \{1, 2, \ldots, N\}$.

Next, we define the critical numbers $b_n(s)$ for all $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$:

$$b_n(s) := \min \left\{ \hat{y} \in [d, \infty) : \; g_n(\hat{y}, s) = \min_{y \in [d,\infty)} g_n(y, s) \right\} \; .$$

Note that by properties (i) and (ii) from the above induction, the minimum of $g_n(\cdot, s)$ over $[d, \infty)$ is achieved, and the set of minimizers over $[d, \infty)$ is a non-empty closed, convex set. Thus, $b_n(s)$ is well-defined. The form of $y_n^*(x, s)$, (5.5), then follows from part (b) of Karush's result, Lemma A.1, with $g_n(y, s)$ playing the role of $f$, $\max(x, d)$ the role of $v$, $x + \frac{P}{c_s}$ the role of $w$, and $b_n(s)$ the role of $S$.

***Part II - Monotonicity of Thresholds in Time:*** In this section, we prove (5.7). We showed above that the optimal action with one time slot remaining is $y_1^*(x, s) = \max(x, d)$, for all $s \in \mathcal{S}$. This is precisely the policy suggested by (5.5) with $b_1(s) = d$, as $\frac{P}{c_s}$ is at least as great as $d$. Thus, we conclude the far right equality in (5.7) holds: $b_1(s) = d, \; \forall s \in \mathcal{S}$.

In order to show the far left inequality in (5.7), we claim more generally that $b_n(s) \leq n \cdot d$, for all $n$ and $s$. This follows from a simple interchange argument, as all packets transmitted beyond $n \cdot d$ incur transmission costs and holding costs for the duration of the horizon; however, they do not satisfy the playout requirements in any remaining slot. Thus, a policy that transmits enough packets to fill the buffer up to $n \cdot d$ at time $n$ is strictly superior to a policy that transmits more packets.

Next, we prove:

$$b_{n+1}(s) \geq b_n(s),\ \forall s \in \mathcal{S},\ \forall n \in \{1, 2, \ldots, N-1\}\ . \tag{A.1}$$

By Topkis' Theorem 2.8.1 [159, pg. 76], in order to show (A.1), it suffices to show that for all $s \in \mathcal{S}$, $n \in \{1, 2, \ldots, N-1\}$, and $y^1, y^2 \in [d, (n+1) \cdot d]$, $y^1 > y^2$ implies:

$$g_{n+1}\left(y^1, s\right) - g_n\left(y^1, s\right) \leq g_{n+1}\left(y^2, s\right) - g_n\left(y^2, s\right)\ . \tag{A.2}$$

We let $s \in \mathcal{S}$ be arbitrary, and proceed by induction on the time slot $n$.

<u>Base Case</u>: $n = 1$

For all $y \in [d, 2d]$,

$$\begin{aligned} g_2(y, s) - g_1(y, s) &= \alpha \cdot \mathbb{E}\left[V_1(y - d, S_1) \mid S_2 = s\right] \\ &= \alpha \cdot \mathbb{E}\left[c_{S_1} | S_2 = s\right] \cdot (2d - y)\ , \end{aligned}$$

which is decreasing in $y$ as $\mathbb{E}\left[c_{S_1} | S_2 = s\right] > 0$.

<u>Induction Step</u>: We assume that (A.2) is true for all $n = 1, 2, \ldots, m-1$ and $s \in \mathcal{S}$. We wish to show it is true for $n = m$. Let $y^1, y^2 \in [d, (m+1) \cdot d]$ be arbitrary, with $y^1 > y^2$. Also, let $\hat{s} \in \mathcal{S}$ be arbitrary. Define:

$$\begin{aligned} \beta_1 &:= \min\left\{ \underset{\max(y^1-d,d) \leq \hat{y} \leq y^1 - d + \frac{P}{c_{\hat{s}}}}{\operatorname{argmin}} \left\{ g_{m-1}(\hat{y}, \hat{s}) \right\} \right\} \\ \text{and}\quad \beta_2 &:= \min\left\{ \underset{\max(y^2-d,d) \leq \hat{y} \leq y^2 - d + \frac{P}{c_{\hat{s}}}}{\operatorname{argmin}} \left\{ g_m(\hat{y}, c_{\hat{s}}) \right\} \right\}\ . \end{aligned}$$

Note that:

$$\max\left(y^1 - d, d\right) \leq \beta_1 \leq \beta_1 \vee \beta_2 \leq y^1 - d + \frac{P}{c_{\hat{s}}}\ ,\ \text{and} \tag{A.3}$$

$$\max\left(y^2 - d, d\right) \leq \beta_1 \wedge \beta_2 \leq \beta_2 \leq y^2 - d + \frac{P}{c_{\hat{s}}}\ . \tag{A.4}$$

Then we have:

$$
\begin{aligned}
&\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{\hat{s}}}}\{g_m(\hat{y},\hat{s})\}-\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{\hat{s}}}}\{g_{m-1}(\hat{y},\hat{s})\}\\
&\le g_m(\beta_1\vee\beta_2,\hat{s})-g_{m-1}(\beta_1,\hat{s}) && \text{(A.5)}\\
&\le g_m(\beta_2,\hat{s})-g_{m-1}(\beta_1\wedge\beta_2,\hat{s}) && \text{(A.6)}\\
&\le \min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{\hat{s}}}}\{g_m(\hat{y},\hat{s})\}-\min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{\hat{s}}}}\{g_{m-1}(\hat{y},\hat{s})\}. && \text{(A.7)}
\end{aligned}
$$

Equation (A.5) follows from (A.3) and (A.7) follows from (A.4). If $\beta_2 \ge \beta_1$, (A.6) holds with equality. Otherwise, it follows from the induction hypothesis. Since $\hat{s}$ was arbitrary, (A.7) holds for all $\hat{s} \in \mathcal{S}$. Therefore, combined with the fact that the Markov process $\{S_n\}_{n=N,N-1,\ldots,1}$ is homogeneous, (A.7) implies:

$$
\begin{aligned}
&\mathbb{E}\left[\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{S_m}}}\{g_m(\hat{y},S_m)\}\mid S_{m+1}=s\right]\\
&-\mathbb{E}\left[\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{S_{m-1}}}}\{g_{m-1}(\hat{y},S_{m-1})\}\mid S_m=s\right]\\
\le\;&\mathbb{E}\left[\min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{S_m}}}\{g_m(\hat{y},S_m)\}\mid S_{m+1}=s\right]\\
&-\mathbb{E}\left[\min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{S_{m-1}}}}\{g_{m-1}(\hat{y},S_{m-1})\}\mid S_m=s\right].
\end{aligned}
\tag{A.8}
$$

Finally, we have:

$$
\begin{aligned}
&g_{m+1}(y^1,s)-g_m(y^1,s)\\
&=\alpha\cdot\mathbb{E}\left[V_m(y^1-d,S_m)\middle|S_{m+1}=s\right]-\alpha\cdot\mathbb{E}\left[V_{m-1}(y^1-d,S_{m-1})\middle|S_m=s\right]\\
&=\alpha\cdot\mathbb{E}\left[\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{S_m}}}\{g_m(\hat{y},S_m)\}\mid S_{m+1}=s\right]\\
&\quad-\alpha\cdot\mathbb{E}\left[\min_{\max(y^1-d,d)\le\hat{y}\le y^1-d+\frac{P}{c_{S_{m-1}}}}\{g_{m-1}(\hat{y},S_{m-1})\}\mid S_m=s\right] && \text{(A.9)}\\
&\le\alpha\cdot\mathbb{E}\left[\min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{S_m}}}\{g_m(\hat{y},S_m)\}\mid S_{m+1}=s\right]\\
&\quad-\alpha\cdot\mathbb{E}\left[\min_{\max(y^2-d,d)\le\hat{y}\le y^2-d+\frac{P}{c_{S_{m-1}}}}\{g_{m-1}(\hat{y},S_{m-1})\}\mid S_m=s\right] && \text{(A.10)}
\end{aligned}
$$

$$= \alpha \cdot \mathbb{E}\left[V_m(y^2 - d, S_m)\middle|S_{m+1} = s\right] - \alpha \cdot \mathbb{E}\left[V_{m-1}(y^2 - d, S_{m-1})\middle|S_m = s\right] \quad \text{(A.11)}$$

$$= g_{m+1}(y^2, s) - g_m(y^2, s) \ .$$

Here, (A.9) and (A.11) follow from the fact that

$$\mathbb{E}\left[c_{S_{m-1}} \mid S_m = s\right] = \mathbb{E}\left[c_{S_m} \mid S_{m+1} = s\right],$$

and (A.10) follows from (A.8). This completes the induction step, and the proof of (5.7).

***Part III - Monotonicity of Thresholds in the Channel Condition:*** Finally, we show (5.8), the monotonicity of the thresholds in the channel condition, when the channel condition process is IID. The far left inequality follows from the same interchange argument described above, showing $b_n(s) \leq n \cdot d$ for all $s$ and $n$. We now show the far right equality of (5.8), $b_n(s_{\text{worst}}) = d$. To satisfy feasibility, we must have $b_n(s) \geq d$ for all $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$. To see that $b_n(s_{\text{worst}}) \leq d$, assume the channel condition at time $n$ is $s_{\text{worst}}$, and consider two control policies satisfying (5.5), with the same critical numbers $b_m(s)$, for all times $m < n$. At time $n$, the first policy, $\boldsymbol{\pi^1}$, transmits according to (5.5), with critical number $b_n(s_{\text{worst}}) = d+\epsilon$ ($\epsilon > 0$), and the second, $\boldsymbol{\pi^2}$, transmits according to (5.5), with critical number $b_n(s_{\text{worst}}) = d$. These two strategies result in the same control action at time $n$ if $x_n \geq d + \epsilon$, and we have already shown it is not optimal to fill the buffer beyond $n \cdot d$, so we only need to consider the case where $x_n < d + \epsilon$ and $\epsilon \leq (n-1) \cdot d$. Let $Z_n^1, Z_{n-1}^1, \ldots, Z_1^1$ and $Z_n^2, Z_{n-1}^2, \ldots, Z_1^2$ be random variables representing the number of packets transmitted at times $n, n-1, \ldots, 1$ by $\boldsymbol{\pi^1}$ and $\boldsymbol{\pi^2}$, respectively. If $d \leq x_n \leq d + \epsilon$, then $Z_n^2 = 0$ and $Z_n^1 - Z_n^2 = Z_n^1 = \min\left\{\frac{P}{c_{\max}}, d + \epsilon - x_n\right\}$. If $x_n < d$, then $Z_n^2 = d - x_n$, $Z_n^1 = \min\left\{\frac{P}{c_{\max}}, d + \epsilon - x_n\right\}$, and $Z_n^1 - Z_n^2 = \min\left\{\frac{P}{c_{\max}} - d + x_n, \epsilon\right\}$. Thus, for all $x_n < d + \epsilon$, we have $Z_n^1 - Z_n^2 \geq 0$. If $Z_n^1 - Z_n^2 = 0$, the two control policies result in the same actions for all remaining times, and therefore result in the same expected cost. So we only need to consider the case where $\lambda := Z_n^1 - Z_n^2 > 0$. Because the critical numbers at times $n-1, n-2, \ldots, 1$ are the same for both policies, for any realization, $\omega$, of the channel condition over future times, we have $Z_m^1(\omega) \leq Z_m^2(\omega),\ \forall m \in \{n-1, \ldots, 1\}$. Moreover, because the scheduler must satisfy the playout requirements for the last $n$ slots, we have $\sum_{m=1}^{n-1}(Z_m^2(\omega) - Z_m^1(\omega)) = \lambda$; i.e., over the remainder of the horizon, an extra $\lambda$ packets are transmitted under the second policy. The total discounted holding costs from time $n$ until the end of the horizon are therefore lower for $\boldsymbol{\pi^2}$ than $\boldsymbol{\pi^1}$, because the number of packets remaining after transmission in each slot is never greater under policy $\boldsymbol{\pi^2}$. Furthermore, the total discounted transmission costs of the extra $\lambda$ packets are also lower for $\boldsymbol{\pi^2}$ as they are transmitted at the maximum cost $c_{\max}$ under $\boldsymbol{\pi^1}$, and transmitted later (and therefore discounted more heavily) under $\boldsymbol{\pi^2}$. Thus, the total discounted transmission plus holding costs are lower for $\boldsymbol{\pi^2}$ under all realizations, and the expected discounted cost of $\boldsymbol{\pi^2}$ is lower than $\boldsymbol{\pi^1}$. We conclude $b_n(s_{\text{worst}}) = d$.

To show $c_{s^1} \leq c_{s^2}$ implies $b_n(s^1) \geq b_n(s^2)$, we follow Kalymon's methodology for the proof of Theorem 1.3 in [81]. For all $y \in [d, \infty)$, we have:

$$\begin{aligned} g_n\left(y, s^2\right) &= c_{s^2} \cdot y + h(y-d) + \alpha \cdot \mathbb{E}\left[V_{n-1}\left(y-d, S_{n-1}\right)\right] \\ &= \left(c_{s^2} - c_{s^1}\right) \cdot y + c_{s^1} \cdot y + h(y-d) + \alpha \cdot \mathbb{E}\left[V_{n-1}\left(y-d, S_{n-1}\right)\right] \\ &= \left(c_{s^2} - c_{s^1}\right) \cdot y + g_n\left(y, s^1\right) \ . \end{aligned} \tag{A.12}$$

Assume $b_n(s^1) < b_n(s^2)$ for some $n \in \{1, 2, \ldots, N\}$ and $s^1, s^2 \in \mathcal{S}$, with $c_{s^1} \leq c_{s^2}$. Substituting first $y = b_n(s^1)$ and then $y = b_n(s^2)$ into (A.12) yields:

$$\begin{aligned} \left(c_{s^2} - c_{s^1}\right) \cdot b_n\left(s^1\right) + g_n\left(b_n\left(s^1\right), s^1\right) &= g_n\left(b_n\left(s^1\right), s^2\right) \\ &\geq g_n\left(b_n\left(s^2\right), s^2\right) \end{aligned}$$

$$= \quad (c_{s^2} - c_{s^1}) \cdot b_n\left(s^2\right)$$
$$+g_n\left(b_n\left(s^2\right), s^1\right). \qquad \text{(A.13)}$$

Yet, $c_{s^1} \leq c_{s^2}$ and $b_n(s^1) < b_n(s^2)$ imply:

$$(c_{s^2} - c_{s^1}) \cdot b_n\left(s^1\right) < (c_{s^2} - c_{s^1}) \cdot b_n\left(s^2\right) \ . \qquad \text{(A.14)}$$

Equations (A.13) and (A.14) imply:

$$g_n\left(b_n\left(s^1\right), s^1\right) > g_n\left(b_n\left(s^2\right), s^1\right) \ ,$$

which clearly contradicts the fact that $b_n\left(s^1\right)$ is a global minimizer of $g_n\left(\cdot, s^1\right)$. We conclude that $c_{s^1} \leq c_{s^2}$ implies $b_n(s^1) \geq b_n(s^2)$, completing the proofs of (5.8) and Theorem 5.1. □

## A.2 Proof of Theorem 5.3

While the proof is similar in spirit to the proof of a finite generalized base-stock policy in [16, pp. 324–334], some key differences include the introduction of (i) stochastic channel conditions (ordering costs); (ii) the underflow constraint $x+z \geq d$; and (iii) the power constraint $z \leq \tilde{z}_{\max}(s)$.

We show by induction on $n$ that the following two statements are true for every $n \in \{1, 2, \ldots, N\}$ and $s \in \mathcal{S}$:

(i) $V_n(x, s)$ is convex in $x$ on $I\!R_+$.

(ii) There exists a nonincreasing sequence of critical numbers $\left\{b_{n,k}(s)\right\}_{k\in\{-1,0,1,\ldots,K\}}$

such that the optimal control action with $n$ slots remaining is given by:

$$z_n^*(x,s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } \; b_{n,k}(s) - \tilde{z}_{k-1}(s) \le x < b_{n,k-1}(s) - \tilde{z}_{k-1}(s) \;, \\ & \quad k \in \{0,1,\ldots,K\} \\ b_{n,k}(s) - x, & \text{if } \; b_{n,k}(s) - \tilde{z}_k(s) \le x < b_{n,k}(s) - \tilde{z}_{k-1}(s) \;, \\ & \quad k \in \{0,1,\ldots,K-1\} \\ b_{n,K}(s) - x, & \text{if } \; b_{n,K}(s) - \tilde{z}_{\max}(s) \le x < b_{n,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } \; 0 \le x < b_{n,K}(s) - \tilde{z}_{\max}(s) \end{cases} \tag{A.15}$$

<u>Base Case</u>: $n = 1$

$$\begin{aligned} V_1(x,s) &= \min_{\max(0,d-x) \le z \le \tilde{z}_{\max}(s)} \{c(z,s) + h(x+z-d)\} \\ &= c\big(\max\{0, d-x\}, s\big) + h\big(\max\{0, x-d\}\big) \;, \end{aligned} \tag{A.16}$$

which is convex because $c(\cdot,s)$ and $h(\cdot)$ are both convex and nondecreasing functions, and $\max\{0,d-x\}$ and $\max\{0,x-d\}$ are both convex functions (see, e.g., [23, Section 3.2] for the relevant results on convexity-preserving operations). Further, let $b_{1,-1}(s) = \infty$ and $b_{1,k}(s) = d$ for all $k \in \{0,1,\ldots,K\}$. Then (A.15) is equivalent to $z_1^*(x,s) = \max\{0, d-x\}$, which clearly achieves the minimum in (A.16).

<u>Induction Step</u>: We now assume (i)-(ii) are true for $n = m-1$ and all $s \in \mathcal{S}$, and show they hold for $n = m$ and an arbitrary $s \in \mathcal{S}$. Let $\breve{x}, \hat{x} \in \mathbb{R}_+$ and $\theta \in [0,1]$ be arbitrary, and define $\bar{x} := \theta \cdot \breve{x} + (1-\theta) \cdot \hat{x}$. We have:

$$\begin{aligned} & V_m(\theta \cdot \breve{x} + (1-\theta) \cdot \hat{x}, s) \\ & = V_m(\bar{x}, s) \\ & = \min_{\max(0,d-\bar{x}) \le z \le \tilde{z}_{\max}(s)} \left\{ \begin{array}{c} (z,s) + h(\bar{x} + z - d) \\ + \alpha \cdot \mathbb{E}\big[V_{m-1}(\bar{x}+z-d, S_{m-1}) \mid S_m = s\big] \end{array} \right\} \end{aligned}$$

$$\leq \min_{\substack{\max\{0,d-\breve{x}\}\leq\breve{z}\leq\tilde{z}_{\max}(s)\\ \max\{0,d-\hat{x}\}\leq\hat{z}\leq\tilde{z}_{\max}(s)}} \left\{ \begin{array}{l} c(\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z},s)+h(\bar{x}+\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z}-d) \\ +\alpha\cdot \mathbb{E}\big[V_{m-1}(\bar{x}+\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z}-d,S_{m-1})\mid S_m=s\big] \end{array} \right\} \tag{A.17}$$

$$\leq \min_{\substack{\max\{0,d-\breve{x}\}\leq\breve{z}\leq\tilde{z}_{\max}(s)\\ \max\{0,d-\hat{x}\}\leq\hat{z}\leq\tilde{z}_{\max}(s)}} \left\{ \begin{array}{l} \theta\cdot c(\breve{z},s)+(1-\theta)\cdot c(\hat{z},s)+ \\ \theta\cdot h(\breve{x}+\breve{z}-d)+(1-\theta)\cdot h(\hat{x}+\hat{z}-d) \\ +\alpha\cdot\theta\cdot \mathbb{E}\big[V_{m-1}(\breve{x}+\breve{z}-d,S_{m-1})\mid S_m=s\big] \\ +\alpha\cdot(1-\theta)\cdot \mathbb{E}\big[V_{m-1}(\hat{x}+\hat{z}-d,S_{m-1})\mid S_m=s\big] \end{array} \right\} \tag{A.18}$$

$$= \theta\cdot \min_{\max\{0,d-\breve{x}\}\leq\breve{z}\leq\tilde{z}_{\max}(s)} \left\{ \begin{array}{l} c(\breve{z},s)+h(\breve{x}+\breve{z}-d) \\ +\alpha\cdot \mathbb{E}\big[V_{m-1}(\breve{x}+\breve{z}-d,S_{m-1})\mid S_m=s\big] \end{array} \right\}$$

$$+(1-\theta)\cdot \min_{\max\{0,d-\hat{x}\}\leq\hat{z}\leq\tilde{z}_{\max}(s)} \left\{ \begin{array}{l} c(\hat{z},s)+h(\hat{x}+\hat{z}-d) \\ +\alpha\cdot \mathbb{E}\big[V_{m-1}(\hat{x}+\hat{z}-d,S_{m-1})\mid S_m=s\big] \end{array} \right\}$$

$$= \theta\cdot V_m(\breve{x},s)+(1-\theta)\cdot V_m(\hat{x},s)\ ,$$

where (A.18) follows from the convexity of $c(\cdot,s)$, $h(\cdot)$, and $\mathbb{E}\big[V_{m-1}(\cdot,S_{m-1})\mid S_m=s\big]$, the last of which follows from the induction hypothesis. Equation (A.17) follows from the fact that for every $\max\{0,d-\breve{x}\}\leq\breve{z}\leq\tilde{z}_{\max}(s)$ and $\max\{0,d-\hat{x}\}\leq\hat{z}\leq\tilde{z}_{\max}(s)$, there exists a $\max\{0,d-\bar{x}\}\leq\bar{z}\leq\tilde{z}_{\max}(s)$ (namely, $\bar{z}:=\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z}$) such that:

$$\begin{aligned} & c(\bar{z},s)+h(\bar{x}+\bar{z}-d)+\alpha\cdot \mathbb{E}\big[V_{m-1}(\bar{x}+\bar{z}-d,S_{m-1})\mid S_m=s\big] \\ &= c(\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z},s)+h(\bar{x}+\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z}-d) \\ &\quad +\alpha\cdot \mathbb{E}\big[V_{m-1}(\bar{x}+\theta\cdot\breve{z}+(1-\theta)\cdot\hat{z}-d,S_{m-1})\mid S_m=s\big]\ . \end{aligned}$$

This concludes the induction step for (i) and we now proceed to (ii).

Note first that $\tilde{g}_m(y,s)=h(y-d)+\alpha\cdot \mathbb{E}\left[V_{m-1}\left(y-d,S_{m-1}\right)|S_m=s\right]$ is convex in $y$, as $h(\cdot)$ is convex, and $V_{m-1}(x,s)$ is convex in $x$ for every $s \in \mathcal{S}$ by the induction hypothesis. Let $b_{m,-1}(s) := \infty$ and

$$b_{m,k}(s) := \max\Big\{d, \inf\big\{b \;\big|\; \tilde{g}_m'^{+}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} \ , \ \forall k \in \{0,1,\ldots,K\} \ ,$$

where $\tilde{g}_m'^{+}(b,s)$ represents the right derivative:

$$\tilde{g}_m'^{+}(b,s) := \lim_{y \downarrow b} \frac{\tilde{g}_m(y,s) - \tilde{g}_m(b,s)}{y-b} \ ,$$

which is nondecreasing and continuous from the right, by the convexity of $\tilde{g}_m(\cdot,s)$ [122, Section 24]. Note that $\big\{b_{m,k}(s)\big\}_{k\in\{-1,0,1,\ldots,K\}}$ is a nonincreasing sequence, because the sequence $\{\tilde{c}_k(s)\}_{k\in\{0,1,\ldots,K\}}$ is nondecreasing. We show the optimal control action $z_m^*(x,s)$ is then given by (A.15), by considering the four exhaustive cases.

<u>Case 1</u>: $b_{m,k}(s) - \tilde{z}_{k-1}(s) \leq x < b_{m,k-1}(s) - \tilde{z}_{k-1}(s) \ , \ k \in \{0,1,\ldots,K\}$

In order to show $z_m^*(x,s)$ is given by (A.15), it suffices to show:

$$c'^{+}(z,s) + \tilde{g}_m'^{+}(x+z,s) < 0 \ , \ \text{for } \max\{0, d-x\} \leq z < \tilde{z}_{k-1}(s), \ \text{and} \tag{A.19}$$

$$c'^{+}(z,s) + \tilde{g}_m'^{+}(x+z,s) \geq 0 \ , \ \text{for } \tilde{z}_{k-1}(s) \leq z \leq \tilde{z}_{\max}(s) \ . \tag{A.20}$$

First, let $z \in \Big[\max\{0,d-x\}, \tilde{z}_{k-1}(s)\Big)$ be arbitrary, and let $j \in \{0,1,\ldots,k-1\}$ be such that $z \in \Big[\tilde{z}_{j-1}(s), \tilde{z}_j(s)\Big)$. If $b_{m,k-1}(s) = d$, then $b_{m,k}(s) = d$, as $d \leq b_{m,k}(s) \leq b_{m,k-1}(s) = d$. Yet, $b_{m,k}(s) = b_{m,k-1}(s) = d$ implies $d - \tilde{z}_{k-1}(s) \leq x < d - \tilde{z}_{k-1}(s)$, which is vacuous. Therefore, we need only consider $b_{m,k-1}(s) = \inf\big\{b \;\big|\; \tilde{g}_m'^{+}(b,s) \geq -\tilde{c}_{k-1}(s)\big\}$. By the construction of the piecewise-linear function $c(\cdot,s)$, $z < \tilde{z}_{k-1}(s)$ implies:

$$c'^{+}(z,s) \leq \tilde{c}_{k-1}(s) \ . \tag{A.21}$$

We also have:

$$x + z < x + \tilde{z}_{k-1}(s) < b_{m,k-1}(s) = \inf\big\{b \;\big|\; \tilde{g}_m'^{+}(b,s) \geq -\tilde{c}_{k-1}(s)\big\} \ ,$$

which implies:

$$\tilde{g}'^{+}_{m}(x+z,s) < -\tilde{c}_{k-1}(s) \ . \tag{A.22}$$

Summing (A.21) and (A.22) yields (A.19).

Next, let $z \in \big[\tilde{z}_{k-1}(s), \tilde{z}_{\max}(s)\big]$ be arbitrary, so that by construction of $c(\cdot, s)$:

$$c'^{+}(z,s) \geq \tilde{c}_{k}(s) \ . \tag{A.23}$$

We also have:

$$x + z \geq x + \tilde{z}_{k-1}(s) \geq b_{m,k}(s) \geq \inf\big\{b \;\big|\; \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_{k}(s)\big\} \ ,$$

which, in combination with the nondecreasing nature of $\tilde{g}'^{+}_{m}(\cdot, s)$, implies:

$$\tilde{g}'^{+}_{m}(x+z,s) \geq \tilde{g}'^{+}_{m}\Big(\inf\big\{b \;\big|\; \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_{k}(s)\big\}, s\Big) \ . \tag{A.24}$$

Because $\tilde{g}'^{+}_{m}(\cdot, s)$ is continuous from the right,

$$\tilde{g}'^{+}_{m}\Big(\inf\big\{b \;\big|\; \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_{k}(s)\big\}, s\Big) \geq -\tilde{c}_{k}(s) \ . \tag{A.25}$$

Combining (A.24) and (A.25), and summing with (A.23) yields (A.20).

<u>Case 2</u>: $b_{m,k}(s) - \tilde{z}_{k}(s) \leq x < b_{m,k}(s) - \tilde{z}_{k-1}(s) \ , \ k \in \{0, 1, \ldots, K-1\}$

In order to show $z^{*}_{m}(x, s)$ is given by (A.15), it suffices to show:

$$c'^{+}(z,s) + \tilde{g}'^{+}_{m}(x+z,s) < 0 \ , \ \text{for} \ \max\{0, d-x\} \leq z < b_{m,k}(s) - x, \tag{A.26}$$

and

$$c'^{+}(z,s) + \tilde{g}'^{+}_{m}(x+z,s) \geq 0 \ , \ \text{for} \ b_{m,k}(s) - x \leq z \leq \tilde{z}_{\max}(s) \ . \tag{A.27}$$

First, let $z \in \Big[\max\{0, d-x\}, b_{m,k}(s) - x\Big)$ be arbitrary. This case is vacuous if $b_{m,k}(s) = d$, so $b_{m,k}(s) = \inf\big\{b \;\big|\; \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_{k}(s)\big\}$. Thus, we have:

$$x + z < b_{m,k}(s) = \inf\big\{b \;\big|\; \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_{k}(s)\big\} \ ,$$

which implies:

$$\tilde{g}'^{+}_{m}(x+z,s) < -\tilde{c}_k(s)\ . \tag{A.28}$$

Furthermore, from $z < b_{m,k}(s) - x \leq \tilde{z}_k(s)$ and the construction of the piecewise-linear function $c(\cdot, s)$,

$$c'^{+}(z,s) \leq \tilde{c}_k(s)\ . \tag{A.29}$$

Summing (A.28) and (A.29) yields (A.26).

Next, let $z \in \big[b_{m,k}(s) - x, \tilde{z}_{\max}(s)\big]$ be arbitrary, so that $z \geq b_{m,k}(s) - x > \tilde{z}_{k-1}(s)$, which by the construction of the piecewise-linear function $c(\cdot, s)$ implies:

$$c'^{+}(z,s) \geq \tilde{c}_k(s)\ . \tag{A.30}$$

We also have $x + z \geq b_{m,k}(s) \geq \inf\big\{b \bigm| \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_k(s)\big\}$. Therefore, because $\tilde{g}'^{+}_{m}(\cdot, s)$ is nondecreasing and continuous from the right,

$$\tilde{g}'^{+}_{m}\,(x+z,s) \geq \tilde{g}'^{+}_{m}\,\Big(\inf\big\{b \bigm| \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_k(s)\big\}, s\Big) \geq -\tilde{c}_k(s)\ . \tag{A.31}$$

Summing (A.30) and (A.31) yields (A.27).

<u>Case 3</u>: $b_{m,K}(s) - \tilde{z}_{\max}(s) \leq x < b_{m,K}(s) - \tilde{z}_{K-1}(s)$

This case is the same as Case 2, with $K$ in place of $k$, and $\tilde{z}_{\max}(s)$ in place of $\tilde{z}_k(s)$.

<u>Case 4</u>: $0 \leq x < b_{m,K}(s) - \tilde{z}_{\max}(s)$

Let $z \in \big[\max\{0, d - x\}, \tilde{z}_{\max}(s)\big)$ be arbitrary. $\tilde{z}_{\max}(s) \geq d$ by assumption, so this case is vacuous if $b_{m,K} = d$. Thus, we have $b_{m,K}(s) = \inf\big\{b \bigm| \tilde{g}'^{+}_{m}(b,s) \geq -\tilde{c}_K(s)\big\}$, which, in combination with $x + z < x + \tilde{z}_{\max} < b_{m,K}(s)$, implies:

$$\tilde{g}'^{+}_{m}\,(x+z,s) < -\tilde{c}_K(s)\ . \tag{A.32}$$

Additionally, $z < \tilde{z}_{\max}(s)$ implies:

$$c'^{+}(z,s) \leq \tilde{c}_K(s)\ . \tag{A.33}$$

Summing (A.32) and (A.33) yields $c'^{+}(z,s)+\tilde{g}'^{+}_{m}(x+z,s)<0$ for all $z \in \big[\max\{0, d-x\}, \tilde{z}_{\max}(s)\big)$, which implies $z^*_m(x,s) = \tilde{z}_{\max}(s)$. □

## A.3 Proof of Theorem 5.4

We proceed in a manner similar to [56], incorporating the per slot peak power constraints and the relaxing the linear ordering costs to piecewise-linear convex ordering costs. Before proving Theorem 5.2, we state and prove two lemmas. Let $\bar{\boldsymbol{\pi}}$ be a strategy that prescribes transmitting according to (5.18).

**Lemma A.2.** *If $\bar{\boldsymbol{\pi}}$ is optimal for periods $m-1, m-2, \ldots, 1$, then*

$$\alpha \cdot \mathbb{E}\left[V_{l-1}\big((r-1)\cdot d+\eta, S\big) - V_{l-1}\big((r-1)\cdot d, S\big)\right] \geq -\eta \cdot \big(\tilde{\gamma}_{l,r+1}+h\big), \quad \text{(A.34)}$$

*for all* $(l,r,\eta) \in \mathcal{Z}_1 := \{(l,r,\eta) \in \mathbb{N} \times \mathbb{N} \times [0,d] : 1 \leq l \leq m, 1 \leq r \leq l\}$.

*Proof.* We proceed by induction on $l$.

<u>Base Case</u>: $l = 1$

$l = 1$ implies $r = 1$, so we have:

$$\begin{aligned}
& \alpha \cdot \mathbb{E}\left[V_{l-1}\big((r-1)\cdot d+\eta, S\big) - V_{l-1}\big((r-1)\cdot d, S\big)\right] \\
= \;& \alpha \cdot \mathbb{E}\left[V_0\big(\eta, S\big) - V_0\big(0, S\big)\right] \\
= \;& 0 \\
\geq \;& -\eta \cdot h \\
= \;& -\eta \cdot (\tilde{\gamma}_{1,2}+h) \ ,
\end{aligned}$$

and we conclude (A.34) holds for $l = 1$.

<u>Induction Step</u>

Assume (A.34) is true for $l = 2, 3, \ldots, t$ and all $r$ and $\eta$ such that $(l,r,\eta) \in \mathcal{Z}_1$. We show (A.34) is true for $l = t+1$ by letting $r$ and $\eta$ be arbitrary such that

$(t+1, r, \eta) \in \mathcal{Z}_1$. Note that $(t+1, r, \eta) \in \mathcal{Z}_1$ implies $t \le m-1$, so $\bar{\boldsymbol{\pi}}$ is optimal at time $t$, and we have:

$$
\begin{aligned}
&\alpha \cdot \mathbb{E}\big[V_t\big((r-1)d+\eta, S\big) - V_t\big((r-1)d, S\big)\big] \\
&= \sum_{\left\{s:\ b_{t,0}(s) \le (r-1)d\right\}} \alpha \cdot p(s) \cdot \left[h \cdot \eta + \alpha \cdot \mathbb{E}\begin{bmatrix} V_{t-1}\big((r-2)d+\eta, S\big) \\ -V_{t-1}\big((r-2)d, S\big) \end{bmatrix}\right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{array}{l} \displaystyle\sum_{\left\{s: \left(r-1+\tilde{L}_{k-1}(s)\right) d < b_{t,k}(s) \le \left(r-1+\tilde{L}_k(s)\right) d\right\}} - \ \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_k(s)\Big) \\ + \displaystyle\sum_{\left\{\substack{s:\ b_{t,k+1}(s) \\ \le \left(r-1+\tilde{L}_k(s)\right) d \\ < b_{t,k}(s)}\right\}} \alpha \cdot p(s) \cdot \begin{bmatrix} h \cdot \eta + \\ \alpha \mathbb{E} \begin{bmatrix} V_{t-1}\Big(\big(r-2+\tilde{L}_k(s)\big)d+\eta, S\Big) \\ -V_{t-1}\Big(\big(r-2+\tilde{L}_k(s)\big)d, S\Big) \end{bmatrix} \end{bmatrix} \end{array} \right\} \\
&+ \sum_{\left\{s: \left(r-1+\tilde{L}_{K-1}(s)\right) d < b_{t,K}(s) \le \left(r-1+\tilde{L}_{\max}(s)\right) d\right\}} - \ \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_K(s)\Big) \\
&+ \sum_{\left\{\substack{s:\ b_{t,K}(s) \\ > \left(r-1+\tilde{L}_{\max}(s)\right) d}\right\}} \alpha \cdot p(s) \cdot \begin{bmatrix} h \cdot \eta \\ +\alpha \cdot \mathbb{E} \begin{bmatrix} V_{t-1}\Big(\big(r-2+\tilde{L}_{\max}(s)\big)d+\eta, S\Big) \\ -V_{t-1}\Big(\big(r-2+\tilde{L}_{\max}(s)\big)d, S\Big) \end{bmatrix} \end{bmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&\ge \sum_{\left\{s:\ b_{t,0}(s) \le (r-1)d\right\}} \alpha \cdot p(s) \cdot \Big[-\eta \cdot \tilde{\gamma}_{t,r}\Big] \\
&\quad + \sum_{k=0}^{K-1} \left\{ \begin{array}{l} \displaystyle\sum_{\left\{s: \left(r-1+\tilde{L}_{k-1}(s)\right) d < b_{t,k}(s) \le \left(r-1+\tilde{L}_k(s)\right) d\right\}} - \ \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_k(s)\Big) \\ + \displaystyle\sum_{\left\{s:\ b_{t,k+1}(s) \le \left(r-1+\tilde{L}_k(s)\right) d < b_{t,k}(s)\right\}} \alpha \cdot p(s) \cdot \Big[-\eta \cdot \tilde{\gamma}_{t,r+\tilde{L}_k(s)}\Big] \end{array} \right\} \\
&\quad + \sum_{\left\{s: \left(r-1+\tilde{L}_{K-1}(s)\right) d < b_{t,K}(s) \le \left(r-1+\tilde{L}_{\max}(s)\right) d\right\}} - \ \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_K(s)\Big) \\
&\quad + \sum_{\left\{s:\ b_{t,K}(s) > \left(r-1+\tilde{L}_{\max}(s)\right) d\right\}} \alpha \cdot p(s) \cdot \Big[-\eta \cdot \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)}\Big]
\end{aligned}
$$

$$
= \; -\alpha\cdot\eta\cdot\left\{\begin{array}{l}
\displaystyle\sum_{s:\ \tilde{c}_0(s)\geq\tilde{\gamma}_{t,r}} p(s)\cdot\tilde{\gamma}_{t,r} \\
\displaystyle +\sum_{k=0}^{K-1}\left\{\begin{array}{l}
\displaystyle\sum_{s:\ \tilde{\gamma}_{t,r+\tilde{L}_k(s)}\leq\tilde{c}_k(s)<\tilde{\gamma}_{t,r+\tilde{L}_{k-1}(s)}} p(s)\cdot\tilde{c}_k(s) \\
\displaystyle +\sum_{s:\ \tilde{c}_k(s)<\tilde{\gamma}_{t,r+\tilde{L}_k(s)}\leq\tilde{c}_{k+1}(s)} p(s)\cdot\tilde{\gamma}_{t,r+\tilde{L}_k(s)}
\end{array}\right\} \\
\displaystyle +\sum_{s:\ \tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)}\leq\tilde{c}_K(s)<\tilde{\gamma}_{t,r+\tilde{L}_{K-1}(s)}} p(s)\cdot\tilde{c}_K(s) \\
\displaystyle +\sum_{s:\ \tilde{c}_K(s)<\tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)}} p(s)\cdot\tilde{\gamma}_{t,r+\tilde{L}_{\max}(s)}
\end{array}\right\}
$$

$$
= \; -\eta\cdot(\tilde{\gamma}_{t+1,r+1}+h)\ ,
$$

where the inequality follows from the induction hypothesis, and the penultimate equality follows from the definition $b_{n,k} := j\cdot d$, if $\tilde{\gamma}_{n,j+1}\leq\tilde{c}_k(s)<\tilde{\gamma}_{n,j}$. This concludes the induction step, and the proof of Lemma A.2. □

**Lemma A.3.** *If $\bar{\boldsymbol{\pi}}$ is optimal for periods $m-1, m-2, \ldots, 1$, then*

$$
\alpha\cdot\mathbb{E}\left[V_{l-1}\big((r-1)\cdot d-\eta, S\big)-V_{l-1}\big((r-1)\cdot d, S\big)\right]\geq\eta\cdot\big(\tilde{\gamma}_{l,r}+h\big), \qquad \text{(A.35)}
$$

*for all $(l,r,\eta)\in\mathcal{Z}_2 := \{(l,r,\eta)\in\mathbb{N}\times\mathbb{N}\times[0,d] : 2\leq l\leq m, 2\leq r\leq l\}$.*

*Proof.* We proceed by induction on $l$.

<u>Base Case</u>: $l=2$

$l=2$ implies $r=2$, so we have:

$$
\begin{aligned}
&\alpha\cdot\mathbb{E}\left[V_{l-1}\big((r-1)\cdot d-\eta, S\big)-V_{l-1}\big((r-1)\cdot d, S\big)\right] \\
&\quad = \; \alpha\cdot\mathbb{E}\left[V_1\big(d-\eta, S\big)-V_1\big(d, S\big)\right] \\
&\quad = \; \alpha\cdot\mathbb{E}\big[c(\eta, S)\big] \\
&\quad = \; \eta\cdot(\tilde{\gamma}_{2,2}+h)\,,
\end{aligned}
$$

where the last equality follows from $\gamma_{2,2} = -h + \alpha \cdot \mathbb{E}[\tilde{c}_0(S)]$, and the fact that $\eta \leq \tilde{z}_0(s)$ for every $s \in \mathcal{S}$. So (A.35) holds with equality for $l = 2$.

Induction Step

Assume (A.35) is true for $l = 2, 3, \ldots, t$ and all $r$ and $\eta$ such that $(l, r, \eta) \in \mathcal{Z}_2$. We show (A.35) is true for $l = t + 1$ by letting $r$ and $\eta$ be arbitrary such that $(t + 1, r, \eta) \in \mathcal{Z}_2$. Note that $(t + 1, r, \eta) \in \mathcal{Z}_2$ implies $t \leq m - 1$, so $\bar{\boldsymbol{\pi}}$ is optimal at time $t$, and we have:

$$
\begin{aligned}
&\alpha \cdot \mathbb{E}\big[V_t\big((r-1)d - \eta, S\big) - V_t\big((r-1)d, S\big)\big] \\
&= \sum_{\left\{s:\ b_{t,0}(s) \leq (r-2)d\right\}} \alpha \cdot p(s) \cdot \left[ -\eta \cdot h + \alpha \cdot \mathbb{E} \left[ \begin{array}{l} V_{t-1}\big((r-2)d - \eta, S\big) \\ -V_{t-1}\big((r-2)d, S\big) \end{array} \right] \right] \\
&+ \sum_{k=0}^{K-1} \left\{ \begin{array}{l} \displaystyle\sum_{\left\{s: \left(r-2+\tilde{L}_{k-1}(s)\right) d < b_{t,k}(s) \leq \left(r-2+\tilde{L}_k(s)\right) d\right\}} \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_k(s)\Big) \\
+ \displaystyle\sum_{\left\{\begin{array}{c} s:\ b_{t,k+1}(s) \\ \leq \left(r-2+\tilde{L}_k(s)\right) d \\ < b_{t,k}(s) \end{array}\right\}} \alpha \cdot p(s) \cdot \left[ \begin{array}{l} -\eta \cdot h + \\ \alpha \mathbb{E} \left[ \begin{array}{l} V_{t-1}\Big(\big(r-2+\tilde{L}_k(s)\big)d - \eta, S\Big) \\ -V_{t-1}\Big(\big(r-2+\tilde{L}_k(s)\big)d, S\Big) \end{array} \right] \end{array} \right] \end{array} \right\} \\
&+ \sum_{\left\{s: \left(r-2+\tilde{L}_{K-1}(s)\right) d < b_{t,K}(s) \leq \left(r-2+\tilde{L}_{\max}(s)\right) d\right\}} \alpha \cdot p(s) \cdot \Big(\eta \cdot \tilde{c}_K(s)\Big) \\
&+ \sum_{\left\{\begin{array}{c} s:\ b_{t,K}(s) \\ > \left(r-2+\tilde{L}_{\max}(s)\right) d \end{array}\right\}} \alpha \cdot p(s) \cdot \left[ \begin{array}{l} -\eta \cdot h \\ +\alpha \cdot \mathbb{E} \left[ \begin{array}{l} V_{t-1}\Big(\big(r-2+\tilde{L}_{\max}(s)\big)d - \eta, S\Big) \\ -V_{t-1}\Big(\big(r-2+\tilde{L}_{\max}(s)\big)d, S\Big) \end{array} \right] \end{array} \right]
\end{aligned}
$$

$$
\begin{aligned}
&\geq \sum_{\left\{s:\ b_{t,0}(s)\leq(r-2)d\right\}} \alpha\cdot p(s)\cdot\Big[\eta\cdot\tilde{\gamma}_{t,r-1}\Big] \\
&\quad + \sum_{k=0}^{K-1}\left\{\begin{array}{l} \displaystyle\sum_{\left\{s:\left(r-2+\tilde{L}_{k-1}(s)\right)d<b_{t,k}(s)\leq\left(r-2+\tilde{L}_k(s)\right)d\right\}} \alpha\cdot p(s)\cdot\Big(\eta\cdot\tilde{c}_k(s)\Big) \\
\displaystyle + \sum_{\left\{s:\ b_{t,k+1}(s)\leq\left(r-2+\tilde{L}_k(s)\right)d<b_{t,k}(s)\right\}} \alpha\cdot p(s)\cdot\Big[\eta\cdot\tilde{\gamma}_{t,r-1+\tilde{L}_k(s)}\Big]\end{array}\right\} \\
&\quad + \sum_{\left\{s:\left(r-2+\tilde{L}_{K-1}(s)\right)d<b_{t,K}(s)\leq\left(r-2+\tilde{L}_{\max}(s)\right)d\right\}} \alpha\cdot p(s)\cdot\Big(\eta\cdot\tilde{c}_K(s)\Big) \\
&\quad + \sum_{\left\{s:\ b_{t,K}(s)>\left(r-2+\tilde{L}_{\max}(s)\right)d\right\}} \alpha\cdot p(s)\cdot\Big[\eta\cdot\tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)}\Big]
\end{aligned}
$$

$$
= \alpha\cdot\eta\cdot\left\{\begin{array}{l}
\displaystyle\sum_{s:\ \tilde{c}_0(s)\geq\tilde{\gamma}_{t,r-1}} p(s)\cdot\tilde{\gamma}_{t,r-1} \\
\displaystyle + \sum_{k=0}^{K-1}\left\{\begin{array}{l}\displaystyle\sum_{s:\ \tilde{\gamma}_{t,r-1+\tilde{L}_k(s)}\leq\tilde{c}_k(s)<\tilde{\gamma}_{t,r-1+\tilde{L}_{k-1}(s)}} p(s)\cdot\tilde{c}_k(s) \\
\displaystyle + \sum_{s:\ \tilde{c}_k(s)<\tilde{\gamma}_{t,r-1+\tilde{L}_k(s)}\leq\tilde{c}_{k+1}(s)} p(s)\cdot\tilde{\gamma}_{t,r-1+\tilde{L}_k(s)}\end{array}\right\} \\
\displaystyle + \sum_{s:\ \tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)}\leq\tilde{c}_K(s)<\tilde{\gamma}_{t,r-1+\tilde{L}_{K-1}(s)}} p(s)\cdot\tilde{c}_K(s) \\
\displaystyle + \sum_{s:\ \tilde{c}_K(s)<\tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)}} p(s)\cdot\tilde{\gamma}_{t,r-1+\tilde{L}_{\max}(s)}
\end{array}\right\}
$$

$$
= \eta\cdot\left(\tilde{\gamma}_{t+1,r}+h\right)\ ,
$$

where the inequality follows from the induction hypothesis, and the penultimate equality again follows from the definition of $b_{n,k}(s)$. This concludes the induction step, and the proof of Lemma A.3. □

We now return to the proof of Theorem 5.2. We first show by induction that $V_n^{\bar{\boldsymbol{\pi}}}(x,s)=V_n(x,s),\forall n\in\{1,2,\ldots,N\}$, $\forall s\in\mathcal{S}$, and $\forall x\in\{0,d,2d,3d,\ldots\}$.

Base Case: $n = 1$

With one slot remaining, we have:

$$\begin{aligned} V_1(x,s) &= \min_{\left\{\max(0,d-x)\leq z_1 \leq \tilde{z}_{\max}(s)\right\}} \{c(z_1,s) + h(x+z_1-d)\} \\ &= c\Big(\max\{0,d-x\},s\Big) + h\Big(\max\{0,(x-d)\}\Big) \ , \end{aligned}$$

where the minimum is achieved by $z_1 = \max\{0, d-x\}$. $\tilde{\gamma}_{1,1} = \infty$ and $\tilde{\gamma}_{1,2} = 0$, so $b_{1,k}(s) = d$ for all $s \in \mathcal{S}$ and $k \in \{0,1,\ldots,K\}$. Thus, according to (5.18), $\bar{z}_1(x,s)$ is also equal to $\max\{0, d-x\}$, the optimal amount.

Induction Step

Assume that for $n = \{1, 2, \ldots, m-1\}$,

$$V_n^{\bar{\boldsymbol{\pi}}}(x,s) = V_n(x,s), \ \forall x \in \{0, d, 2d, 3d, \ldots\}, \ \forall s \in \mathcal{S}.$$

We show this is also true for $n = m$ by considering first any strategy that transmits more than $\bar{\boldsymbol{\pi}}$ at time $m$, and then any strategy that transmits less than $\bar{\boldsymbol{\pi}}$ at time $m$. Let $s \in \mathcal{S}$ be arbitrary. with $\tilde{\gamma}_{m,j_k+1} \leq \tilde{c}_k(s) < \tilde{\gamma}_{m,j_k}$ so that $\bar{\boldsymbol{\pi}}$ prescribes $b_{m,k}(s) = j_k \cdot d$ for $k \in \{0,1,\ldots,K\}$. Let $\boldsymbol{\pi}^q$ be a strategy that at time $m$ transmits enough to satisfy the demands of slots $m, m-1, m-2, \ldots, q+1$, and $q$, and transmits optimally at times $m-1$,$m-2, \ldots, 1$.

Part I: Do not transmit more than suggested by $\bar{\boldsymbol{\pi}}$ at time $m$

Let $\boldsymbol{\pi}'(\epsilon)$ be a feasible strategy with $z'_m = \bar{z}_m + \epsilon$, where $\epsilon > 0$, and the optimal transmission policy at times $m-1$,$m-2, \ldots, 1$. We consider four cases for the current buffer level $x$.

Case (a): $j_k \cdot d - \tilde{z}_{k-1}(s) < x \leq j_{k-1} \cdot d - \tilde{z}_{k-1}(s), \ k \in \{0,1,\ldots,K\}$

In this case, $\bar{z}_m = \tilde{z}_{k-1}(s)$. Let $p$ be the integer such that $x + \tilde{z}_{k-1}(s) = p \cdot d$. Let

$q,\ \eta$ be such that $z'_m = \tilde{z}_{k-1}(s) + \epsilon = q \cdot d + \eta - x$ and $0 \leq \eta < d$ $\Big($i.e., $q = \left\lfloor \frac{z'_m + x}{d} \right\rfloor$ and $\eta = z'_m + x - q \cdot d\Big)$. Thus, we have $q \geq p \geq j_k$.

Then we have:

$$
\begin{aligned}
V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s) - V_m^{\boldsymbol{\pi}^q}(x,s) &= c\Big(z'_m, s\Big) - c\Big(z'_m - \eta, s\Big) \\
&\quad + \eta \cdot h \\
&\quad + \alpha \cdot \mathbb{E}\Big[V_{m-1}\big((q-1)\cdot d + \eta, S\big) - V_{m-1}\big((q-1)\cdot d, S\big)\Big] \\
&\geq c\Big(z'_m, s\Big) - c\Big(z'_m - \eta, s\Big) \\
&\quad - \eta \cdot \tilde{\gamma}_{m,q+1} \qquad \text{(A.36)} \\
&\geq c\Big(z'_m, s\Big) - c\Big(z'_m - \eta, s\Big) \\
&\quad - \eta \cdot \tilde{\gamma}_{m,j_k+1} \qquad \text{(A.37)} \\
&\geq \eta \cdot \left(\tilde{c}_k(s) - \tilde{\gamma}_{m,j_k+1}\right) \qquad \text{(A.38)} \\
&\geq 0. \qquad \text{(A.39)}
\end{aligned}
$$

Equation (A.36) follows from Lemma A.2, with $l = m$, $r = q$, and $\eta = \eta$. Equation (A.37) follows from $q + 1 \geq j_k + 1$, which implies $\tilde{\gamma}_{m,q+1} \leq \tilde{\gamma}_{m,j_k+1}$. Equation (A.38) follows from $z'_m - \eta \geq \tilde{z}_{k-1}(s)$ and the construction of $c(\cdot, s)$. Finally, (A.39) follows from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m,j_k+1}$, by construction of $j_k$, and we conclude:

$$V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s) \geq V_m^{\boldsymbol{\pi}^q}(x,s)\ . \qquad \text{(A.40)}$$

Now let $t \in \{q+1, q+2, \ldots, m-p, m-p+1\}$ be arbitrary. We have:

$$
\begin{aligned}
V_m^{\boldsymbol{\pi}^{t-1}}(x,s) - V_m^{\boldsymbol{\pi}^t}(x,s) &= c\Big((m-t+2)\cdot d - x, s\Big) \\
&\quad - c\Big((m-t+1)\cdot d - x, s\Big) + d \cdot h \\
&\quad + \alpha \cdot \mathbb{E}\left[\begin{array}{l} V_{m-1}\big((m-t+1)\cdot d, S\big) \\ -V_{m-1}\big((m-t)\cdot d, S\big) \end{array}\right]
\end{aligned}
$$

$$
\begin{aligned}
&\geq c\Big((m-t+2)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad - d\cdot \tilde{\gamma}_{m,m-t+2} && \text{(A.41)} \\
&\geq c\Big((m-t+2)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad - d\cdot \tilde{\gamma}_{m,j_k+1} && \text{(A.42)} \\
&\geq d\cdot \big(\tilde{c}_k(s) - \tilde{\gamma}_{m,j_k+1}\big) && \text{(A.43)} \\
&\geq 0. && \text{(A.44)}
\end{aligned}
$$

Equation (A.41) follows from Lemma A.2, with $l = m$, $r = m-t+1 \leq m-q \leq m = l$, and $\eta = d$. Equation (A.42) follows from:

$$
\begin{aligned}
t \leq m - p + 1 \ \Leftrightarrow\ p + 1 \leq m - t + 2 \ &\Rightarrow j_k + 1 \leq m - t + 2 \\
&\Rightarrow\ \tilde{\gamma}_{m,j_k+1} \geq \tilde{\gamma}_{m,m-t+2}.
\end{aligned}
$$

Equation (A.43) follows from the construction of $c(\cdot, s)$ and the fact that:

$$
(m-t+1)\cdot d - x \geq \Big[m - (m-p+1) + 1\Big]\cdot d - x = p\cdot d - x = \tilde{z}_{k-1}(s)\ .
$$

Finally, (A.44) follows once again from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m,j_k+1}$, by construction of $j_k$. Rearranging (A.44) yields:

$$
V_m^{\boldsymbol{\pi}^{t-1}}(x,s) \geq V_m^{\boldsymbol{\pi}^{t}}(x,s),\ \forall t \in \{q+1, q+2, \ldots, m-p, m-p+1\}\,. \qquad \text{(A.45)}
$$

Noting that $V_m^{\bar{\boldsymbol{\pi}}}(x,s) = V_m^{\boldsymbol{\pi}^{m-p+1}}(x,s)$, (A.40) and repeated application of (A.45) imply:

$$
\begin{aligned}
&V_m^{\bar{\boldsymbol{\pi}}}(x,s) \\
&= V_m^{\boldsymbol{\pi}^{m-p+1}}(x,s) \leq V_m^{\boldsymbol{\pi}^{m-p}}(x,s) \leq \ldots \leq V_m^{\boldsymbol{\pi}^{q+1}}(x,s) \leq V_m^{\boldsymbol{\pi}^{q}}(x,s) \\
&\qquad \leq V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s)\ ,
\end{aligned}
$$

and we conclude $\bar{\boldsymbol{\pi}}$ is at least as good as $\boldsymbol{\pi}'(\epsilon)$.

Case (b): $j_k \cdot d - \tilde{z}_k(s) < x \le j_k \cdot d - \tilde{z}_{k-1}(s),\ k \in \{0, 1, \ldots, K-1\}$

Let $q, \eta$ be such that $z'_m = (m-q+1)\cdot d + \eta - x$ and $0 \le \eta < d$ $\Big($i.e., $q = m+1-\left\lfloor \frac{z'_m+x}{d} \right\rfloor$ and $\eta = z'_m - (m-q+1)\cdot d - x\Big)$. Note that $m-q+1 \ge j_k$ by the assumption that $z'_m \ge \bar{z}_m = j_k \cdot d - x$. Additionally, because $x \le j_k \cdot d - \tilde{z}_{k-1}(s)$ and $m-q+1 \ge j_k$, we have:

$$(m-q+1)\cdot d - x \ge (m-q+1-j_k)\cdot d + \tilde{z}_{k-1}(s) \ge \tilde{z}_{k-1}(s)\ ,$$

which implies:

$$c\Big((m-q+1)\cdot d + \eta - x, s\Big) - c\Big((m-q+1)\cdot d - x, s\Big) \ge \eta \cdot \tilde{c}_k(s)\ . \qquad \text{(A.46)}$$

Then we have:

$$\begin{aligned}
V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s) - V_m^{\boldsymbol{\pi}^q}(x,s) &= c\Big((m-q+1)\cdot d + \eta - x, s\Big) \\
&\quad - c\Big((m-q+1)\cdot d - x, s\Big) + \eta \cdot h \\
&\quad + \alpha \cdot I\!\!E \left[ \begin{array}{l} V_{m-1}\big((m-q)\cdot d + \eta, S\big) \\ -V_{m-1}\big((m-q)\cdot d, S\big) \end{array} \right] \\
&\ge c\Big((m-q+1)\cdot d + \eta - x, s\Big) \\
&\quad - c\Big((m-q+1)\cdot d - x, s\Big) - \eta \cdot \tilde{\gamma}_{m,m-q+2} && \text{(A.47)} \\
&\ge c\Big((m-q+1)\cdot d + \eta - x, s\Big) \\
&\quad - c\Big((m-q+1)\cdot d - x, s\Big) - \eta \cdot \tilde{\gamma}_{m,j_k+1} && \text{(A.48)} \\
&\ge \eta \cdot \big(\tilde{c}_k(s) - \tilde{\gamma}_{m,j_k+1}\big) && \text{(A.49)} \\
&\ge 0. && \text{(A.50)}
\end{aligned}$$

Equation (A.47) follows from Lemma A.2, with $l = m$, $r = m-q \le m-1$, and $\eta = \eta$. Equation (A.48) follows from $m-q+2 \ge j_k+1$, which implies $\tilde{\gamma}_{m,m-q+2} \le \tilde{\gamma}_{m,j_k+1}$. Equation (A.49) follows from (A.46). Finally, (A.50) follows from $\tilde{c}_k(s) \ge \tilde{\gamma}_{m,j_k+1}$, by construction of $j_k$, and we conclude:

$$V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s) \geq V_m^{\boldsymbol{\pi}^q}(x,s)\ . \tag{A.51}$$

Now let $t \in \{q+1, q+2, \ldots, m-j_k, m-j_k+1\}$ be arbitrary. We have:

$$\begin{aligned}
V_m^{\boldsymbol{\pi}^{t-1}}(x,s) - V_m^{\boldsymbol{\pi}^t}(x,s) &= c\Big((m-t+2)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad + d \cdot h \\
&\quad + \alpha \cdot I\!\!E \left[ \begin{array}{l} V_{m-1}\big((m-t+1)\cdot d, S\big) \\ -V_{m-1}\big((m-t)\cdot d, S\big) \end{array} \right] \\
&\geq c\Big((m-t+2)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad - d \cdot \tilde{\gamma}_{m,m-t+2} \qquad\qquad (A.52) \\
&\geq c\Big((m-t+2)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad - d \cdot \tilde{\gamma}_{m,j_k+1} \qquad\qquad (A.53) \\
&\geq d \cdot \big(\tilde{c}_k(s) - \tilde{\gamma}_{m,j_k+1}\big) \qquad\qquad (A.54) \\
&\geq 0. \qquad\qquad (A.55)
\end{aligned}$$

Equation (A.52) follows from Lemma A.2, with $l = m$, $r = m-t+1 \leq m-q \leq m = l$, and $\eta = d$. Equation (A.53) follows from:

$$t \leq m - j_k + 1 \ \Leftrightarrow\ j_k + 1 \leq m - t + 2 \ \Rightarrow\ \tilde{\gamma}_{m,j_k+1} \geq \tilde{\gamma}_{m,m-t+2}\ .$$

Similarly to (A.46), equation (A.54) follows from the fact that:

$$(m-t+1)\cdot d - x \geq \Big[m - (m - j_k + 1) + 1\Big] \cdot d - x = j_k \cdot d - x \geq \tilde{z}_{k-1}(s)\ .$$

Finally, (A.55) follows once again from $\tilde{c}_k(s) \geq \tilde{\gamma}_{m,j_k+1}$, by construction of $j_k$. Rearranging (A.55) yields:

$$V_m^{\boldsymbol{\pi}^{t-1}}(x,s) \geq V_m^{\boldsymbol{\pi}^t}(x,s),\ \forall t \in \{q+1, q+2, \ldots, m-j_k, m-j_k+1\}\,. \tag{A.56}$$

Noting that $V_m^{\bar{\boldsymbol{\pi}}}(x,s) = V_m^{\boldsymbol{\pi}^{m-j_k+1}}(x,s)$, (A.51) and repeated application of (A.56) imply:

$$
\begin{aligned}
V_m^{\bar{\boldsymbol{\pi}}}(x,s) &= V_m^{\boldsymbol{\pi}^{m-j_k+1}}(x,s) \\
&\leq V_m^{\boldsymbol{\pi}^{m-j_k}}(x,s) \leq \ldots \leq V_m^{\boldsymbol{\pi}^{q+1}}(x,s) \leq V_m^{\boldsymbol{\pi}^{q}}(x,s) \leq V_m^{\boldsymbol{\pi}'(\epsilon)}(x,s) \ ,
\end{aligned}
$$

and we conclude $\bar{\boldsymbol{\pi}}$ is at least as good as $\boldsymbol{\pi}'(\epsilon)$.

Case (c): $j_K \cdot d - \tilde{z}_{\max}(s) < x \leq j_K \cdot d - \tilde{z}_{K-1}(s)$

Same as Case (b) with $K$ replacing $k$.

Case (d): $0 \leq x \leq j_K \cdot d - \tilde{z}_{\max}(s)$

$\bar{z}_m(x,s) = \tilde{z}_{\max}(s)$, the upper bound of the action space, so it is not feasible to transmit more.

Part II: Do not transmit less than suggested by $\bar{\boldsymbol{\pi}}$ at time $m$

Let $\boldsymbol{\pi}''(\epsilon)$ be a feasible strategy with $z''_m = \bar{z}_m - \epsilon$, where $\epsilon > 0$, and the optimal transmission policy at times $m-1$,$m-2,\ldots,1$. To satisfy feasibility, we require $\bar{z}_m - \epsilon \geq \max(0, d-x)$. Define $\eta := \epsilon - \left\lfloor \frac{\epsilon}{d} \right\rfloor \cdot d$, and note that $\eta \in [0,d)$. Let $\boldsymbol{\pi}_\theta^l$ be a strategy that at time $m$ satisfies the demands of periods $m, m-1, \ldots, l$, except for $\theta$ units of the demand of period $l$, where $0 \leq \theta \leq d$, and behaves optimally in slots $m-1, m-2, \ldots, 1$. We consider four exhaustive cases for the current buffer level $x$.

Case (a): $x > j_0 \cdot d$

$\bar{z}_m(x,s) = 0$, the lower bound of the action space, so it is not feasible to transmit less.

Case (b): $j_k \cdot d - \tilde{z}_k(s) < x \leq j_k \cdot d - \tilde{z}_{k-1}(s)$, $k \in \{0,1,\ldots,K\}$, where we define $\tilde{z}_K(s) := \tilde{z}_{\max}(s)$

Define $q := m - j_k + 1 + \left\lfloor \frac{\epsilon}{d} \right\rfloor$. By the feasibility of $\boldsymbol{\pi}''(\epsilon)$ and $\epsilon > 0$, we have

$q \in \{m - j_k + 1, m - j_k + 2, \dots, m-2, m-1\}$. Furthermore, we have:

$$[m-q+1] \cdot d - x = \Big[m - \Big(m - j_k + 1 + \Big\lfloor \frac{\epsilon}{d} \Big\rfloor\Big) + 1\Big] \cdot d - x \le j_k \cdot d - x \le \tilde{z}_k(s) \ ,$$

which, by the construction of $c(\cdot, s)$, implies:

$$c\Big((m-q+1)\cdot d - \eta - x, s\Big) - c\Big((m-q+1)\cdot d - x, s\Big) \ge -\eta \cdot \tilde{c}_k(s) \ . \quad \text{(A.57)}$$

We now compare $\boldsymbol{\pi}^q_\eta$ and $\boldsymbol{\pi}^q_0$:

$$\begin{aligned}
V_m^{\boldsymbol{\pi}^q_\eta}(x,s) - V_m^{\boldsymbol{\pi}^q_0}(x,s) &= c\Big((m-q+1)\cdot d - \eta - x, s\Big) - c\Big((m-q+1)\cdot d - x, s\Big) \\
&\quad -h \cdot \eta \\
&\quad +\alpha \cdot I\!\!E\Big[V_{m-1}\big((m-q)\cdot d - \eta, S\big) - V_{m-1}\big((m-q)\cdot d, S\big)\Big] \\
&\ge c\Big((m-q+1)\cdot d - \eta - x, s\Big) - c\Big((m-q+1)\cdot d - x, s\Big) \\
&\quad +\eta \cdot \tilde{\gamma}_{m,m-q+1} && \text{(A.58)} \\
&\ge c\Big((m-q+1)\cdot d - \eta - x, s\Big) - c\Big((m-q+1)\cdot d - x, s\Big) \\
&\quad +\eta \cdot \tilde{\gamma}_{m,j_k} && \text{(A.59)} \\
&\ge \eta \cdot \Big[\tilde{\gamma}_{m,j_k} - \tilde{c}_k(s)\Big] && \text{(A.60)} \\
&\ge 0. && \text{(A.61)}
\end{aligned}$$

Equation (A.58) follows from Lemma A.3 with $r = m - q + 1 \le m = l$ and $\eta = \eta$. Equation (A.59) follows from:

$$q \ge m - j_k + 1 \ \Leftrightarrow \ m - q + 1 \le j_k \ \Rightarrow \ \tilde{\gamma}_{m,j_k} \le \tilde{\gamma}_{m,m-q+1} \ .$$

Equation (A.60) follows from (A.57). Finally, (A.61) follows from $\tilde{c}_k(s) < \tilde{\gamma}_{m,j_k}$. Rearranging (A.61) yields:

$$V_m^{\boldsymbol{\pi}^q_0}(x,s) \le V_m^{\boldsymbol{\pi}^q_\eta}(x,s) \ . \quad \text{(A.62)}$$

Next, let $t \in \{m - j_k + 1, m - j_k + 2, \ldots, m - 1\}$ be arbitrary. We have:

$$\begin{aligned}
V_m^{\boldsymbol{\pi}_0^{t+1}}(x,s) - V_m^{\boldsymbol{\pi}_0^{t}}(x,s) &= c\Big((m-t)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad -h \cdot d \\
&\quad +\alpha \cdot I\!\!E \left[ \begin{array}{l} V_{m-1}\big((m-t-1)\cdot d, S\big) \\ -V_{m-1}\big((m-t)\cdot d, S\big) \end{array} \right] \\
&\geq c\Big((m-t)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad +d \cdot \tilde{\gamma}_{m,m-t+1} && \text{(A.63)} \\
&\geq c\Big((m-t)\cdot d - x, s\Big) - c\Big((m-t+1)\cdot d - x, s\Big) \\
&\quad +d \cdot \tilde{\gamma}_{m,j_k} && \text{(A.64)} \\
&\geq d \cdot \Big[\tilde{\gamma}_{m,j_k} - \tilde{c}_k(s)\Big] && \text{(A.65)} \\
&\geq 0. && \text{(A.66)}
\end{aligned}$$

Equation (A.63) follows from Lemma A.3 with $r = m - t \leq m = l$ and $\eta = d$. Equation (A.64) follows from:

$$t \geq m - j_k + 1 \;\Leftrightarrow\; m - t + 1 \leq j_k \;\Rightarrow\; \tilde{\gamma}_{m,j_k} \leq \tilde{\gamma}_{m-t+1} \ .$$

Equation (A.65) follows from construction of $c(\cdot, s)$ and the fact that:

$$(m-t+1)\cdot d - x \leq \Big(m - (m - j_k + 1) + 1\Big)\cdot d - x = j_k \cdot d - x \leq \tilde{z}_k(s) \ .$$

Finally, (A.66) follows from $\tilde{c}_k(s) < \tilde{\gamma}_{m,j_k}$. Rearranging (A.66) yields:

$$V_m^{\boldsymbol{\pi}_0^{t}}(x,s) \leq V_m^{\boldsymbol{\pi}_0^{t+1}}(x,s)\ \forall t \in \{m - j_k + 1, m - j_k + 2, \ldots, m - 1\} \ . \qquad \text{(A.67)}$$

Noting that $\bar{\boldsymbol{\pi}} = \boldsymbol{\pi}_0^{m-j_k+1}$, (A.62) and repeated application of (A.67) imply:

$$
\begin{aligned}
V_m^{\bar{\boldsymbol{\pi}}}(x,s) &= V_m^{{\boldsymbol{\pi}_0}^{m-j_k+1}}(x,s) \\
&\leq V_m^{{\boldsymbol{\pi}_0}^{m-j_k+2}}(x,s) \leq \ldots \leq V_m^{\boldsymbol{\pi}_0^q}(x,s) \leq V_m^{\boldsymbol{\pi}_\eta^q}(x,s) \\
&\qquad\qquad\qquad\qquad\qquad\qquad = V_m^{\boldsymbol{\pi}''(\epsilon)}(x,s),
\end{aligned} \tag{A.68}
$$

and we conclude $\bar{\boldsymbol{\pi}}$ is at least as good as $\boldsymbol{\pi}''(\epsilon)$.

<u>Case (c)</u>: $j_k \cdot d - \tilde{z}_{k-1}(s) < x \leq j_{k-1} \cdot d - \tilde{z}_{k-1}(s),\ k \in \{1, \ldots, K\}$

In this case, $\bar{\boldsymbol{\pi}} = \boldsymbol{\pi}_0^{m+1-\frac{x+\tilde{z}_{k-1(s)}}{d}}$. Define $p := m+1-\frac{x+\tilde{z}_{k-1(s)}}{d}$, and $q := p + \left\lfloor \frac{\epsilon}{d} \right\rfloor$.

Again, we start by comparing $\boldsymbol{\pi}_\eta^q$ and $\boldsymbol{\pi}_0^q$:

$$
\begin{aligned}
V_m^{\boldsymbol{\pi}_\eta^q}(x,s) - V_m^{\boldsymbol{\pi}_0^q}(x,s) &= c\Big(z''_m - \eta, s\Big) - c\Big(z''_m, s\Big) \\
&\quad -h \cdot \eta \\
&\quad +\alpha \cdot I\!\!E \left[ \begin{array}{l} V_{m-1}\big((m-q)\cdot d - \eta, S\big) \\ -V_{m-1}\big((m-q)\cdot d, S\big) \end{array} \right] \\
&\geq c\Big(z''_m - \eta, s\Big) - c\Big(z''_m, s\Big) \\
&\quad +\eta \cdot \tilde{\gamma}_{m,m-q+1} && \text{(A.69)} \\
&\geq c\Big(z''_m - \eta, s\Big) - c\Big(z''_m, s\Big) \\
&\quad +\eta \cdot \tilde{\gamma}_{m,j_{k-1}} && \text{(A.70)} \\
&\geq \eta \cdot \Big[\tilde{\gamma}_{m,j_{k-1}} - \tilde{c}_{k-1}(s)\Big] && \text{(A.71)} \\
&\geq 0. && \text{(A.72)}
\end{aligned}
$$

Equation (A.69) follows from Lemma A.3 with $r = m-q+1 \leq m = l$ and $\eta = \eta$. Equation (A.70) follows from:

$$
m-q+1 = m - \left(p + \left\lfloor \frac{\epsilon}{d} \right\rfloor\right) + 1 = \frac{x+\tilde{z}_{k-1(s)}}{d} - \left\lfloor \frac{\epsilon}{d} \right\rfloor \leq \frac{x+\tilde{z}_{k-1(s)}}{d} \leq j_{k-1}\ ,
$$

which implies $\tilde{\gamma}_{m,j_{k-1}} \leq \tilde{\gamma}_{m,m-q+1}$. Equation (A.71) follows from $z''_m < \tilde{z}_{k-1}(s)$ and the construction of $c(\cdot, s)$. Finally, (A.72) follows from $\tilde{c}_{k-1}(s) < \tilde{\gamma}_{m,j_{k-1}}$. Rearranging

(A.72) yields:

$$V_m^{\boldsymbol{\pi}_0^q}(x,s) \leq V_m^{\boldsymbol{\pi}_\eta^q}(x,s)\ . \tag{A.73}$$

Next, let $\hat{t} \in \{p, p+1, \ldots, q-1\}$ be arbitrary. We have:

$$\begin{aligned}
V_m^{\boldsymbol{\pi}_0^{\hat{t}+1}}(x,s) - V_m^{\boldsymbol{\pi}_0^{\hat{t}}}(x,s) &= c\Big((m-\hat{t})\cdot d - x, s\Big) - c\Big((m-\hat{t}+1)\cdot d - x, s\Big) \\
&\quad -h\cdot d \\
&\quad +\alpha \cdot I\!\!E\begin{bmatrix} V_{m-1}\big((m-\hat{t}-1)\cdot d, S\big) \\ -V_{m-1}\big((m-\hat{t})\cdot d, S\big)\end{bmatrix} \\
&\geq c\Big((m-\hat{t})\cdot d - x, s\Big) - c\Big((m-\hat{t}+1)\cdot d - x, s\Big) \\
&\quad +d\cdot \tilde{\gamma}_{m,m-\hat{t}+1} && \text{(A.74)} \\
&\geq c\Big((m-\hat{t})\cdot d - x, s\Big) - c\Big((m-\hat{t}+1)\cdot d - x, s\Big) \\
&\quad +d\cdot \tilde{\gamma}_{m,j_{k-1}} && \text{(A.75)} \\
&\geq d\cdot\Big[\tilde{\gamma}_{m,j_{k-1}} - \tilde{c}_{k-1}(s)\Big] && \text{(A.76)} \\
&\geq 0. && \text{(A.77)}
\end{aligned}$$

Equation (A.74) follows from Lemma A.3 with $r = m - \hat{t} \leq m = l$ and $\eta = d$. Equation (A.75) follows from:

$$\hat{t} \geq p \;\Rightarrow\; m - \hat{t} + 1 \leq m - p + 1 = \frac{x + \tilde{z}_{k-1}(s)}{d} \leq j_{k-1} \;\Rightarrow\; \tilde{\gamma}_{m,j_{k-1}} \leq \tilde{\gamma}_{m-\hat{t}+1}\ .$$

Equation (A.76) follows from construction of $c(\cdot, s)$ and the fact that:

$$(m-\hat{t}+1)\cdot d - x \leq (m-p+1)\cdot d - x = \tilde{z}_{k-1}(s)\ .$$

Finally, (A.77) follows from $\tilde{c}_{k-1}(s) < \tilde{\gamma}_{m,j_{k-1}}$. Rearranging (A.77) yields:

$$V_m^{\boldsymbol{\pi}_0^{\hat{t}}}(x,s) \leq V_m^{\boldsymbol{\pi}_0^{\hat{t}+1}}(x,s)\ \forall \hat{t} \in \{p, p+1, \ldots, q-1\}\ . \tag{A.78}$$

Then (A.73) and repeated application of (A.78) yield:

$$
\begin{aligned}
V_m^{\bar{\boldsymbol{\pi}}}(x,s) &= V_m^{\boldsymbol{\pi}_0^p}(x,s) \\
&\leq V_m^{\boldsymbol{\pi}_0^{p+1}}(x,s) \leq \ldots \leq V_m^{\boldsymbol{\pi}_0^{q-1}}(x,s) \leq V_m^{\boldsymbol{\pi}_0^q}(x,s) \leq V_m^{\boldsymbol{\pi}_\eta^q}(x,s) \\
&\qquad = V_m^{\boldsymbol{\pi}''(\epsilon)}(x,s) \ ,
\end{aligned}
$$

and we conclude $\bar{\boldsymbol{\pi}}$ is at least as good as $\boldsymbol{\pi}''(\epsilon)$.

Case (d): $0 \leq x \leq j_K \cdot d - \tilde{z}_{\max}$

The same argument as Case (c) applies with $k$ replaced by $K+1$ and $\tilde{z}_K(s) = \tilde{z}_{\max}(s)$. This completes Part II.

From Parts I and II, we conclude $\bar{\boldsymbol{\pi}}$ is optimal if the starting queue level is an integer multiple of the demand $d$. By assumption, the starting queue level $x$ at time $N$ is zero. Thus, $\bar{\boldsymbol{\pi}}$ is optimal at time $N$. $z_N^*(x,s) = \bar{z}_N(x,s)$ will also be an integer multiple of demand as $b_{N,k}(s)$, and $\{\tilde{z}_k(s)\}_{k=0,1,\ldots,K}$ are all integer multiples of $d$. It follows that the queue level at the end of slot $N$ (equal to the queue level at the beginning of slot $N-1$), $z_N^*(x,s) - d$, will also be an integer multiple of $d$. Continuing this logic, if the strategy $\bar{\boldsymbol{\pi}}$ is used, the queue level at the beginning of each subsequent time slot will be an integer multiple of demand. Thus, $\bar{\boldsymbol{\pi}}$ is optimal.

□

## A.4 Proof of Theorem 5.8

We prove statements (i)-(v) by joint induction on the time remaining, $n$.

Base Case: $n = 1$

$V_0(\mathbf{x}, \mathbf{s}_0) = 0$, for all $\mathbf{s}_0$, so (i) and (ii) hold trivially. Let $\mathbf{s}_1 \in \mathcal{S}$ be arbitrary. $G_1(\mathbf{y}_1, \mathbf{s}_1) = \mathbf{c}_{\mathbf{s}_1}^{\mathrm{T}} \mathbf{y}_1 + h(\mathbf{y}_1 - \mathbf{d})$, which is convex and supermodular. Thus, (iii) and

(iv) are true. Additionally,

$$G_1(\mathbf{y}_1, \mathbf{s}_1) = \sum_{m=1}^{2} \left\{ c_s^m \cdot y_1^m + h^m \left( y_1^m - d^m \right) \right\} ,$$

so $\inf \left\{ \underset{y_1^2 \in [d^2, \infty)}{\operatorname{argmin}} \left\{ G_1\left(y_1^1, y_1^2, s_1^1, s_1^2\right) \right\} \right\}$ is independent of $y_1^1$, and vice versa. Thus, (v) is true for $n = 1$, completing the base case.

Induction Step

Assume statements (i)-(v) are true for $n = 2, 3, \ldots, l-1$. We want to show they are true for $n = l$. We let $\mathbf{s} \in \mathcal{S}$ be arbitrary, and proceed in order.

(i) Consider two arbitrary points, $\bar{\mathbf{x}}, \tilde{\mathbf{x}} \in I\!R_+^2$. Let $\lambda \in [0, 1]$ be arbitrary, and define $\hat{\mathbf{x}} := \lambda \bar{\mathbf{x}} + (1 - \lambda) \tilde{\mathbf{x}}$. Let $\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s})$, $\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s})$, and $\mathbf{y}^*(\hat{\mathbf{x}}, \mathbf{s})$ be optimal buffer levels after transmission in slot $l - 1$, for each of the respective starting points. We have:

$$\begin{aligned}
\lambda \cdot V_{l-1}(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda) \cdot V_{l-1}(\tilde{\mathbf{x}}, \mathbf{s}) &= -\mathbf{c}_\mathbf{s}^{\mathrm{T}} \hat{\mathbf{x}} + \lambda \cdot G_{l-1}\big(\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}), \mathbf{s}\big) \\
&\quad + (1-\lambda) \cdot G_{l-1}\big(\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\big) \\
&\geq -\mathbf{c}_\mathbf{s}^{\mathrm{T}} \hat{\mathbf{x}} \\
&\quad + G_{l-1}\big(\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda) \mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\big) \\
&\geq -\mathbf{c}_\mathbf{s}^{\mathrm{T}} \hat{\mathbf{x}} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\hat{\mathbf{x}}, \mathbf{s})} \left\{ G_{l-1}(\mathbf{y}, \mathbf{s}) \right\} \\
&= V_{l-1}(\hat{\mathbf{x}}, \mathbf{s}) \\
&= V_{l-1}(\lambda \bar{\mathbf{x}} + (1-\lambda) \tilde{\mathbf{x}}, \mathbf{s}) \ , \qquad \text{(A.79)}
\end{aligned}$$

where the first inequality follows from the convexity of $G_{l-1}(\cdot, \mathbf{s})$ from the induction hypothesis. The second inequality follows from the following argument. $\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})$ implies:

$$\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee \bar{\mathbf{x}} \ \text{ and } \ \mathbf{c}_\mathbf{s}^{\mathrm{T}} \left[ \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \right] \leq P \ . \qquad \text{(A.80)}$$

Similarly, $\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})$ implies:

$$\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee \tilde{\mathbf{x}} \quad \text{and} \quad \mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}\right] \leq P \ . \tag{A.81}$$

Multiplying the equations in (A.80) by $\lambda$ and the equations in (A.81) by $1 - \lambda$, and summing, we have:

$$\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda)\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \succeq \lambda(\mathbf{d} \vee \bar{\mathbf{x}}) + (1-\lambda)(\mathbf{d} \vee \tilde{\mathbf{x}}) \succeq \mathbf{d} \vee \hat{\mathbf{x}}, \tag{A.82}$$

and

$$\begin{aligned} &\mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda)\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \hat{\mathbf{x}}\right] \\ &= \lambda \mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}}\right] + (1-\lambda)\mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}\right] \leq P \ . \end{aligned} \tag{A.83}$$

From (A.82) and (A.83), we conclude $\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda)\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\hat{\mathbf{x}}, \mathbf{s})$, as shown in Figure A.1. Thus, the value of $G_{l-1}(\cdot, \mathbf{s})$ at this point is greater than or equal to the minimum of $G_l(\cdot, \mathbf{s})$ over the region $\tilde{\mathcal{A}}^{\mathbf{d}}(\hat{\mathbf{x}}, \mathbf{s})$. From (A.79), we conclude $V_{l-1}(\cdot, \mathbf{s})$ is convex. This is a similar argument to the one used by Evans to show convexity in [45].

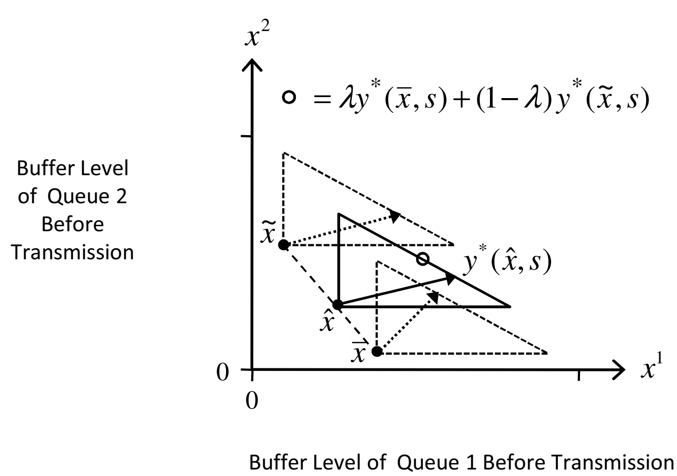


Figure A.1. Diagram showing $\lambda \mathbf{y}^*(\bar{\mathbf{x}}, \mathbf{s}) + (1-\lambda)\mathbf{y}^*(\tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\hat{\mathbf{x}}, \mathbf{s})$ in the proof of the convexity of $V_{l-1}(\cdot, \mathbf{s})$.

(ii) Recall that $V_{l-1}(\mathbf{x},\mathbf{s}) = -\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{x} + \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\}$. The first term, $-\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{x}$, is clearly supermodular in $\mathbf{x}$, so it suffices to show that the second term, $\min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\}$, is also supermodular in $\mathbf{x}$. Let $\bar{\mathbf{x}},\tilde{\mathbf{x}} \in I\!R^2$ be arbitrary. We want to show:

$$\min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\} + \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\}$$
$$\leq \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\} + \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\}\,. \tag{A.84}$$

If $\bar{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ are comparable (i.e., $\tilde{x}^1 \geq \bar{x}^1$ and $\tilde{x}^2 \geq \bar{x}^2$ or $\tilde{x}^1 \leq \bar{x}^1$ and $\tilde{x}^2 \leq \bar{x}^2$), then (A.84) is trivial. So we assume they are not comparable, and also assume without loss of generality that $\bar{x}^1 < \tilde{x}^1$ and $\tilde{x}^2 < \bar{x}^2$. We begin with a quick lemma.

**Lemma A.4.** *There exist optimal buffer levels after transmission in slot $l-1$, $\boldsymbol{y}^*(\bar{\boldsymbol{x}}\wedge\tilde{\boldsymbol{x}},\boldsymbol{s})$ and $\boldsymbol{y}^*(\bar{\boldsymbol{x}}\vee\tilde{\boldsymbol{x}},\boldsymbol{s})$, such that $\boldsymbol{y}^*(\bar{\boldsymbol{x}}\wedge\tilde{\boldsymbol{x}},\boldsymbol{s}) \not\succ \boldsymbol{y}^*(\bar{\boldsymbol{x}}\vee\tilde{\boldsymbol{x}},\boldsymbol{s})$; i.e., such that ${y^*}^1(\bar{\boldsymbol{x}}\wedge\tilde{\boldsymbol{x}},\boldsymbol{s}) \leq {y^*}^1(\bar{\boldsymbol{x}}\vee\tilde{\boldsymbol{x}},\boldsymbol{s})$ or ${y^*}^2(\bar{\boldsymbol{x}}\wedge\tilde{\boldsymbol{x}},\boldsymbol{s}) \leq {y^*}^2(\bar{\boldsymbol{x}}\vee\tilde{\boldsymbol{x}},\boldsymbol{s})$.*

*Proof.* Fix a choice of $\mathbf{y}^*(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s})$ such that

$$G_{l-1}\big(\mathbf{y}^*\left(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}\right),\mathbf{s}\big) = \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s})}\{G_{l-1}(\mathbf{y},\mathbf{s})\}\,.$$

Assume that for all optimal choices of $\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})$, we have $\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) \succ \mathbf{y}^*(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s})$. Fix one such choice of $\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})$, and we have:

$$\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) \succ \mathbf{y}^*(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) \succeq \mathbf{d}\vee\left(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}}\right)\,. \tag{A.85}$$

Further, $\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})$ implies $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\left[\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{\mathbf{x}}\wedge\tilde{\mathbf{x}}\right] \leq P$, and thus:

$$\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\left[\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{\mathbf{x}}\vee\tilde{\mathbf{x}}\right] \leq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\left[\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{\mathbf{x}}\wedge\tilde{\mathbf{x}}\right] \leq P\,. \tag{A.86}$$

Equations (A.85) and (A.86) imply $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, and thus:

$$
\begin{aligned}
& G_{l-1}\big(\mathbf{y}^*\left(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}\right), \mathbf{s}\big) \\
&= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \leq G_{l-1}\big(\mathbf{y}^*\left(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}\right), \mathbf{s}\big).
\end{aligned} \tag{A.87}
$$

However, we also have:

$$
\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{d} \vee (\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}) \succeq \mathbf{d} \vee (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) \ , \tag{A.88}
$$

and

$$
\mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}\right] \leq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \left[\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}\right] \leq P \ . \tag{A.89}
$$

Equations (A.88) and (A.89) imply $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$, which, in combination with (A.87), implies it is optimal to move from $\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}$ to $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, contradicting the assumption that $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \succ \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ for all possible choices of $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$. $\square$

Now let $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ be arbitrary optimal actions such that $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \not\succ \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. We show (A.84) by considering two exhaustive cases.

<u>Case 1</u>: $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \succeq \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$

We start with another lemma.

**Lemma A.5.** *Let $f : [d^1, \infty) \times [d^2, \infty) \to I\!R$ be convex and supermodular, let $\sigma, \beta \in [0,1]$ be arbitrary, and let $\boldsymbol{z} = (z_1, z_2) \preceq (\hat{z}_1, \hat{z}_2) = \hat{\boldsymbol{z}}$. Define*

$$
\boldsymbol{z}^{\lambda_1, \lambda_2} := \Big(\lambda_1 \hat{z}_1 + (1 - \lambda_1) z_1, \lambda_2 \hat{z}_2 + (1 - \lambda_2) z_2\Big).
$$

*Then*

$$
f(\boldsymbol{z}) + f(\hat{\boldsymbol{z}}) \geq f(\boldsymbol{z}^{\sigma, \beta}) + f(\boldsymbol{z}^{1-\sigma, 1-\beta}) \ . \tag{A.90}
$$

*Proof.*

Step 1: Assume $\sigma, \beta \le \frac{1}{2}$. Assume without loss of generality that $\sigma \le \beta$. By the convexity of $f(\cdot)$, we have:

$$f(\mathbf{z}) + f(\hat{\mathbf{z}}) \ge f(\mathbf{z}^{\sigma,\sigma}) + f(\mathbf{z}^{1-\sigma,1-\sigma}) \ , \tag{A.91}$$

and

$$f(\mathbf{z}^{1-\sigma,1-\sigma}) + f(\mathbf{z}^{1-\sigma,\sigma}) \ge f(\mathbf{z}^{1-\sigma,\beta}) + f(\mathbf{z}^{1-\sigma,1-\beta}) \ . \tag{A.92}$$

By the supermodularity of $f(\cdot)$, we have:

$$f(\mathbf{z}^{1-\sigma,\beta}) + f(\mathbf{z}^{\sigma,\sigma}) \ge f(\mathbf{z}^{\sigma,\beta}) + f(\mathbf{z}^{1-\sigma,\sigma}) \ . \tag{A.93}$$

Figure A.2 shows these relationships. Combining (A.91)-(A.93), we have:

$$\begin{aligned} f(\mathbf{z}) + f(\hat{\mathbf{z}}) &\ge f(\mathbf{z}^{\sigma,\sigma}) + f(\mathbf{z}^{1-\sigma,1-\sigma}) \\ &\ge f(\mathbf{z}^{\sigma,\sigma}) + f(\mathbf{z}^{1-\sigma,\beta}) - f(\mathbf{z}^{1-\sigma,\sigma}) + f(\mathbf{z}^{1-\sigma,1-\beta}) \\ &\ge f(\mathbf{z}^{\sigma,\beta}) + f(\mathbf{z}^{1-\sigma,1-\beta}) \ . \end{aligned}$$

Step 2: Now let $\sigma, \beta \in [0,1]$, and define $\hat{\sigma} := \min\{\sigma, 1-\sigma\}$ and $\hat{\beta} := \min\{\beta, 1-\beta\}$. Then $\hat{\sigma}, \hat{\beta} \le \frac{1}{2}$, so by Step 1, we have:

$$f(\mathbf{z}) + f(\hat{\mathbf{z}}) \ge f(\mathbf{z}^{\hat{\sigma},\hat{\beta}}) + f(\mathbf{z}^{1-\hat{\sigma},1-\hat{\beta}}) \ . \tag{A.94}$$

Note that $\mathbf{z}^{\sigma,\beta} \wedge \mathbf{z}^{1-\sigma,1-\beta} = \mathbf{z}^{\hat{\sigma},\hat{\beta}}$, and $\mathbf{z}^{\sigma,\beta} \vee \mathbf{z}^{1-\sigma,1-\beta} = \mathbf{z}^{1-\hat{\sigma},1-\hat{\beta}}$, so by the supermodularity of $f(\cdot)$, we have:

$$f(\mathbf{z}^{\hat{\sigma},\hat{\beta}}) + f(\mathbf{z}^{1-\hat{\sigma},1-\hat{\beta}}) \ge f(\mathbf{z}^{\sigma,\beta}) + f(\mathbf{z}^{1-\sigma,1-\beta}) \ . \tag{A.95}$$

Combining (A.94) and (A.95) yields the desire result, (A.90). □

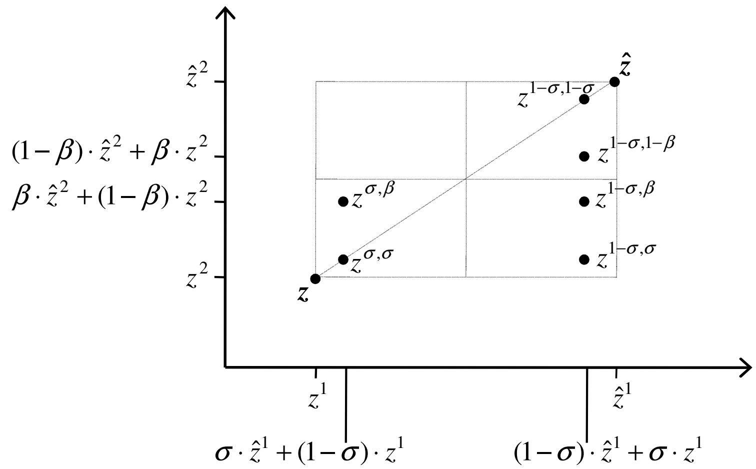


Figure A.2. Diagram of the points referred to in Step 1 of the proof of Lemma A.5.

Next, define the following points, shown in Figure A.3:

$$\bar{\mathbf{y}} := \begin{pmatrix} \bar{x}^1 + \max\left\{y^{*^1}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^1, y^{*^1}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^1\right\}, \\ \bar{x}^2 + \min\left\{y^{*^2}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^2, y^{*^2}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^2\right\} \end{pmatrix}, \text{ and}$$

$$\tilde{\mathbf{y}} := \begin{pmatrix} \tilde{x}^1 + \min\left\{y^{*^1}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^1, y^{*^1}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^1\right\}, \\ \tilde{x}^2 + \max\left\{y^{*^2}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^2, y^{*^2}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^2\right\} \end{pmatrix}.$$

Note that $\bar{\mathbf{y}} \succeq \mathbf{d}\vee\bar{\mathbf{x}}$ and $\tilde{\mathbf{y}} \succeq \mathbf{d}\vee\tilde{\mathbf{x}}$. Furthermore, we have:

$$\begin{aligned}
\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}(\bar{\mathbf{y}}-\bar{\mathbf{x}}) &= \mathbf{c}_{\mathbf{s}}^{\mathrm{T}} \begin{pmatrix} \max\left\{y^{*^1}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^1, y^{*^1}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^1\right\}, \\ \min\left\{y^{*^2}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^2, y^{*^2}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^2\right\} \end{pmatrix} \\
&\le \max\left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\left(y^{*^1}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^1, y^{*^2}(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^2\right), \\ \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\left(y^{*^1}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{x}^1, y^{*^2}(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \bar{x}^2\right) \end{array} \right\} \\
&= \max\left\{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\Big(\mathbf{y}^*(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) - (\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}})\Big), \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\Big(\mathbf{y}^*(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - (\bar{\mathbf{x}}\vee\tilde{\mathbf{x}})\Big)\right\} \\
&\le P.
\end{aligned}$$

By a similar argument, $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}(\tilde{\mathbf{y}}-\tilde{\mathbf{x}}) \le P$, and thus $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}},\mathbf{s})$, and $\tilde{\mathbf{y}} \in$

$\tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})$. So we have:

$$\min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}},\mathbf{s})} \{G_{l-1}(\mathbf{y},\mathbf{s})\} + \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}},\mathbf{s})} \{G_{l-1}(\mathbf{y},\mathbf{s})\}$$
$$\leq G_{l-1}(\bar{\mathbf{y}},\mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}},\mathbf{s}). \tag{A.96}$$

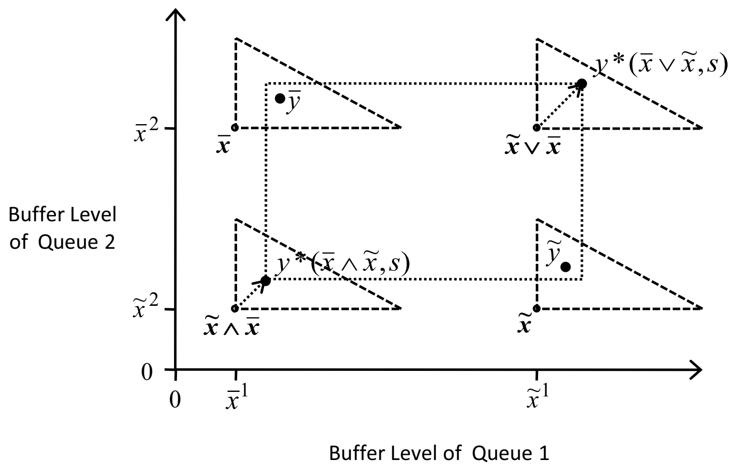


Figure A.3. Construction of feasible points $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ in Case 1 of the proof of supermodularity of $V_{l-1}(\cdot, \mathbf{s})$.

Now define[1]:

$$\sigma := \frac{{y^*}^1(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{y}^1}{{y^*}^1(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - {y^*}^1(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})}\ , \text{ and}$$

$$\beta := \frac{{y^*}^2(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - \tilde{y}^2}{{y^*}^2(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - {y^*}^2(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s})}\ .$$

Rearranging the definitions of $\sigma$ and $\beta$ yields:

$$\tilde{\mathbf{y}} = \begin{pmatrix} (1-\sigma)\cdot {y^*}^1(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) + \sigma\cdot {y^*}^1(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}), \\ (1-\beta)\cdot {y^*}^2(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) + \beta\cdot {y^*}^2(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) \end{pmatrix}.$$

It is also straightforward to check that:

$$\bar{\mathbf{y}} = \begin{pmatrix} \sigma\cdot {y^*}^1(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) + (1-\sigma)\cdot {y^*}^1(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}), \\ \beta\cdot {y^*}^2(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) + (1-\beta)\cdot {y^*}^2(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) \end{pmatrix}.$$

[1] If ${y^*}^1(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - {y^*}^1(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) = 0$, let $\sigma$ be arbitrary in $[0,1]$. Similarly, if ${y^*}^2(\bar{\mathbf{x}}\vee\tilde{\mathbf{x}},\mathbf{s}) - {y^*}^2(\bar{\mathbf{x}}\wedge\tilde{\mathbf{x}},\mathbf{s}) = 0$, let $\beta$ be arbitrary in $[0,1]$.

Note also that

$$
\begin{aligned}
y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) &= \min\left\{y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (\tilde{x}^1 - \tilde{x}^2)\right\} \\
&\leq \min\left\{y^{*^1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (\tilde{x}^1 - \tilde{x}^2)\right\} \\
&= \tilde{y}^1 \\
&\leq y^{*^1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \ ,
\end{aligned}
$$

and thus, $\sigma \in [0,1]$. Similarly, $y^{*^2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq \tilde{y}^2 \leq y^{*^2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$, and thus, $\beta \in [0,1]$.

Since $G_{l-1}(\cdot, \mathbf{s})$ is convex and supermodular, we can now apply Lemma A.5, with $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ playing the role of $\mathbf{z}$; $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ the role of $\hat{\mathbf{z}}$; $\bar{\mathbf{y}}$ the role of $\mathbf{z}^{\sigma,\beta}$; and $\tilde{\mathbf{y}}$ the role of $\mathbf{z}^{1-\sigma,1-\beta}$, to get:

$$
\begin{aligned}
G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) &\leq G_{l-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\Big) + G_{l-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\Big) \\
&= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \left\{G_{l-1}(\mathbf{y}, \mathbf{s})\right\} \\
&\quad + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \left\{G_{l-1}(\mathbf{y}, \mathbf{s})\right\} . \qquad \text{(A.97)}
\end{aligned}
$$

Combining equations (A.96) and (A.97) yields the desired result, (A.84).

Case 2: $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \not\geq \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \not> \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$

There are two possibilities for this case. The first possibility is that $y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*^1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ and $y^{*^2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq y^{*^2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. The second possibility is that $y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \leq y^{*^1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ and $y^{*^2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*^2}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. We show (A.84) under the first possibility, and a symmetric argument can be used to show (A.84) under the second possibility. We have:

$$
y^{*^1}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) > y^{*^1}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \geq \max\left\{(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}})^1, d^1\right\} = \max\left\{\tilde{x}^1, d^1\right\}, \quad \text{(A.98)}
$$

$$
y^{*^2}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \geq \max\left\{(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})^2, d^2\right\} = \max\left\{\tilde{x}^2, d^2\right\}, \qquad \text{(A.99)}
$$

and

$$\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\Big[\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - \tilde{\mathbf{x}}\Big] \leq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\Big[\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) - (\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}})\Big] \leq P \ . \tag{A.100}$$

Equations (A.98), (A.99), and (A.100) imply $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})$. If it also happens that $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})$, then we have:

$$\begin{aligned}
&\min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \\
&\leq G_{l-1}\big(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\big) + G_{l-1}\big(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\big) \\
&= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \, .
\end{aligned}$$

Otherwise, define:

$$\gamma := \frac{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\big[\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \bar{\mathbf{x}}\big] - P}{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\big[\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) - \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})\big]} \ .$$

From $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) \notin \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$, we know:

$$\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}) > \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\bar{\mathbf{x}} + P \geq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}) + P \geq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) \ . \tag{A.101}$$

It is clear from (A.101) that the numerator and denominator of $\gamma$ are positive, and $\gamma \in [0, 1]$. Now define:

$$\begin{aligned}
&\bar{\mathbf{y}} := \gamma \mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + (1-\gamma)\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \text{ and} \\
&\tilde{\mathbf{y}} := (1-\gamma)\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}) + \gamma \mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}).
\end{aligned}$$

It is somewhat tedious but straightforward to show that $\bar{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})$, and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})$. Thus, we have:

$$\begin{aligned}
&\min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \\
&\leq G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}).
\end{aligned} \tag{A.102}$$

In Figure A.4, $\bar{\mathbf{y}}$ is the point where the line segment connecting $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ and $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$ intersects the budget constraint (hypotenuse) of $\tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}}, \mathbf{s})$, and $\tilde{\mathbf{y}}$ is a point along this line segment the same distance from $\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})$ as $\bar{\mathbf{y}}$ is from $\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})$. By the convexity of $G_{l-1}(\cdot, \mathbf{s})$ along this line segment, we have:

$$\begin{aligned} G_{l-1}(\bar{\mathbf{y}}, \mathbf{s}) + G_{l-1}(\tilde{\mathbf{y}}, \mathbf{s}) &\leq G_{l-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\Big) + G_{l-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s}), \mathbf{s}\Big) \\ &= \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \wedge \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\} \\ &\quad + \min_{\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\bar{\mathbf{x}} \vee \tilde{\mathbf{x}}, \mathbf{s})} \{G_{l-1}(\mathbf{y}, \mathbf{s})\}. \end{aligned} \tag{A.103}$$

Combining (A.102) and (A.103) yields the desired result, (A.84).

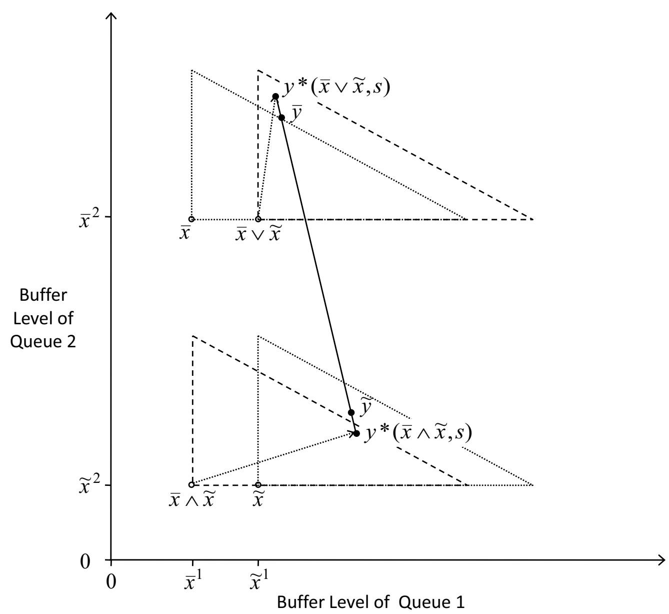


Figure A.4. Construction of feasible points $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ in Case 2 of the proof of supermodularity of $V_{l-1}(\cdot, \mathbf{s})$.

(iii) $G_l(\mathbf{y}, \mathbf{s}) = \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{y} + h(\mathbf{y} - \mathbf{d}) + \alpha \cdot \mathbb{E}\big[V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{S})\big]$. By (i), for all $\mathbf{s}$, $V_{l-1}(\mathbf{x}, \mathbf{s})$ is convex in $\mathbf{x}$; thus, $V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{s})$ is convex in $\mathbf{y}$ as it is the composition of a convex function with an affine function. $\mathbb{E}\big[V_{l-1}(\mathbf{y} - \mathbf{d}, \mathbf{S})\big]$ is also convex as it is the nonnegative weighted sum/integral of convex functions. It follows that $G_l(\mathbf{y}, \mathbf{s})$, the sum of convex functions, is convex in $\mathbf{y}$.

(iv) Supermodularity of $G_l(\mathbf{y}, \mathbf{s})$ follows from the same series of arguments as (iii), because, like convexity, supermodularity is preserved under addition and scalar multiplication (Smith and McCardle refer to these as *closed convex cone properties* [149]).

(v) This step basically follows from Topkis' Theorem 2.8.1 [159, pg. 76], but, for the reader's benefit, we reproduce the proof here with our notation. Let $y^2, \hat{y}^2 \in [d^2, \infty)$ be arbitrary with $y^2 < \hat{y}^2$. Let $\bar{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1,\infty)} \left\{ G_l\left(y^1, y^2, \mathbf{s}\right) \right\}$ and $\tilde{y}^1 \in \operatorname{argmin}_{y^1 \in [d^1,\infty)} \left\{ G_l\left(y^1, \hat{y}^2, \mathbf{s}\right) \right\}$ be arbitrary. We want to show:

$$\bar{y}^1 \wedge \tilde{y}^1 \in \operatorname*{argmin}_{y^1 \in [d^1,\infty)} \left\{ G_l\left(y^1, \hat{y}^2, \mathbf{s}\right) \right\} .$$

If $\tilde{y}^1 \leq \bar{y}^1$, this is trivial, so we check that it is true for $\tilde{y}^1 > \bar{y}^1$. Since $\bar{y}^1$ is a minimizer of $G_l\left(\cdot, y^2, \mathbf{s}\right)$, we have:

$$G_l\left(\bar{y}^1, y^2, \mathbf{s}\right) \leq G_l\left(\tilde{y}^1, y^2, \mathbf{s}\right) \ , \tag{A.104}$$

and since $\tilde{y}^1$ is a minimizer of $G_l\left(\cdot, \hat{y}^2, \mathbf{s}\right)$, we have:

$$G_l\left(\tilde{y}^1, \hat{y}^2, \mathbf{s}\right) \leq G_l\left(\bar{y}^1, \hat{y}^2, \mathbf{s}\right) \ . \tag{A.105}$$

By the supermodularity of $G_l(\cdot, \mathbf{s})$, we have:

$$\begin{aligned} G_l\left(\tilde{y}^1, y^2, \mathbf{s}\right) + G_l\left(\bar{y}^1, \hat{y}^2, \mathbf{s}\right) \ &\leq \ G_l\left(\tilde{y}^1 \wedge \bar{y}^1, y^2 \wedge \hat{y}^2, \mathbf{s}\right) \\ &\quad + G_l\left(\tilde{y}^1 \vee \bar{y}^1, y^2 \vee \hat{y}^2, \mathbf{s}\right) \\ &= \ G_l\left(\bar{y}^1, y^2, \mathbf{s}\right) + G_l\left(\tilde{y}^1, \hat{y}^2, \mathbf{s}\right) \ , \end{aligned}$$

or, rearranging terms:

$$G_l\left(\tilde{y}^1, y^2, \mathbf{s}\right) - G_l\left(\bar{y}^1, y^2, \mathbf{s}\right) \leq G_l\left(\tilde{y}^1, \hat{y}^2, \mathbf{s}\right) - G_l\left(\bar{y}^1, \hat{y}^2, \mathbf{s}\right) \ . \tag{A.106}$$

Combining (A.104), (A.105), and (A.106) yields:

$$0 \leq G_l\left(\tilde{y}^1, y^2, \mathbf{s}\right) - G_l\left(\bar{y}^1, y^2, \mathbf{s}\right) \leq G_l\left(\tilde{y}^1, \hat{y}^2, \mathbf{s}\right) - G_l\left(\bar{y}^1, \hat{y}^2, \mathbf{s}\right) \leq 0. \quad \text{(A.107)}$$

So (A.107) holds with equality throughout, implying

$G_l\left(\tilde{y}^1, \hat{y}^2, \mathbf{s}\right) = G_l\left(\bar{y}^1, \hat{y}^2, \mathbf{s}\right)$, and we conclude:

$$\tilde{y}^1 \wedge \bar{y}^1 = \bar{y}^1 \in \operatorname*{argmin}_{y^1 \in [d^1, \infty)} \left\{ G_l\left(y^1, \hat{y}^2, \mathbf{s}\right) \right\}.$$

Since $\bar{y}^1$ and $\tilde{y}^1$ were chosen arbitrarily, we have:

$$\inf \left\{ \operatorname*{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n\left(y_n^1, y_n^2, s^1, s^2\right) \right\} \right\} \geq \inf \left\{ \operatorname*{argmin}_{y_n^1 \in [d^1, \infty)} \left\{ G_n\left(y_n^1, \hat{y}_n^2, s^1, s^2\right) \right\} \right\}.$$

The first implication in (v) follows from a symmetric argument. $\square$

## A.5 Proof of Theorem 5.9

Let $n \in \{1, 2, \ldots, N\}$ and $\mathbf{s} \in \mathcal{S}$ be arbitrary. We start by proving (5.25). First, let $\mathbf{x} \in \mathcal{R}_I(n, \mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$ be arbitrary. We know from Theorem 5.8 that $G_n(\cdot, \mathbf{s})$ is convex on $[d^1, \infty) \times [d^2, \infty)$, which implies that $G_n(\cdot, \mathbf{s})$ is also convex on any line segment in $[d^1, \infty) \times [d^2, \infty)$ (see, e.g., [122, Theorem 4.1]). Specifically, by the convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^1 = \hat{y}^1$ and the fact that $\hat{y}^2 \geq x^2 \geq f_n^2(\hat{y}^1, \mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n\big((\hat{y}^1, x^2), \mathbf{s}\big) \geq G_n\Big(\big(\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})\big), \mathbf{s}\Big) \ . \quad \text{(A.108)}$$

Similarly, by the convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^2 = x^2$ and the fact that $\hat{y}^1 \geq x^1 \geq f_n^1(x^2, \mathbf{s})$, we have:

$$G_n\big((\hat{y}^1, x^2), \mathbf{s}\big) \geq G_n(\mathbf{x}, \mathbf{s}) \geq G_n\Big(\big(f_n^1(x^2, \mathbf{s}), x^2\big), \mathbf{s}\Big) \ . \quad \text{(A.109)}$$

Combining (A.108) and (A.109) yields:

$$G_n(\hat{\mathbf{y}}, \mathbf{s}) \geq G_n\Big(\big(\hat{y}^1, x^2\big), \mathbf{s}\Big) \geq G_n(\mathbf{x}, \mathbf{s}) \ ,$$

and we conclude $G_n(\mathbf{x},\mathbf{s}) = \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \{G_n(\mathbf{y},\mathbf{s})\}$.

Second, let $\mathbf{x} \in \mathcal{R}_{II}(n,\mathbf{s})$ be arbitrary. Then $\mathbf{b}_n(\mathbf{s}) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ and $\mathbf{b}_n(\mathbf{s})$ is a global minimizer of $G_n(\cdot,\mathbf{s})$, so it is clearly optimal to transmit to bring the receivers' buffer levels up to $\mathbf{b}_n(\mathbf{s})$.

Next, let $\mathbf{x} \in \mathcal{R}_{III-A}(n,\mathbf{s})$ and $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ be arbitrary. By definition of $f_n^1(\cdot,\mathbf{s})$, we have:

$$G_n(\tilde{\mathbf{y}},\mathbf{s}) \geq G_n\Big(\big(f_n^1(\tilde{y}^2,\mathbf{s}),\tilde{y}^2\big),\mathbf{s}\Big)\ . \tag{A.110}$$

Furthermore, the function $\min_{y^1\in[d^1,\infty)}\Big\{G_n\Big((y^1,y^2),\mathbf{s}\Big)\Big\}$ is convex in $y^2$ since $[d^1,\infty)$ is a convex set (see, e.g., [23, pp. 101-102]). Thus, $\tilde{y}^2 \geq x^2 \geq b_n^2(\mathbf{s})$ implies:

$$\begin{aligned} &G_n\Big(\big(f_n^1(\tilde{y}^2,\mathbf{s}),\tilde{y}^2\big),\mathbf{s}\Big) \\ &\geq G_n\Big(\big(f_n^1(x^2,\mathbf{s}),x^2\big),\mathbf{s}\Big) \\ &\geq G_n\Big(\big(f_n^1(b_n^2(\mathbf{s}),\mathbf{s}),b_n^2(\mathbf{s})\big),\mathbf{s}\Big) \\ &= G_n\Big(\mathbf{b}_n(\mathbf{s}),\mathbf{s}\Big)\ . \end{aligned} \tag{A.111}$$

Combining (A.110) and (A.111) yields:

$$G_n(\tilde{\mathbf{y}},\mathbf{s}) \geq G_n\Big(\big(f_n^1(x^2,\mathbf{s}),x^2\big),\mathbf{s}\Big)\ ,$$

and $\mathbf{x} \in \mathcal{R}_{III-A}(n,\mathbf{s})$ implies $\Big(f_n^1(x^2,\mathbf{s}),x^2\Big) \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$. Since $\tilde{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ was arbitrary, we conclude $\mathbf{y}_n^*(\mathbf{x},\mathbf{s}) = \Big(f_n^1(x^2,\mathbf{s}),x^2\Big)$ is optimal.

The optimality of $\mathbf{y}_n^*(\mathbf{x},\mathbf{s}) = \Big(x^1,f_n^2(x^1,\mathbf{s})\Big)$ for $\mathbf{x} \in \mathcal{R}_{III-B}(n,\mathbf{s})$ follows from a symmetric argument, using the convexity of $G_n(\cdot,\mathbf{s})$ along the curve $\Big(x^1,f_n^2(x^1,\mathbf{s})\Big)$.

Finally, we prove (5.26). Define:

$$\mathcal{H}^{\mathbf{d}}(\mathbf{x},\mathbf{s}) := \Big\{\mathbf{y} \in [d^1,\infty)\times[d^2,\infty)\ :\ \mathbf{y} \succeq \mathbf{x} \text{ and } \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] = P\Big\} \subset \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})\ .$$

First, let $\mathbf{x} \in \mathcal{R}_{IV-B}(n,\mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ be arbitrary such that $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$. Define

$$\lambda_0 := \frac{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{b}_n(\mathbf{s}) - \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{x} - P}{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{b}_n(\mathbf{s}) - \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\hat{\mathbf{y}}} \ .$$

Note that $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$ and $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{b}_n(\mathbf{s}) - \mathbf{x}] > P$ imply $\lambda_0 \in (0,1)$. Then define:

$$\tilde{\mathbf{y}} := \lambda_0\hat{\mathbf{y}} + (1-\lambda_0)\mathbf{b}_n(\mathbf{s}) \ .$$

By the convexity of $G_n(\cdot,\mathbf{s})$ along the line segment from $\hat{\mathbf{y}}$ to $\mathbf{b}_n(\mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}},\mathbf{s}) \geq G_n(\tilde{\mathbf{y}},\mathbf{s}) \geq G_n\Big(\mathbf{b}_n(\mathbf{s}),\mathbf{s}\Big) \ .$$

Since $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ was arbitrary, we conclude:

$$\min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})}\Big\{G_n(\mathbf{y},\mathbf{s})\Big\} = \min_{\mathbf{y}\in\mathcal{H}^{\mathbf{d}}(\mathbf{x},\mathbf{s})}\Big\{G_n(\mathbf{y},\mathbf{s})\Big\} \ .$$

Next, let $\mathbf{x} \in \mathcal{R}_{IV-C}(n,\mathbf{s})$ and $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ be arbitrary such that $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$. We consider two exhaustive cases, and for each case, we construct a $\tilde{\mathbf{y}} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ such that $G_n(\tilde{\mathbf{y}},\mathbf{s}) \leq G_n(\hat{\mathbf{y}},\mathbf{s})$.

<u>Case 1</u>: $\hat{y}^2 < f_n^2(\hat{y}^1,\mathbf{s})$ and $\bar{\mathbf{y}} := \Big(\hat{y}^1, f_n^2(\hat{y}^1,\mathbf{s})\Big) \notin \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$

Let $\tilde{\mathbf{y}} := \Big(\hat{y}^1, x^2 + \frac{P - c_{s1}\cdot[\hat{y}^1 - x^1]}{c_{s2}}\Big)$. Then, by the convexity of $G_n(\cdot,\mathbf{s})$ along $y^1 = \hat{y}^1$, the definition of $f_n^2(\hat{y}^1,\mathbf{s})$, and $\hat{y}^2 \leq \tilde{y}^2 \leq f_n^2(\hat{y}^1,\mathbf{s})$, we have:

$$G_n(\bar{\mathbf{y}},\mathbf{s}) = G_n\Big(\big(\hat{y}^1, f_n^2(\hat{y}^1,\mathbf{s})\big),\mathbf{s}\Big) \leq G_n(\tilde{\mathbf{y}},\mathbf{s}) \leq G_n(\hat{\mathbf{y}},\mathbf{s}) \ .$$

It is also straightforward to check that $\tilde{\mathbf{y}} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$, as desired.

<u>Case 2</u>: All other $\hat{\mathbf{y}} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ such that $\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\hat{\mathbf{y}} - \mathbf{x}] < P$

By the definition of $f_n^2(\hat{y}^1,\mathbf{s})$, we have:

$$G_n(\hat{\mathbf{y}},\mathbf{s}) \geq G_n\Big(\big(\hat{y}^1, f_n^2(\hat{y}^1,\mathbf{s})\big),\mathbf{s}\Big) \ . \tag{A.112}$$

Define:

$$\tilde{y}^1 := \sup\Big\{y^1 \in \big[x^1, \hat{y}^1\big) : \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\Big(y^1, f_n^2\left(y^1, \mathbf{s}\right)\Big) \geq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{x} + P\Big\}\ , \text{ and}$$
$$\tilde{y}^2 := \frac{P - c_{s^1} \cdot [\tilde{y}^1 - x^1]}{c_{s^2}}\ .$$

By the convexity of $G_n(\cdot, \mathbf{s})$ along $\Big(y^1, f_n^2\left(y^1, \mathbf{s}\right)\Big)$, we have:

$$G_n\Big(\big(\hat{y}^1, f_n^2(\hat{y}^1, \mathbf{s})\big), \mathbf{s}\Big) \geq G_n\Big(\big(\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})\big), \mathbf{s}\Big)\ . \tag{A.113}$$

Furthermore, we have:

$$G_n\Big(\big(\tilde{y}^1, f_n^2(\tilde{y}^1, \mathbf{s})\big), \mathbf{s}\Big) = G_n\Big(\big(\tilde{y}^1, \tilde{y}^2\big), \mathbf{s}\Big) = G_n\left(\tilde{\mathbf{y}}, \mathbf{s}\right)\ . \tag{A.114}$$

If $\tilde{y}^2 = f_n^2\left(\tilde{y}^1, \mathbf{s}\right)$, (A.114) is trivial. Otherwise, there is a discontinuity in $f_n^2(\cdot, \mathbf{s})$ at $\tilde{y}^1$, and we have:

$$\lim_{y^1 \nearrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right) \geq \tilde{y}^2 \geq \lim_{y^1 \searrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right)\ , \tag{A.115}$$

with at least one of the inequalities being strict. Nonetheless, $G_n\Big(\big(y^1, f_n^2(y^1, \mathbf{s})\big), \mathbf{s}\Big)$ is a continuous function of $y^1$, and therefore:

$$\begin{aligned} G_n\Big(\big(\tilde{y}^1, \lim_{y^1 \nearrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right)\big), \mathbf{s}\Big) &= G_n\Big(\big(\tilde{y}^1, \lim_{y^1 \searrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right)\big), \mathbf{s}\Big) \\ &= G_n\Big(\big(\tilde{y}^1, f_n^2\left(\tilde{y}^1, \mathbf{s}\right)\big), \mathbf{s}\Big). \end{aligned} \tag{A.116}$$

The convexity of $G_n(\cdot, \mathbf{s})$ along the line $y^1 = \tilde{y}^1$ and (A.116) imply:

$$G_n\Big(\big(\tilde{y}^1, y^2\big), \mathbf{s}\Big) = G_n\Big(\big(\tilde{y}^1, f_n^2\left(\tilde{y}^1, \mathbf{s}\right)\big), \mathbf{s}\Big)\ ,\ \forall y^2 \in \left[\lim_{y^1 \searrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right), \lim_{y^1 \nearrow \tilde{y}^1} f_n^2\left(y^1, \mathbf{s}\right)\right],$$

which in combination with (A.115) implies (A.114). Combining (A.112)-(A.114) yields the desired result: $G_n\left(\tilde{\mathbf{y}}, \mathbf{s}\right) \leq G_n\left(\hat{\mathbf{y}}, \mathbf{s}\right)$ for a $\tilde{\mathbf{y}} \in \mathcal{H}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$.

The validity of (5.26) for $\mathbf{x} \in \mathcal{R}_{IV-A}(n, \mathbf{s})$ follows from a symmetric argument, completing the proof of (5.26) and Theorem 5.9. $\square$

# *Appendix B*

## Infinite Horizon Discounted Expected Cost Proofs for Problem (P5.2)

### B.1 Proof of Theorem 5.5

Our line of analysis is similar in spirit to [48], [69, Chapter 8], and [77]. Let $x \in I\!R_+$ and $s \in \mathcal{S}$ be arbitrary. First, we show inductively that $V_1(x,s) \leq V_2(x,s) \leq \ldots \leq V_n(x,s) \leq V_{n+1}(x,s) \leq \ldots$.

Base Case: $n = 1$

$$
\begin{aligned}
V_1(x,s) &= \min_{\left\{\max(0,d-x)\leq z\leq \tilde{z}_{\max}(s)\right\}} \left\{c(z,s) + h(x+z-d)\right\} \\
&\leq \min_{\left\{\max(0,d-x)\leq z\leq \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ +\alpha \cdot I\!E\big[V_1(x+z-d,S_1) \;\big|\; S_2 = s\big] \end{array} \right\} \\
&= V_2(x,s) \ ,
\end{aligned}
$$

where the inequality follows from $V_1(x,s) \geq 0, \ \forall x, \forall s$.

Induction Step: Assume $V_n(x,s) \leq V_{n+1}(x,s)$ for $n = 1, 2, \ldots, m-1$. We show it is true for $n = m$:

$$
V_m(x,s) = \min_{\left\{\max(0,d-x)\leq z\leq \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ +\alpha \cdot I\!E\big[V_{m-1}(x+z-d,S_{m-1}) \;\big|\; S_m = s\big] \end{array} \right\}
$$

$$
\begin{aligned}
&\le \min_{\left\{\max(0,d-x)\le z\le \tilde{z}_{\max}(s)\right\}} \left\{ \begin{array}{l} c(z,s)+h(x+z-d) \\ +\alpha\cdot \mathbb{E}\left[V_m(x+z-d,S_m)\;\middle|\; S_{m+1}=s\right] \end{array} \right\} \\
&= V_{m+1}(x,s)\ ,
\end{aligned}
$$

where the inequality follows from the induction hypothesis and the homogeneity of the Markov process representing the channel condition. So, for every $x \in \mathbb{R}_+$ and $s \in \mathcal{S}$, $\{V_n(x,s)\}_{n=1,2,\ldots}$ is a nondecreasing sequence.

Next, consider a policy $\boldsymbol{\pi}^d$ transmitting $d$ packets in every slot, regardless of channel condition. Define:

$$
\tilde{c}_{\max} := \sup_{\substack{s\in\mathcal{S} \\ k\in\{0,1,\ldots,K\}}} \{\tilde{c}_k(s)\} < \infty \ . \tag{B.1}
$$

Then we have:

$$
V_n(x,s) \le V_n^{\boldsymbol{\pi}^d}(x,s) \le \Big(\tilde{c}_{\max}\cdot d + h(x)\Big)\frac{1-\alpha^n}{1-\alpha} \le \Big(\tilde{c}_{\max}\cdot d + h(x)\Big)\frac{1}{1-\alpha} < \infty \ ,
$$

so $\{V_n(x,s)\}_{n=1,2,\ldots}$ is a bounded nondecreasing sequence, implying $\lim_{n\to\infty} V_n(x,s)$ exists and is finite, $\forall x \in \mathbb{R}_+, \forall s \in \mathcal{S}$.

We now move on to part (b). Recall from Section A.1 that $V_n(x,s)$ is convex in $x$, for all $n$ and all $s$. Define $V_\infty(x,s) := \lim_{n\to\infty} V_n(x,s)$. Let $s \in \mathcal{S}$ be arbitrary, but fixed. $V_\infty(x,s) = \sup_{n\in\mathbb{N}} V_n(x,s)$, so $V_\infty(x,s)$ is convex in $x$ as it is the pointwise supremum of the convex functions $\{V_n(x,s)\}_{n=1,2,\ldots}$.

Define $\tilde{g}_\infty : [d,\infty)\times\mathcal{S} \to \mathbb{R}_+$ by

$$
\begin{aligned}
\tilde{g}_\infty(y,s) &:= h(y-d) + \alpha\cdot\mathbb{E}\left[V_\infty(y-d,S') \mid S=s\right] \\
&= h(y-d) + \alpha\cdot\mathbb{E}\left[\lim_{n\to\infty} V_n(y-d,S') \mid S=s\right] \\
&= h(y-d) + \alpha\cdot\lim_{n\to\infty}\mathbb{E}\left[V_n(y-d,S') \mid S=s\right] \qquad \text{(B.2)} \\
&= \lim_{n\to\infty} \tilde{g}_n(y,s)\ ,
\end{aligned}
$$

where (B.2) follows from the homogeneity of the Markov process representing the channel condition and the Monotone Convergence Theorem. Furthermore, for each $s \in \mathcal{S}$, $\tilde{g}_\infty(y,s)$ is convex in $y$ and $\lim_{y\to\infty} \tilde{g}_\infty(y,s) \geq \lim_{y\to\infty} h(y-d) = \infty$. Thus, for every $s$, at least one finite number achieves the global minimum of $\tilde{g}_\infty(y,s)$.

Next, we proceed to part (d), and let $s \in \mathcal{S}$ be arbitrary. Define $b_{\infty,-1}(s) := \infty$ and

$$b_{\infty,k}(s) := \max\Big\{d, \inf\big\{b \mid \tilde{g}'^{+}_{\infty}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} \ , \ \forall k \in \{0,1,\ldots,K\} \ .$$

Clearly, $b_{\infty,-1}(s) = \lim_{n\to\infty} b_{n,-1}(s)$, as $b_{n,-1}(s) := \infty$ for every $n$. Let $k \in \{0,1,\ldots,K\}$ be arbitrary. We want to show:

$$\begin{aligned}
\lim_{n\to\infty} b_{n,k}(s) &= \lim_{n\to\infty} \max\Big\{d, \inf\big\{b \mid \tilde{g}'^{+}_{n}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} \\
&= \max\Big\{d, \inf\big\{b \mid \tilde{g}'^{+}_{\infty}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} := b_{\infty,k}(s) \ .
\end{aligned}$$

By the continuity of $\max\{d,\cdot\}$, it suffices to show:

$$\lim_{n\to\infty} \Big\{\inf\big\{b \mid \tilde{g}'^{+}_{n}(b,s) \geq -\tilde{c}_k(s)\big\}\Big\} = \inf\big\{b \mid \tilde{g}'^{+}_{\infty}(b,s) \geq -\tilde{c}_k(s)\big\} \ . \tag{B.3}$$

Before proceeding to show (B.3), we present a lemma due to Sobel [151, Lemma 3, pg. 732], which is also presented in [69, Lemma 8-5, pg. 425].

**Lemma B.1** (Sobel, 1971)**.** *Let $g, g_1, g_2, \ldots$ be convex functions on an open convex subset $X$ of $I\!R$ such that $g_n(x) \to g(x)$ as $n \to \infty$ and $g_n(x) \leq g_{n+1}(x)$ for all $n$ and $x$. Let $g'^{-}_n(x)$ and $g'^{-}(x)$ denote derivatives from the left and $g'^{+}_n(x)$ and $g'^{+}(x)$ denote derivatives from the right. Then for all $x \in X$:*

$$g'^{-}(x) \leq \liminf_{n\to\infty} g'^{-}_n(x) \leq \limsup_{n\to\infty} g'^{+}_n(x) \leq g'^{+}(x) \ . \tag{B.4}$$

We now prove (B.3) by contradiction. Define:

$$\hat{b}_{n,k}(s) \ := \ \inf\big\{b \mid \tilde{g}'^{+}_{n}(b,s) \geq -\tilde{c}_k(s)\big\}, \text{ and}$$

$$\hat{b}_{\infty,k}(s) \;:=\; \inf\left\{b \;\middle|\; \tilde{g}'^{+}_{\infty}(b,s) \geq -\tilde{c}_k(s)\right\} .$$

First, assume $\liminf_{n\to\infty} \hat{b}_{n,k}(s) < \hat{b}_{\infty,k}(s)$, so there exists an $x_0 \in \mathbb{R}_+$ such that $d < x_0 < \hat{b}_{\infty,k}(s)$, and a sequence $\{n_i\}_{i=1,2,\ldots}$ such that $\lim_{i\to\infty} \hat{b}_{n_i,k}(s) = x_0$. Then we have:

$$\begin{aligned} -\tilde{c}_k(s) &\leq \lim_{i\to\infty} \tilde{g}'^{+}_{n_i}(x_0,s) && \text{(B.5)} \\ &\leq \limsup_{n\to\infty} \tilde{g}'^{+}_{n}(x_0,s) \\ &\leq \tilde{g}'^{+}_{\infty}(x_0,s) \; . && \text{(B.6)} \end{aligned}$$

Here, (B.5) follows from $\lim_{i\to\infty} \hat{b}_{n_i,k}(s) = x_0$, and the fact that $\tilde{g}'^{+}_{n}(\cdot,s)$ is continuous from the right. Equation (B.6) follows from Lemma B.1. Yet, $\tilde{g}'^{+}_{\infty}(x_0,s) \geq -\tilde{c}_k(s)$ implies $\hat{b}_{\infty,k}(s) \leq x_0$, which is a contradiction. We conclude:

$$\liminf_{n\to\infty} \hat{b}_{n,k}(s) \geq \hat{b}_{\infty,k}(s). \tag{B.7}$$

Next, assume $\limsup_{n\to\infty} \hat{b}_{n,k}(s) > \hat{b}_{\infty,k}(s) \geq d$, and define:

$$x_1 := \frac{\limsup_{n\to\infty} \hat{b}_{n,k}(s) + \hat{b}_{\infty,k}(s)}{2} \; .$$

Then we have:

$$\begin{aligned} -\tilde{c}_k(s) &\leq \tilde{g}'^{+}_{\infty}\left(\hat{b}_{\infty,k}(s), s\right) && \text{(B.8)} \\ &\leq \tilde{g}'^{-}_{\infty}\left(x_1, s\right) && \text{(B.9)} \\ &\leq \liminf_{n\to\infty} \tilde{g}'^{-}_{n}\left(x_1, s\right) && \text{(B.10)} \\ &\leq \liminf_{n\to\infty} \tilde{g}'^{+}_{n}\left(x_1, s\right) && \text{(B.11)} \end{aligned}$$

Here, (B.8) follows from the fact that $\tilde{g}'^{+}_{\infty}(\cdot,s)$ is continuous from the right; (B.10) follows from Lemma B.1[1]; and (B.9) and (B.11) follow from the fact (see, e.g., [122,

[1] One hypothesis of Lemma B.1 is that all functions are defined on an *open* convex subset of $\mathbb{R}$. While our functions $\tilde{g}_\infty(\cdot,s)$ and $\{\tilde{g}_n(\cdot,s)\}_{n\in\mathbb{N}}$ are defined on $[d,\infty)$, we only apply Lemma B.1 at the points $x_0, x_1 \in (d,\infty)$. Thus, equations (B.6) and (B.10) follow from the application of Lemma B.1 to the restrictions of the functions $\tilde{g}_\infty(\cdot,s)$ and $\{\tilde{g}_n(\cdot,s)\}_{n\in\mathbb{N}}$ to the domain of $(d,\infty)$.

pg. 228]) that for a proper convex function $f$ on $I\!R$, $z_1 < x < z_2$ implies:

$$f'^{+}(z_1) \leq f'^{-}(x) \leq f'^{+}(x) \leq f'^{-}(z_2) \ .$$

$\liminf\limits_{n\to\infty} \tilde{g}_n'^{+}\big(x_1, s\big) \geq -\tilde{c}_k(s)$ implies that for every sequence $\{n_j\}_{j=1,2,\ldots}$, we have:

$$\lim_{j\to\infty} \tilde{g}_{n_j}'^{+}\big(x_1, s\big) \geq -\tilde{c}_k(s) \ ,$$

and, in turn:

$$\lim_{j\to\infty} \hat{b}_{n_j,k}(s) \leq x_1 \ .$$

Therefore, $\limsup\limits_{n\to\infty} \hat{b}_{n,k}(s) \leq x_1$, which is a contradiction. We conclude:

$$\limsup_{n\to\infty} \hat{b}_{n,k}(s) \leq \hat{b}_{\infty,k}(s) \ . \tag{B.12}$$

Equations (B.7) and (B.12) imply (B.3).

We are now ready to prove parts (e) and (f) of Theorem 5.5. Define

$$z_\infty^*(x,s) := \begin{cases} \tilde{z}_{k-1}(s), & \text{if } \ b_{\infty,k}(s) - \tilde{z}_{k-1}(s) < x \leq b_{\infty,k-1}(s) - \tilde{z}_{k-1}(s) \ , \\ & \quad k \in \{0, 1, \ldots, K\} \\ b_{\infty,k}(s) - x, & \text{if } \ b_{\infty,k}(s) - \tilde{z}_k(s) < x \leq b_{\infty,k}(s) - \tilde{z}_{k-1}(s) \ , \\ & \quad k \in \{0, 1, \ldots, K-1\} \\ b_{\infty,K}(s) - x, & \text{if } \ b_{\infty,K}(s) - \tilde{z}_{\max}(s) < x \leq b_{\infty,K}(s) - \tilde{z}_{K-1}(s) \\ \tilde{z}_{\max}(s), & \text{if } \ 0 \leq x \leq b_{\infty,K}(s) - \tilde{z}_{\max}(s) \end{cases}$$

Clearly, $\lim\limits_{n\to\infty} b_{n,k}(s) = b_{\infty,k}(s)$ implies

$$\lim_{n\to\infty} z_n^*(x,s) = z_\infty^*(x,s), \ \forall x \in I\!R+, \forall s \in \mathcal{S} \ .$$

Furthermore, $\tilde{g}_n(y,s) \to \tilde{g}_\infty(y,s)$ and $z_n^*(x,s) \to z_\infty^*(x,s)$ as $n \to \infty$ imply:

$$\lim_{n\to\infty} \tilde{g}_n\big(x + z_n^*(x,s)\big) = \tilde{g}_\infty\big(x + z_\infty^*(x,s)\big), \ \forall x \in I\!R+, \forall s \in \mathcal{S} \ . \tag{B.13}$$

So for all $x \in I\!\!R+$ and $s \in \mathcal{S}$, we have:

$$\begin{aligned}
V_\infty(x,s) &= \lim_{n\to\infty} V_n(x,s) \\
&= \lim_{n\to\infty} \min_{\{\max(0,d-x)\le z\le \tilde{z}_{\max}(s)\}} \Big\{ c(z,s) + \tilde{g}_n(x+z,s) \Big\} \\
&= \lim_{n\to\infty} \Big\{ c\big(z_n^*(x,s),s\big) + \tilde{g}_n\big(x+z_n^*(x,s),s\big) \Big\} && \text{(B.14)} \\
&= c\big(z_\infty^*(x,s),s\big) + \tilde{g}_\infty\big(x+z_\infty^*(x,s),s\big) && \text{(B.15)} \\
&= \min_{\{\max(0,d-x)\le z\le \tilde{z}_{\max}(s)\}} \Big\{ c(z,s) + \tilde{g}_\infty(x+z,s) \Big\} && \text{(B.16)} \\
&= \min_{\{\max(0,d-x)\le z\le \tilde{z}_{\max}(s)\}} \left\{ \begin{array}{l} c(z,s) + h(x+z-d) \\ +\alpha \cdot I\!\!E\big[V_\infty(x+z-d,S') \;\big|\; S=s\big] \end{array} \right\}.
\end{aligned}$$

Equation (B.14) follows from Theorem 5.1, and (B.15) follows from (B.13) and the continuity of $c(\cdot,s)$. Equation (B.16) follows from the same line of analysis as part (ii) of the induction step in the proof of Theorem 5.3, with $\tilde{g}_\infty(\cdot,s)$, $b_{\infty,k}(s)$, and $z_\infty^*(\cdot,s)$ replacing $\tilde{g}_m(\cdot,s)$, $b_{m,k}(s)$, and $z_m^*(\cdot,s)$, respectively. Thus, $V_\infty(\cdot,\cdot)$, the limit of the finite horizon value functions, satisfies the $\alpha$-DCOE (5.20) and is also equal to the infinite horizon discounted expected cost-to-go resulting from the stationary policy $\boldsymbol{\pi}_\infty^* := (z_\infty^*, z_\infty^*, \ldots)$. We conclude $\boldsymbol{\pi}_\infty^*$, the natural extension of the finite horizon optimal policy, is optimal for the infinite horizon problem (see, for example, [20, Propositions 9.12 and 9.16]). $\square$

## B.2 Proof of Theorem 5.10

We follow the same line of analysis as the proof of Theorem 5.5. Let $\mathbf{x} \in I\!\!R_+^2$ and $\mathbf{s} \in \mathcal{S}$ be arbitrary. First, we show inductively that $V_1(\mathbf{x},\mathbf{s}) \le V_2(\mathbf{x},\mathbf{s}) \le \ldots \le V_n(\mathbf{x},\mathbf{s}) \le V_{n+1}(\mathbf{x},\mathbf{s}) \le \ldots$.

Base Case: $n = 1$

$$\begin{aligned} V_1(\mathbf{x},\mathbf{s}) &= \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{\mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}+\mathbf{z}-\mathbf{d})\right\} \\ &\le \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}+\mathbf{z}-\mathbf{d}) \\ +\alpha \cdot I\!\!E\big[V_1(\mathbf{x}+\mathbf{z}-\mathbf{d},\mathbf{S}_1) \bigm| \mathbf{S}_2 = \mathbf{s}\big] \end{array} \right\} \\ &= V_2(\mathbf{x},\mathbf{s}) \ , \end{aligned}$$

where the inequality follows from $V_1(\mathbf{x},\mathbf{s}) \ge 0,\ \forall\mathbf{x},\forall\mathbf{s}$.

Induction Step: Assume $V_n(\mathbf{x},\mathbf{s}) \le V_{n+1}(\mathbf{x},\mathbf{s})$ for $n = 1, 2, \ldots, m-1$. We show it is true for $n = m$:

$$\begin{aligned} V_m(\mathbf{x},\mathbf{s}) &= \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}+\mathbf{z}-\mathbf{d}) \\ +\alpha \cdot I\!\!E\big[V_{m-1}(\mathbf{x}+\mathbf{z}-\mathbf{d},\mathbf{S}_{m-1}) \bigm| \mathbf{S}_m = \mathbf{s}\big] \end{array} \right\} \\ &\le \min_{\mathbf{z}\in\mathcal{A}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}\mathbf{z} + h(\mathbf{x}+\mathbf{z}-\mathbf{d}) \\ +\alpha \cdot I\!\!E\big[V_m(\mathbf{x}+\mathbf{z}-\mathbf{d},\mathbf{S}_m) \bigm| \mathbf{S}_{m+1} = \mathbf{s}\big] \end{array} \right\} \\ &= V_{m+1}(\mathbf{x},\mathbf{s}) \ , \end{aligned}$$

where the inequality follows from the induction hypothesis and the homogeneity of the Markov process representing the channel condition. So, for every $\mathbf{x} \in I\!\!R_+^2$ and $\mathbf{s} \in \mathcal{S}$, $\{V_n(\mathbf{x},\mathbf{s})\}_{n=1,2,\ldots}$ is a nondecreasing sequence.

Next, consider a policy $\boldsymbol{\pi}^{\mathbf{d}}$ transmitting $d^1$ packets to user 1 and $d^2$ packets to user 2 in every slot, regardless of channel condition. Define:

$$\mathbf{c}_{\max}^{\mathrm{T}} := \left(c_{\max}^1, c_{\max}^2\right)^{\mathrm{T}},\ \text{where } c_{\max}^i := \sup_{s^i\in\mathcal{S}^i} \{c_{s^i}\} < \infty \ . \tag{B.17}$$

Then we have:

$$V_n(\mathbf{x},\mathbf{s}) \le V_n^{\boldsymbol{\pi}^{\mathbf{d}}}(\mathbf{x},\mathbf{s}) \le \left(\mathbf{c}_{\max}^{\mathrm{T}}\mathbf{d} + h(\mathbf{x})\right)\frac{1-\alpha^n}{1-\alpha} \le \left(\mathbf{c}_{\max}^{\mathrm{T}}\mathbf{d} + h(\mathbf{x})\right)\frac{1}{1-\alpha} < \infty \ ,$$

so $\{V_n(\mathbf{x},\mathbf{s})\}_{n=1,2,\ldots}$ is a bounded nondecreasing sequence, implying $\lim_{n\to\infty} V_n(\mathbf{x},\mathbf{s})$ exists and is finite, $\forall\mathbf{x} \in I\!\!R_+^2, \forall\mathbf{s} \in \mathcal{S}$.

Next, recall from Theorem 5.8 that $V_n(\mathbf{x},\mathbf{s})$ is convex and supermodular in $\mathbf{x}$, for all $n$ and all $\mathbf{s}$. Define $V_\infty(\mathbf{x},\mathbf{s}) := \lim_{n\to\infty} V_n(\mathbf{x},\mathbf{s})$. Let $\mathbf{s} \in \mathcal{S}$ be arbitrary, but fixed. $V_\infty(\mathbf{x},\mathbf{s}) = \sup_{n\in I\!N} V_n(\mathbf{x},\mathbf{s})$, so $V_\infty(\mathbf{x},\mathbf{s})$ is convex in $\mathbf{x}$ as it is the pointwise supremum of the convex functions $\{V_n(\mathbf{x},\mathbf{s})\}_{n=1,2,\ldots}$. Furthermore, the pointwise limit of supermodular functions is supermodular (see, e.g., [159, Lemma 2.6.1]), so $V_\infty(\mathbf{x},\mathbf{s})$ is also supermodular in $\mathbf{x}$.

Define $G_\infty : [d^1,\infty) \times [d^2,\infty) \times \mathcal{S} \to I\!R_+$ by

$$\begin{aligned} G_\infty(\mathbf{y},\mathbf{s}) &:= \mathbf{c}_\mathbf{s}^\mathrm{T}\mathbf{y} + h(\mathbf{y}-\mathbf{d}) + \alpha \cdot I\!E\big[V_\infty(\mathbf{y}-\mathbf{d},\mathbf{S}') \bigm| \mathbf{S}=\mathbf{s}\big] \\ &= \mathbf{c}_\mathbf{s}^\mathrm{T}\mathbf{y} + h(\mathbf{y}-\mathbf{d}) + \alpha \cdot I\!E\big[\lim_{n\to\infty} V_n(\mathbf{y}-\mathbf{d},\mathbf{S}') \bigm| \mathbf{S}=\mathbf{s}\big] \\ &= \mathbf{c}_\mathbf{s}^\mathrm{T}\mathbf{y} + h(\mathbf{y}-\mathbf{d}) + \alpha \cdot \lim_{n\to\infty} I\!E\big[V_n(\mathbf{y}-\mathbf{d},\mathbf{S}') \bigm| \mathbf{S}=\mathbf{s}\big] \qquad \text{(B.18)} \\ &= \lim_{n\to\infty} G_n(\mathbf{y},\mathbf{s}) \ , \qquad \text{(B.19)} \end{aligned}$$

where (B.18) follows from the homogeneity of the Markov process representing the channel condition and the Monotone Convergence Theorem. Furthermore, for each $\mathbf{s} \in \mathcal{S}$, $G_\infty(\mathbf{y},\mathbf{s})$ is convex and supermodular in $\mathbf{y}$ as it is the sum of an affine function of $\mathbf{y}$, a convex separable function of $\mathbf{y}-\mathbf{d}$ and a weighted sum of the convex supermodular functions $V_\infty(\mathbf{y}-\mathbf{d},\mathbf{s}')$. Additionally, $\lim\limits_{||\mathbf{y}||\to\infty} G_\infty(y,s) \geq \lim\limits_{||\mathbf{y}||\to\infty} \mathbf{c}_\mathbf{s}^\mathrm{T}\mathbf{y} = \infty$. Thus, for every $\mathbf{s}$, at least one finite vector achieves the global minimum of $G_\infty(\mathbf{y},\mathbf{s})$; $\mathcal{B}_\infty(\mathbf{s})$ is a non-empty closed convex set; and $\mathbf{b}_\infty(\mathbf{s})$, $f_\infty^1(\cdot,\mathbf{s})$, and $f_\infty^2(\cdot,\mathbf{s})$ are well-defined. The structure of the optimal policy outlined in (b) then follows from the same line of analysis used to prove the the structure of the optimal policy in the induction step of Theorem 5.9.

Moreover, since for a fixed $\mathbf{s} \in \mathcal{S}$ and $x^2 \in [d^2,\infty)$,

$$f_n^1(x^2,\mathbf{s}) := \inf\left\{\underset{y^1\in[d^1,\infty)}{\operatorname{argmin}}\left\{G_n\left(y^1,x^2,s^1,s^2\right)\right\}\right\} = \inf\left\{b^1 \bigm| G_n'^{+}\left(b^1,x^2,s^1,s^2\right) \geq 0\right\} \ ,$$

the convergence of $f_n^1(x^2,\mathbf{s})$ to $f_\infty^1(x^2,\mathbf{s})$ follows from the same argument used to show (B.3). The convergence of $f_n^2(x^1,\mathbf{s})$ to $f_\infty^2(x^1,\mathbf{s})$ follows from a symmetric argument.

For all $\mathbf{s}\in\mathcal{S}$ and $x^1\in[d^1,\infty)$, define:

$$\Psi_n(x^1,\mathbf{s}) := \min_{x^2\in[d^2,\infty)}\left\{G_n(x^1,x^2,s^1,s^2)\right\} = G_n\big(x^1,f_n^2(x^1,\mathbf{s}),s^1,s^2\big)\ ,\ \forall n\in I\!N,$$

and

$$\Psi_\infty(x^1,\mathbf{s}) := \min_{x^2\in[d^2,\infty)}\left\{G_\infty(x^1,x^2,s^1,s^2)\right\} = G_\infty\big(x^1,f_\infty^2(x^1,\mathbf{s}),s^1,s^2\big)\ .$$

For fixed but arbitrary $x^1$ and $\mathbf{s}$, $f_n^2(x^1,\mathbf{s})$ converges to $f_\infty^2(x^1,\mathbf{s})$, and, by Dini's Theorem, $G_n(x^1,\cdot,\mathbf{s})$ converges to $G_\infty(x^1,\cdot,\mathbf{s})$ uniformly on a compact interval containing $f_\infty^2(x^1,\mathbf{s})$. Thus, $\Psi_n(x^1,\mathbf{s})$ converges pointwise to $\Psi_\infty(x^1,\mathbf{s})$. Moreover, for every $\mathbf{s}$, $\{\Psi_n(x^1,\mathbf{s})\}_{n\in I\!N}$ and $\Psi_\infty(x^1,\mathbf{s})$ are all convex in $x^1$ with the limit as $x^1$ approaches infinity equal to infinity. Therefore, by the same argument used to show (B.3), $b_n^1(\mathbf{s})$ converges pointwise to $b_\infty^1(\mathbf{s})$.

For all $\mathbf{s}\in\mathcal{S}$ and $x^2\in[d^2,\infty)$, define:

$$\tilde{\Psi}_n(x^2\mathbf{s}) := G_n\big(b_n^1(\mathbf{s}),x^2,s^1,s^2\big)\ ,\ \forall n\in I\!N,$$

and

$$\tilde{\Psi}_n(x^2\mathbf{s}) := G_\infty\big(b_\infty^1(\mathbf{s}),x^2,s^1,s^2\big)\ .$$

For fixed but arbitrary $x^2$ and $\mathbf{s}$, $b_n^1(\mathbf{s})$ converges to $b_\infty^1(\mathbf{s})$, and, by Dini's Theorem, $G_n(\cdot,x^2,\mathbf{s})$ converges to $G_\infty(\cdot,x^2,\mathbf{s})$ uniformly on a compact interval around $b_\infty^1(\mathbf{s})$. Thus, $\tilde{\Psi}_n(x^2,\mathbf{s})$ converges pointwise to $\tilde{\Psi}_\infty(x^2,\mathbf{s})$. Moreover, for every $\mathbf{s}$, $\{\tilde{\Psi}_n(x^2,\mathbf{s})\}_{n\in I\!N}$ and $\tilde{\Psi}_\infty(x^2,\mathbf{s})$ are all convex in $x^2$ with the limit as $x^2$ approaches infinity equal to infinity. Therefore, by the same argument used to show (B.3), $b_n^2(\mathbf{s})$ converges pointwise to $b_\infty^2(\mathbf{s})$, and we conclude $\mathbf{b}_\infty(\mathbf{s}) = \lim_{n\to\infty}\mathbf{b}_n(\mathbf{s})$. $\square$

# *Appendix C*

## Infinite Horizon Average Expected Cost Proofs for Problem (P5.3)

In this appendix, we prove Theorem 5.11 using the vanishing discount approach (see, e.g., [68]). The proof of Theorem 5.6 is nearly identical, and we note the few key differences.

Substituting (5.28) and (5.30) into the $\alpha$-DCOE (5.27) and rearranging yields:

$$\begin{aligned}
&(1-\alpha)\cdot m_{\infty,\alpha} + w_{\infty,\alpha}(\mathbf{x},\mathbf{s}) \\
&= \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \Big\{ \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] + \mathbf{h}\left(\mathbf{y}-\mathbf{d}\right) + \alpha\cdot \mathbb{E}\big[w_{\infty,\alpha}(\mathbf{y}-\mathbf{d},\mathbf{S}') \;\big|\; \mathbf{S}=\mathbf{s}\big]\Big\}, \\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall \mathbf{x}\in\mathbb{R}_+^2, \forall \mathbf{s}\in\mathcal{S}. \qquad \text{(C.1)}
\end{aligned}$$

The main idea of the vanishing discount approach is to take the limit as $\alpha$ goes to 1, and show that (C.1) converges to the ACOE (5.31).

We start by presenting five conditions from the literature on the vanishing discount approach.

**Condition (G).** $\rho := \inf_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \inf_{\substack{\mathbf{x}\in\mathbb{R}_+^2\\ \mathbf{s}\in\mathcal{S}}} \left\{ \limsup_{N\to\infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}}(\mathbf{x},\mathbf{s}) \right\} < \infty.$

**Condition (W). (i)** The state space $\mathbb{R}_+^2 \times \mathcal{S}$ is a locally compact space with countable base.

**(ii)** The action space $\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ is a non-empty compact subset of the state space $I\!R_+^2 \times \mathcal{S}$, and the multifunction $\phi : (\mathbf{x},\mathbf{s}) \mapsto \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})$ is upper semicontinuous; that is, $\phi^{-1}(F)$ is closed in $I\!R_+^2 \times \mathcal{S}$ for every closed set $F \subset I\!R_+^2$.

**(iii)** The transition law is weakly continuous (see, e.g., [68, Appendix C]).

**(iv)** The one-stage cost $c(\mathbf{z},\mathbf{s})+h(\mathbf{x}+\mathbf{z}-\mathbf{d})$ is lower semicontinuous and nonnegative.

**Condition (B).** $\sup_{\alpha<1} w_{\infty,\alpha}(\mathbf{x},\mathbf{s}) < \infty$ for all $\mathbf{x} \in I\!R_+^2$ and $\mathbf{s} \in \mathcal{S}$.

**Condition (B2).** There is a measurable function $\bar{\kappa} : I\!R_+^2 \times \mathcal{S} \to I\!R_+$ such that $\bar{\kappa} \geq w_{\infty,\alpha}$ for all $\alpha \in [0,1)$, and:

$$I\!E\left[\bar{\kappa}(\mathbf{y}-\mathbf{d},\mathbf{S}') \;\middle|\; \mathbf{S}=\mathbf{s}\right] < \infty,\ \forall(\mathbf{x},\mathbf{s}) \in I\!R_+^2 \times \mathcal{S},\ \forall \mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})\ . \tag{C.2}$$

**Condition (E).** For every increasing sequence of discount factors $\{\alpha(l)\}_{l=1,2,\ldots}$ approaching 1, the sequence $\left\{w_{\infty,\alpha(l)}\right\}_{l=1,2,\ldots}$ is equicontinuous.

We show below that our model satisfies these five conditions, but first we show how they lead to Theorem 5.11. Parts (b), (c), and (e) of Theorem 5.11 follow directly from the following theorem due to Schäl [130, Theorem 3.8] and adapted to our notation.

**Theorem C.1** (Schäl, 1993)**.** *Suppose conditions (G), (W), and (B) hold. Then the minimum average cost* $\rho^* = \inf_{\boldsymbol{\pi}\in\mathbf{\Pi}} \inf_{\substack{\boldsymbol{x}\in I\!R_+^2 \\ \boldsymbol{s}\in\mathcal{S}}} \left\{ \limsup_{N\to\infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}}(\boldsymbol{x},\boldsymbol{s}) \right\} = \lim_{\alpha\nearrow 1} (1-\alpha)\cdot m_{\infty,\alpha}$. *Moreover, there exists an optimal selector* $\boldsymbol{y}^*_{\infty,1}(\cdot,\cdot)$ *such that:*

$$\begin{aligned} \rho^* + w_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) &\geq \min_{\boldsymbol{y}\in\tilde{\mathcal{A}}^{\boldsymbol{d}}(\boldsymbol{x},\boldsymbol{s})} \left\{ \begin{array}{l} \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}[\boldsymbol{y}-\boldsymbol{x}] + \boldsymbol{h}(\boldsymbol{y}-\boldsymbol{d}) \\ +I\!E\left[w_{\infty,1}(\boldsymbol{y}-\boldsymbol{d},\boldsymbol{S}') \;\middle|\; \boldsymbol{S}=\boldsymbol{s}\right] \end{array} \right\} \\ &= \boldsymbol{c}_{\boldsymbol{s}}^{\mathrm{T}}\left[\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{x}\right] + \boldsymbol{h}\left(\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{d}\right) \\ &\quad + I\!E\left[w_{\infty,1}\left(\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x},\boldsymbol{s}) - \boldsymbol{d},\boldsymbol{S}'\right)\middle|\boldsymbol{S}=\boldsymbol{s}\right],\ \forall \boldsymbol{x}\in I\!R_+^2,\ \forall \boldsymbol{s}\in\mathcal{S}, \end{aligned} \tag{C.3}$$

*where for every* $(\boldsymbol{x}, \boldsymbol{s}) \in I\!R_+^2 \times \mathcal{S}$ *and any increasing sequence of discount factors* $\{\alpha(l)\}_{l=1,2,\ldots}$ *approaching 1,*

$$w_{\infty,1}(\boldsymbol{x}, \boldsymbol{s}) := \liminf_{l\to\infty} w_{\infty,\alpha(l)}(\boldsymbol{x}, \boldsymbol{s}) \ . \tag{C.4}$$

*Furthermore, for every* $(\boldsymbol{x}, \boldsymbol{s}) \in I\!R_+^2 \times \mathcal{S}$ *and any increasing sequence of discount factors* $\{\alpha(l)\}_{l=1,2,\ldots}$ *approaching 1, there exists a subsequence* $\{\alpha(l_i)\}_{i=1,2,\ldots}$ *approaching 1 and a sequence* $\{\boldsymbol{x}(i)\}_{i=1,2,\ldots}$ *approaching* $\boldsymbol{x}$ *such that:*

$$\boldsymbol{y}^*_{\infty,1}(\boldsymbol{x}, \boldsymbol{s}) \ = \lim_{i\to\infty} \boldsymbol{y}^*_{\infty,\alpha(l_i)}(\boldsymbol{x}(i), \boldsymbol{s}) \ .$$

To get the opposite inequality from (C.3), we use a method from [50] and [111, Theorem 4.1] (which is presented in [68, Section 5.5]). Namely, for every $\mathbf{x} \in I\!R_+^2$, $\mathbf{s} \in \mathcal{S}$, $\mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s})$, and $\alpha(l)$ from (C.4), (C.1) implies:

$$\begin{aligned}&(1-\alpha(l)) \cdot m_{\infty,\alpha(l)} + w_{\infty,\alpha(l)}(\mathbf{x}, \mathbf{s}) \\ &\leq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] + \mathbf{h}\left(\mathbf{y}-\mathbf{d}\right) + \alpha(l) \cdot I\!E\big[w_{\infty,\alpha(l)}(\mathbf{y}-\mathbf{d}, \mathbf{S}') \;\big|\; \mathbf{S}=\mathbf{s}\big] \ .\end{aligned} \tag{C.5}$$

Furthermore, in combination with Conditions (B) and (E), the Arzelá-Ascoli Theorem implies there exists a subsequence $\{\alpha(l_i)\}_{i=1,2,\ldots}$ of $\{\alpha(l)\}_{l=1,2,\ldots}$ such that:

$$w_{\infty,1}(\mathbf{x}, \mathbf{s}) = \lim_{i\to\infty} w_{\infty,\alpha(l_i)}(\mathbf{x}, \mathbf{s}) \ , \forall \mathbf{x} \in I\!R_+^2, \ \forall \mathbf{s} \in \mathcal{S} \ . \tag{C.6}$$

Then, taking the limit of (C.5) as $\alpha$ goes to 1 along the sequence $\{\alpha(l_i)\}_{i=1,2,\ldots}$, (5.29), (C.6), Condition (B2), and the Lebesgue Dominated Convergence Theorem imply:

$$\begin{gathered}\rho^* + w_{\infty,1}(\mathbf{x}, \mathbf{s}) \leq \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] + \mathbf{h}\left(\mathbf{y}-\mathbf{d}\right) + I\!E\big[w_{\infty,1}(\mathbf{y}-\mathbf{d}, \mathbf{S}') \;\big|\; \mathbf{S}=\mathbf{s}\big] \ , \\ \forall \mathbf{x} \in I\!R_+^2, \ \forall \mathbf{s} \in \mathcal{S}, \ \forall \mathbf{y} \in \tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x}, \mathbf{s}) \ ,\end{gathered}$$

which implies:

$$\rho^* + w_{\infty,1}(\mathbf{x},\mathbf{s}) \leq \min_{\mathbf{y}\in\tilde{\mathcal{A}}^{\mathbf{d}}(\mathbf{x},\mathbf{s})} \left\{ \begin{array}{l} \mathbf{c}_{\mathbf{s}}^{\mathrm{T}}[\mathbf{y}-\mathbf{x}] + \mathbf{h}\,(\mathbf{y}-\mathbf{d}) \\ + \mathbb{E}\big[w_{\infty,1}(\mathbf{y}-\mathbf{d},\mathbf{S}') \,\big|\, \mathbf{S}=\mathbf{s}\big] \end{array} \right\}, \;\; \forall \mathbf{x} \in \mathbb{R}_+^2,\; \forall \mathbf{s} \in \mathcal{S} \; . \tag{C.7}$$

Equations (C.3) and (C.7) yield the ACOE (5.31). Moreover, from (C.6) and the fact that convexity and supermodularity are preserved under pointwise limits, we conclude that for every $\mathbf{s} \in \mathcal{S}$, $w_{\infty,1}(\mathbf{x},\mathbf{s})$ is convex and supermodular in $\mathbf{x}$. Then, by the same argument as the one used in Theorems 5.9 and 5.10, there exists an optimal stationary policy with the same structure as statement (b) in Theorem 5.10 that minimizes the right hand side of the ACOE.

Thus, it just remains to show our model satisfies the five conditions. We proceed in order, beginning with Condition (G). Consider again the policy $\boldsymbol{\pi}^{\mathbf{d}}$ transmitting $d^1$ packets to user 1 and $d^2$ packets to user 2 in every slot, regardless of channel condition. Let the initial vector of buffer levels $\mathbf{x}_0 = (0,0)$, and let the initial vector of channel conditions $\mathbf{s}_0$ be arbitrary. Then we have:

$$\rho := \inf_{\boldsymbol{\pi}\in\boldsymbol{\Pi}} \inf_{\substack{x\in\mathbb{R}_+^2 \\ s\in\mathcal{S}}} \left\{ \limsup_{N\to\infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}}(\mathbf{x},\mathbf{s}) \right\} \leq \limsup_{N\to\infty} \frac{1}{N} V_{N,1}^{\boldsymbol{\pi}^{\mathbf{d}}}(\mathbf{x}_0,\mathbf{s}_0) \leq \mathbf{c}_{\max}^{\mathrm{T}} \mathbf{d} < \infty \; ,$$

where $\mathbf{c}_{\max}^{\mathrm{T}}$ is defined in (B.17).[1]

The only nontrivial statement in Condition (W) is the weak continuity of the transition law. Let $\{\mathbf{x}_i\}_{i=1,2,\ldots}$, $\{\mathbf{s}_i\}_{i=1,2,\ldots}$, and $\{\mathbf{y}_i\}_{i=1,2,\ldots}$ be sequences approaching $\mathbf{x}, \mathbf{s}$, and $\mathbf{y}$, respectively, and let $\Gamma$ be a bounded, continuous function on $\mathbb{R}_+^2 \times \mathcal{S}$. We need to show:

$$\lim_{i\to\infty} \mathbb{E}\left[\Gamma\left(\mathbf{X}',\mathbf{S}'\right) | \mathbf{X}=\mathbf{x}_i, \mathbf{S}=\mathbf{s}_i, \mathbf{Y}=\mathbf{y}_i\right] = \mathbb{E}\left[\Gamma\left(\mathbf{X}',\mathbf{S}'\right) | \mathbf{X}=\mathbf{x}, \mathbf{S}=\mathbf{s}, \mathbf{Y}=\mathbf{y}\right] \; .$$

[1] For the proof of Theorem 5.6, we use $\tilde{c}_{\max}$ defined in (B.1) instead.

This is true, as

$$
\begin{aligned}
&\lim_{i\to\infty} \mathbb{E}\left[\Gamma\left(\mathbf{X}',\mathbf{S}'\right)|\mathbf{X}=\mathbf{x}_i,\mathbf{S}=\mathbf{s}_i,\mathbf{Y}=\mathbf{y}_i\right]\\
&=\lim_{i\to\infty}\sum_{\mathbf{s}'\in\mathcal{S}}\Pr\left(\mathbf{S}'=\mathbf{s}' \mid \mathbf{S}=\mathbf{s}_i\right)\cdot\Gamma\left(\mathbf{y}_i-\mathbf{d},\mathbf{s}'\right)\\
&=\sum_{\mathbf{s}'\in\mathcal{S}}\left[\lim_{i\to\infty}\Pr\left(\mathbf{S}'=\mathbf{s}' \mid \mathbf{S}=\mathbf{s}_i\right)\right]\cdot\left[\lim_{i\to\infty}\Gamma\left(\mathbf{y}_i-\mathbf{d},\mathbf{s}'\right)\right]\\
&=\sum_{\mathbf{s}'\in\mathcal{S}}\Pr\left(\mathbf{S}'=\mathbf{s}' \mid \mathbf{S}=\mathbf{s}\right)\cdot\Gamma\left(\mathbf{y}-\mathbf{d},\mathbf{s}'\right)\\
&=\mathbb{E}\left[\Gamma\left(\mathbf{X}',\mathbf{S}'\right)|\mathbf{X}=\mathbf{x},\mathbf{S}=\mathbf{s},\mathbf{Y}=\mathbf{y}\right]\ .
\end{aligned}
$$

Next, we prove Conditions (B) and (B2). Let $\alpha\in[0,1)$ be arbitrary. For every $\mathbf{s}\in\mathcal{S}$, $V_{\infty,\alpha}(\mathbf{x},\mathbf{s})$ is convex in $\mathbf{x}$, and

$$
\lim_{||\mathbf{x}||\to\infty} V_{\infty,\alpha}(\mathbf{x},\mathbf{s}) \geq \lim_{||\mathbf{x}||\to\infty} h(\mathbf{x}-\mathbf{d})=\infty\ ,
$$

so there exists an $\mathbf{x}^*(\mathbf{s})\in\mathbb{R}^2_+$ such that:

$$
\min_{\mathbf{x}\in\mathbb{R}^2_+}\left\{V_{\infty,\alpha}(\mathbf{x},\mathbf{s})\right\}=V_{\infty,\alpha}\big(\mathbf{x}^*(\mathbf{s}),\mathbf{s}\big)\ .
$$

Define:

$$
\mathbf{s}^* := \operatorname*{argmin}_{\mathbf{s}\in\mathcal{S}}\left\{V_{\infty,\alpha}(\mathbf{x}^*(\mathbf{s}),\mathbf{s})\right\}\ ,
$$

so that

$$
m_{\infty,\alpha}=V_{\infty,\alpha}(\mathbf{x}^*(\mathbf{s}^*),\mathbf{s}^*)\ .
$$

Define also the stationary policy $\breve{\boldsymbol{\pi}}=(\breve{\mathbf{y}},\breve{\mathbf{y}},\ldots)$, where $\breve{\mathbf{y}}(\mathbf{x},\mathbf{s}):=\Big(\breve{y}^1(x^1,s^1),\breve{y}^2(x^2,s^2)\Big)$, and for $m\in\{1,2\}$,

$$
\breve{y}^m(x^m,s^m):=\begin{cases}
x^m, & \text{if } \ x^{m^*}(\mathbf{s}^*)+d^m\leq x^m\\
x^{m^*}(\mathbf{s}^*)+d^m, & \text{if } \ x^{m^*}(\mathbf{s}^*)+d^m-\frac{\frac{P}{2}}{c_{s^m}}\leq x^m< x^{m^*}(\mathbf{s}^*)+d^m\\
x^m+\frac{\frac{P}{2}}{c_{s^m}}, & \text{if } \ x^m< x^{m^*}(\mathbf{s}^*)+d^m-\frac{\frac{P}{2}}{c_{s^m}}
\end{cases}\ .
$$

The stationary policy $\breve{\boldsymbol{\pi}}$ calls for the scheduler to allocate at most $\frac{P}{2}$ units of power for transmission to each user, and tries to bring receiver $m$'s buffer towards $x^{m^*}(\mathbf{s}^*) + d^m$ (before transmission), regardless of the random channel conditions.[2] For $m \in \{1,2\}$, let $\tau^m(x^m, s^m)$ be the random number of time slots until receiver $m$'s buffer level at the beginning of a slot reaches $x^{m^*}(\mathbf{s}^*)$ under policy $\breve{\boldsymbol{\pi}}$, starting from state $(x^m, s^m)$. Define also $\tau_{\max}(\mathbf{x}, \mathbf{s}) := \max\{\tau^1(x^1, s^1), \tau^2(x^2, s^2)\}$, and $\tau_{\min} := \min\{\tau^1(x^1, s^1), \tau^2(x^2, s^2)\}$. Note that if $x^m > x^{m^*}(\mathbf{s}^*)$, then $\tau^m(x^m, s^m) = \left\lceil \frac{x - x^{m^*}(\mathbf{s}^*)}{d^m} \right\rceil$, and the total discounted expected transmission and holding cost associated with receiver $m$ for the first $\tau^m(x^m, s^m)$ slots is upper bounded by:

$$\alpha^{\tau^m(x^m,s^m)-1} \cdot c^m_{\max} \cdot d^m + \sum_{t=1}^{\tau^m(x^m,s^m)} \alpha^{t-1} \cdot h^m\left(x - t \cdot d^m\right)$$
$$\leq c^m_{\max} \cdot d^m + \sum_{t=1}^{\left\lceil \frac{x - x^{m^*}(\mathbf{s}^*)}{d^m} \right\rceil} h^m\left(x - t \cdot d^m\right) \ . \tag{C.8}$$

On the other hand, if $x^m \leq x^{m^*}(\mathbf{s}^*)$, $\mathbb{E}[\tau^m(x^m, s^m)]$ is finite.[3] Therefore, by Wald's Lemma, the total discounted expected transmission and holding cost associated with receiver $m$ for the first $\tau^m(x^m, s^m)$ slots is upper bounded by:

$$\sum_{t=1}^{\tau^m(x^m,s^m)} \alpha^{t-1} \cdot \left[\frac{P}{2} + h^m\left(x^{m^*}(\mathbf{s}^*)\right)\right] \leq \mathbb{E}[\tau^m(x^m, s^m)] \cdot \left[\frac{P}{2} + h^m\left(x^{m^*}(\mathbf{s}^*)\right)\right] . \tag{C.9}$$

So for $m \in \{1,2\}$, we define:

$$\bar{\kappa}^m(x^m, s^m) := \begin{cases} c^m_{\max} \cdot d^m + \sum_{t=1}^{\left\lceil \frac{x - x^{m^*}(\mathbf{s}^*)}{d^m} \right\rceil} h^m\left(x - t \cdot d^m\right), & \text{if} \ \ x^{m^*}(\mathbf{s}^*) < x^m \\ \mathbb{E}[\tau^m(x^m, s^m)] \cdot \left[\frac{P}{2} + h^m\left(x^{m^*}(\mathbf{s}^*)\right)\right], & \text{if} \ \ x^m \leq x^{m^*}(\mathbf{s}^*) \end{cases} .$$

Next, let $\tau_{\text{switch}}(\mathbf{x}, \mathbf{s})$ be the random number of time slots until the state $\left(\mathbf{x}^*(\mathbf{s}^*), \mathbf{s}^*\right)$ is reached at the beginning of a slot under policy $\breve{\boldsymbol{\pi}}$, starting from state $(\mathbf{x}, \mathbf{s})$. We

[2] For the proof of Theorem 5.6, the policy $\breve{\boldsymbol{\pi}}$ calls for the scheduler to allocate the full $P$ units of power for transmission to the single receiver when its buffer is below $x^*(s^*) + d$. The bounds are adjusted accordingly.

[3] In order to guarantee $\mathbb{E}[\tau^m(x^m, s^m)]$ is finite, we actually need an additional assumption that $\Pr\left(\frac{\frac{P}{2}}{c^m_{\max}} = d^m\right) < 1$. However, this assumption is harmless, for if it is not true, the channel condition does not vary over time, a scenario outside of our scope of interest.

define a new policy $\bar{\boldsymbol{\pi}}$ that follows $\breve{\boldsymbol{\pi}}$ for $\tau_{\text{switch}}(\mathbf{x},\mathbf{s})$ slots (a random stopping time), and then behaves optimally. Then we have:

$$V_{\infty,\alpha}(\mathbf{x},\mathbf{s}) \leq V^{\bar{\boldsymbol{\pi}}}_{\infty,\alpha}(\mathbf{x},\mathbf{s}) \leq \bar{\kappa}(\mathbf{x},\mathbf{s}) + V_{\infty,\alpha}\big(\mathbf{x}^*(\mathbf{s}^*),\mathbf{s}^*\big) \ , \tag{C.10}$$

where

$$\begin{aligned}\bar{\kappa}(\mathbf{x},\mathbf{s}) :=& \bar{\kappa}^1(x^1,s^1) + \bar{\kappa}^2(x^2,s^2) \\ &+ I\!\!E\big[\tau_{\text{switch}}(\mathbf{x},\mathbf{s}) - \tau_{\min}(\mathbf{x},\mathbf{s})\big] \cdot \big[\mathbf{c}^{\mathrm{T}}_{\max}\mathbf{d} + h^1\big(x^{1^*}(\mathbf{s}^*)\big) + h^2\big(x^{2^*}(\mathbf{s}^*)\big)\big] \ .\end{aligned} \tag{C.11}$$

The third term in (C.11) is an upper bound on the transmission and holding costs required to keep the vector of buffer levels at $\mathbf{x}^*(\mathbf{s}^*)$ while waiting for the vector of channel condition realizations to reach $\mathbf{s}^*$. Since the vector of channel conditions is a finite-state ergodic Markov process, this quantity is finite. Equation (C.10) implies:

$$\begin{aligned} w_{\infty,\alpha}(\mathbf{x},\mathbf{s}) &= V_{\infty,\alpha}(\mathbf{x},\mathbf{s}) - m_{\infty,\alpha} \\ &= V_{\infty,\alpha}(\mathbf{x},\mathbf{s}) - V_{\infty,\alpha}(\mathbf{x}^*(\mathbf{s}^*),\mathbf{s}^*) \\ &\leq \bar{\kappa}(\mathbf{x},\mathbf{s}) < \infty \ . \end{aligned}$$

The important thing to note here is that the bounding function $\bar{\kappa}(\mathbf{x},\mathbf{s})$ is independent of $\alpha$, so Condition (B) holds. The function $\bar{\kappa}(\mathbf{x},\mathbf{s})$ is also measurable and satisfies (C.2), so Condition (B2) also holds.

Finally, Condition (E) follows from the fact that for every $l \in \{1,2,\ldots\}$ and $\mathbf{s} \in \mathcal{S}$, $w_{\infty,\alpha(l)}(\cdot,\mathbf{s})$ is convex. Thus, by the finiteness of $\mathcal{S}$ and essentially the same argument used by Fernández-Gaucherand, Marcus, and Arapostathis in [50, pp. 178-179], $\big\{w_{\infty,\alpha(l)}(\cdot,\cdot)\big\}_{l=1,2,\ldots}$ is locally equi-Lipschitzian and equicontinuous. □

# *Appendix D*

## Proof of Proposition 7.2

We prove (7.2) for the submodular case. The other statements follow from symmetric arguments. First, assume $f(\cdot,\cdot)$ is $dv(\mathbf{c},2)$-submodular; i.e., for every $\mathbf{x} \in I\!R^2$, $\Delta_1 > 0$, $\Delta_2 > 0$, and $\hat{\Delta}_2 := \frac{c^2}{c^1} \cdot \Delta_2$, we have:

$$\begin{aligned} & f(x^1 + \hat{\Delta}_2, x^2) + f(x^1 + \Delta_1, x^2 + \Delta_2) \\ & \leq f(x^1, x^2 + \Delta_2) + f(x^1 + \Delta_1 + \hat{\Delta}_2, x^2) \ . \end{aligned} \tag{D.1}$$

Rearranging (D.1) yields:

$$\begin{aligned} & f(x^1 + \Delta_1 + \hat{\Delta}_2, x^2) - f(x^1 + \hat{\Delta}_2, x^2) \\ & \geq f(x^1 + \Delta_1, x^2 + \Delta_2) - f(x^1, x^2 + \Delta_2) \ . \end{aligned} \tag{D.2}$$

Add $f(x^1, x^2) - f(x^1 + \Delta_1, x^2)$ to both sides of (D.2), and divide both sides by $\Delta_1 \cdot \hat{\Delta}_2$ to get:

$$\begin{aligned} & \frac{\Big[f(x^1 + \Delta_1 + \hat{\Delta}_2, x^2) - f(x^1 + \hat{\Delta}_2, x^2)\Big] - \Big[f(x^1 + \Delta_1, x^2) - f(x^1, x^2)\Big]}{\Delta_1 \cdot \hat{\Delta}_2} \\ & \geq \frac{\big[f(x^1 + \Delta_1, x^2 + \Delta_2) - f(x^1 + \Delta_1, x^2)\big] - \big[f(x^1, x^2 + \Delta_2) - f(x^1, x^2)\big]}{\Delta_1 \cdot \hat{\Delta}_2} \ . \end{aligned} \tag{D.3}$$

Now, take the limits of (D.3) as $\Delta_1$ and $\hat{\Delta}_2$ go to 0:

$$\begin{aligned}
&\frac{\partial^2 f}{\partial x^1 \partial x^1} \\
&= \lim_{\Delta_1 \to 0, \hat{\Delta}_2 \to 0} \frac{\Big[f(x^1+\Delta_1+\hat{\Delta}_2, x^2) - f(x^1+\hat{\Delta}_2, x^2)\Big] - \Big[f(x^1+\Delta_1, x^2) - f(x^1, x^2)\Big]}{\Delta_1 \cdot \hat{\Delta}_2} \\
&\geq \lim_{\Delta_1 \to 0, \hat{\Delta}_2 \to 0} \frac{\Big[f(x^1+\Delta_1, x^2+\Delta_2) - f(x^1+\Delta_1, x^2)\Big] - \Big[f(x^1, x^2+\Delta_2) - f(x^1, x^2)\Big]}{\Delta_1 \cdot \hat{\Delta}_2} \\
&= \lim_{\Delta_1 \to 0, \Delta_2 \to 0} \frac{c^1}{c^2} \cdot \frac{\Big[f(x^1+\Delta_1, x^2+\Delta_2) - f(x^1+\Delta_1, x^2)\Big] - \Big[f(x^1, x^2+\Delta_2) - f(x^1, x^2)\Big]}{\Delta_1 \cdot \Delta_2} \\
&= \frac{c^1}{c^2} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^2} \ ,
\end{aligned}$$

where the second to last equality follows from the substitution $\hat{\Delta}_2 = \frac{c^2}{c^1} \cdot \Delta_2$.

Next, assume

$$\frac{\partial^2 f}{\partial x^1 \partial x^1} \geq \frac{c^1}{c^2} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^2}, \ \forall \mathbf{x} \in I\!R^2 \ . \tag{D.4}$$

Let $\mathbf{y} \in I\!R^2$, $\Delta_1 > 0$, and $\Delta_2 > 0$ be arbitrary, and define $\hat{\Delta}_2 := \frac{c^2}{c^1} \cdot \Delta_2$. Then we have:

$$\begin{aligned}
& f(y^1+\hat{\Delta}_2, y^2) + f(y^1+\Delta_1, y^2+\Delta_2) - f(y^1, y^2+\Delta_2) - f(y^1+\hat{\Delta}_2+\Delta_1, y^2) \\
&= f\big(g(y^1+\hat{\Delta}_2, y^2)\big) + f\big(g(y^1+\hat{\Delta}_2+\Delta_1, y^2+\Delta_2)\big) \\
&\quad - f\big(g(y^1+\hat{\Delta}_2, y^2+\Delta_2)\big) - f\big(g(y^1+\hat{\Delta}_2+\Delta_1, y^2)\big) \\
&= h(y^1+\hat{\Delta}_2, y^2) + h(y^1+\hat{\Delta}_2+\Delta_1, y^2+\Delta_2) \\
&\quad - h(y^1+\hat{\Delta}_2, y^2+\Delta_2) - h(y^1+\hat{\Delta}_2+\Delta_1, y^2) \ ,
\end{aligned} \tag{D.5}$$

where $h := f \circ g$ and the change of variables $g : I\!R^2 \to I\!R^2$ is defined by:

$$g(x^1, x^2) := \left(x^1 - \frac{c^2}{c^1} \cdot (x^2 - y^2), x^2\right) \ .$$

Moreover,

$$\begin{aligned}
&h(y^1+\hat{\Delta}_2, y^2) + h(y^1+\hat{\Delta}_2+\Delta_1, y^2+\Delta_2) \\
&\quad - h(y^1+\hat{\Delta}_2, y^2+\Delta_2) - h(y^1+\hat{\Delta}_2+\Delta_1, y^2) \\
&= \int_{y^1+\hat{\Delta}_2}^{y^1+\hat{\Delta}_2+\Delta_1} \int_{y^2}^{y^2+\Delta_2} \frac{\partial^2 h}{\partial x^1 \partial x^2}(u,v)\ dv\ du\ .
\end{aligned} \tag{D.6}$$

Straightforward application of the chain rule yields:

$$\frac{\partial^2 h}{\partial x^1 \partial x^2}(u,v) = \frac{\partial^2 f}{\partial x^1 \partial x^2}\big(g(u,v)\big) - \frac{c^2}{c^1} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^1}\big(g(u,v)\big)\ . \tag{D.7}$$

Combining (D.5), (D.6), and (D.7), we have:

$$\begin{aligned}
&f(y^1+\hat{\Delta}_2, y^2) + f(y^1+\Delta_1, y^2+\Delta_2) - f(y^1, y^2+\Delta_2) - f(y^1+\hat{\Delta}_2+\Delta_1, y^2) \\
&= \int_{y^1+\hat{\Delta}_2}^{y^1+\hat{\Delta}_2+\Delta_1} \int_{y^2}^{y^2+\Delta_2} \frac{\partial^2 f}{\partial x^1 \partial x^2}\big(g(u,v)\big) - \frac{c^2}{c^1} \cdot \frac{\partial^2 f}{\partial x^1 \partial x^1}\big(g(u,v)\big)\ dv\ du \\
&\leq\ 0\ ,
\end{aligned} \tag{D.8}$$

where the inequality in (D.8) follows from (D.4). $\square$

# *Appendix E*

## Proofs for Problem (P7.1)

### E.1 Proof of Theorem 7.3

We prove the statements by joint induction on the time remaining, $n$. The proofs of statements (i), (ii), (iv), (v), and (vi) are essentially the same as the proofs of the corresponding statements in Theorem 5.8, except that the functions no longer depend on the ordering prices, which are time-invariant. So we prove statements (iii) and (vii) here.

Base Case: $n = 1$

$V_0(\mathbf{x}) = 0$ for all $\mathbf{x} \in I\!R^2$, so (iii) holds trivially. Let $\bar{\mathbf{y}}, \tilde{\mathbf{y}} \in I\!R^2$ be arbitrary. We show $G_1(\mathbf{y}_1)$ is $dv(\mathbf{c}, 1)$-submodular by showing:

$$G_1(\bar{\mathbf{y}}) + G_1(\tilde{\mathbf{y}}) \geq G_1(\bar{\mathbf{y}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{y}}) + G_1(\bar{\mathbf{y}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{y}}) \;, \tag{E.1}$$

and that $G_1(\mathbf{y}_1)$ is $dv(\mathbf{c}, 2)$-submodular follows by a symmetric argument. If $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ are $dv(\mathbf{c}, 1)$ comparable (i.e., $\bar{\mathbf{y}} \preceq_{dv(\mathbf{c},1)} \tilde{\mathbf{y}}$ or $\tilde{\mathbf{y}} \preceq_{dv(\mathbf{c},1)} \bar{\mathbf{y}}$), then (E.1) holds trivially. Assume $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ are not $dv(\mathbf{c}, 1)$ comparable, and assume without loss of generality that $\tilde{y}^1 > \bar{y}^1$. Thus, in order for $\bar{\mathbf{y}}$ and $\tilde{\mathbf{y}}$ to not be $dv(\mathbf{c}, 1)$ comparable, we also have

$\mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} > \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}}$, and therefore

$$\bar{\mathbf{y}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{y}} = \left(\bar{y}^1, \tfrac{\mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^1 \cdot \bar{y}^1}{c^2}\right) \text{ , and}$$

$$\bar{\mathbf{y}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{y}} = \left(\tilde{y}^1, \tfrac{\mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} - c^1 \cdot \tilde{y}^1}{c^2}\right) .$$

These locations of the meet and join imply:

$$\begin{aligned}
\mathbf{c}^{\mathrm{T}}(\bar{\mathbf{y}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{y}}) + \mathbf{c}^{\mathrm{T}}(\bar{\mathbf{y}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{y}}) &= \mathbf{c}^{\mathrm{T}}\left(\bar{y}^1, \frac{\mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^1 \cdot \bar{y}^1}{c^2}\right) + \mathbf{c}^{\mathrm{T}}\left(\tilde{y}^1, \frac{\mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} - c^1 \cdot \tilde{y}^1}{c^2}\right) \\
&= c^1 \cdot \bar{y}^1 + \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^1 \cdot \bar{y}^1 + c^1 \cdot \tilde{y}^1 + \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} - c^1 \cdot \tilde{y}^1 \\
&= \mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} + \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} .
\end{aligned}$$

Thus, $\mathbf{c}^{\mathrm{T}}\mathbf{y}$ is $dv(\mathbf{c},1)$-submodular (it is actually a $dv(\mathbf{c},1)$-valuation). Note that $\tilde{y}^1 > \bar{y}^1$ and $\mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} > \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}}$ imply:

$$\bar{y}^2 - \tilde{y}^2 \geq \frac{c^1}{c^2} \cdot \left(\tilde{y}^1 - \bar{y}^1\right) \geq 0 . \tag{E.2}$$

Next, let $\mathbf{d}$ be an arbitrary realization of the random vector $\mathbf{D}$. Then we have:

$$\begin{aligned}
& l\left(\bar{\mathbf{y}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{y}} - \mathbf{d}\right) + l\left(\bar{\mathbf{y}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{y}} - \mathbf{d}\right) \\
&= l^1\left(\bar{y}^1 - d^1\right) + l^2\left(\frac{\mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^1 \cdot \bar{y}^1}{c^2} - d^2\right) \\
&\quad + l^1\left(\tilde{y}^1 - d^1\right) + l^2\left(\frac{\mathbf{c}^{\mathrm{T}}\bar{\mathbf{y}} - c^1 \cdot \tilde{y}^1}{c^2} - d^2\right) \\
&= l^1\left(\bar{y}^1 - d^1\right) + l^1\left(\tilde{y}^1 - d^1\right) \\
&\quad + l^2\left(\tilde{y}^2 - d^2 + \frac{c^1}{c^2} \cdot \left(\tilde{y}^1 - \bar{y}^1\right)\right) \\
&\quad + l^2\left(\bar{y}^2 - d^2 - \frac{c^1}{c^2} \cdot \left(\tilde{y}^1 - \bar{y}^1\right)\right) \\
&\leq l^1\left(\bar{y}^1 - d^1\right) + l^1\left(\tilde{y}^1 - d^1\right) \\
&\quad + l^2\left(\bar{y}^2 - d^2\right) + l^2\left(\tilde{y}^2 - d^2\right) \\
&= l\left(\bar{\mathbf{y}} - \mathbf{d}\right) + l\left(\tilde{\mathbf{y}} - \mathbf{d}\right) \ ,
\end{aligned}$$

where the inequality follows from the convexity of $l(\cdot)$, and (E.2). Since $dv(\mathbf{c},1)$-submodularity is preserved under addition and positive scalar multiplication, the $dv(\mathbf{c},1)$-submodularity of the function $l\left(\mathbf{y}-\mathbf{d}\right)$ implies $I\!E\big[l\left(\mathbf{y}-\mathbf{D}\right)\big]$ is also $dv(\mathbf{c},1)$-submodular, and therefore, $G_1(\mathbf{y}_1)$, the sum of $dv(\mathbf{c},1)$-submodular functions, is also $dv(\mathbf{c},1)$-submodular. This completes the base case.

Induction Step

Assume statements (i) - (vii) are true for $n = 2, 3, \ldots, l-1$. We now show (iii) and (vii) are true for $n = l$. We show (iii) for $i = 1$, and it follows for $i = 2$ by a symmetric argument. As discussed in Section 7.3, it suffices to construct $\hat{\mathbf{y}}$ and $\check{\mathbf{y}}$, depending on the relative locations of $\bar{\mathbf{x}}$, $\tilde{\mathbf{x}}$, $\mathbf{y}^*(\bar{\mathbf{x}})$, and $\mathbf{y}^*(\tilde{\mathbf{x}})$, so as to ensure:

$$
\begin{aligned}
& \min_{\mathbf{y}\in\hat{\mathcal{A}}(\bar{\mathbf{x}})} \left\{\hat{G}_{l-1}(\mathbf{y})\right\} + \min_{\mathbf{y}\in\hat{\mathcal{A}}(\tilde{\mathbf{x}})} \left\{\hat{G}_{l-1}(\mathbf{y})\right\} \\
= \; & \hat{G}_{l-1}\Big(\mathbf{y}^*(\bar{\mathbf{x}})\Big) + \hat{G}_{l-1}\Big(\mathbf{y}^*(\tilde{\mathbf{x}})\Big) \\
\geq \; & \hat{G}_{l-1}(\hat{\mathbf{y}}) + \hat{G}_{l-1}(\check{\mathbf{y}}) \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{(E.3)} \\
\geq \; & \min_{\mathbf{y}\in\hat{\mathcal{A}}\left(\bar{\mathbf{x}}\wedge_{dv(\mathbf{c},1)}\tilde{\mathbf{x}}\right)} \left\{\hat{G}_{l-1}(\mathbf{y})\right\} + \min_{\mathbf{y}\in\hat{\mathcal{A}}\left(\bar{\mathbf{x}}\vee_{dv(\mathbf{c},1)}\tilde{\mathbf{x}}\right)} \left\{\hat{G}_{l-1}(\mathbf{y})\right\}.
\end{aligned}
$$

If $\bar{\mathbf{x}}$ and $\tilde{\mathbf{x}}$ are $dv(\mathbf{c},1)$-comparable, the desired conclusion is trivial, so we assume without loss of generality that $\bar{x}^1 \leq \tilde{x}^1$ and $\mathbf{c}^{\mathrm{T}}\bar{\mathbf{x}} \geq \mathbf{c}^{\mathrm{T}}\tilde{\mathbf{x}}$. For $\mathbf{x} = \bar{\mathbf{x}}, \tilde{\mathbf{x}}$, let

$$
\mathbf{y}^*(\mathbf{x}) = \Big(y^{1^*}(\mathbf{x}), y^{2^*}(\mathbf{x})\Big) = \min_{\mathbf{y}\in\hat{\mathcal{A}}(\mathbf{x})} \left\{\hat{G}_{l-1}(\mathbf{y})\right\} \; .
$$

Before proceeding, we note that for two vectors $\bar{\mathbf{z}}, \tilde{\mathbf{z}} \in I\!R^2$,

$$
\bar{\mathbf{z}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{z}} = \begin{pmatrix} \max\left\{\bar{z}^1, \tilde{z}^1\right\}, \\ \max\left\{\bar{z}^2 - \frac{c^1}{c^2}\cdot\max\{0, \tilde{z}^1 - \bar{z}^1\}, \tilde{z}^2 - \frac{c^1}{c^2}\cdot\max\{0, \bar{z}^1 - \tilde{z}^1\}\right\} \end{pmatrix},
$$

$$
\text{and } \bar{\mathbf{z}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{z}} = \begin{pmatrix} \min\left\{\bar{z}^1, \tilde{z}^1\right\}, \\ \min\left\{\bar{z}^2 + \frac{c^1}{c^2}\cdot\max\{0, \bar{z}^1 - \tilde{z}^1\}, \tilde{z}^2 + \frac{c^1}{c^2}\cdot\max\{0, \tilde{z}^1 - \bar{z}^1\}\right\} \end{pmatrix}.
$$

We now consider four exhaustive cases for the relative locations of $\bar{\mathbf{x}}, \tilde{\mathbf{x}}, \mathbf{y}^*(\bar{\mathbf{x}})$, and $\mathbf{y}^*(\tilde{\mathbf{x}})$.

Case 1: $y^{1^*}(\bar{\mathbf{x}}) \leq y^{1^*}(\tilde{\mathbf{x}}) - (\tilde{x}^1 - \bar{x}^1)$ and $\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\bar{\mathbf{x}}) \geq \mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}})$

Define:

$$\hat{\mathbf{y}} := \bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}} + \mathbf{y}^*(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}, \text{ and}$$

$$\check{\mathbf{y}} := \bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}} + \mathbf{y}^*(\bar{\mathbf{x}}) - \bar{\mathbf{x}} \ .$$

Clearly $\hat{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$ and $\check{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$, so we just need to show:

$$\hat{G}_{l-1}\big(\mathbf{y}^*(\bar{\mathbf{x}})\big) + \hat{G}_{l-1}\big(\mathbf{y}^*(\tilde{\mathbf{x}})\big) \geq \hat{G}_{l-1}(\hat{\mathbf{y}}) + \hat{G}_{l-1}(\check{\mathbf{y}}) \ . \tag{E.4}$$

For that matter, define:

$$\mathbf{z}_1 := \left(y^{1^*}(\bar{\mathbf{x}}), \frac{\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}}) - c^1 \cdot y^{1^*}(\bar{\mathbf{x}})}{c^2}\right), \text{ and}$$
$$\mathbf{z}_2 := \left(y^{1^*}(\bar{\mathbf{x}}) + \tilde{x}^1 - \bar{x}^1, \frac{\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}}) - c^1 \cdot \left(y^{1^*}(\bar{\mathbf{x}}) + \tilde{x}^1 - \bar{x}^1\right)}{c^2}\right).$$

See Figure E.1 for a diagram of these points. Note that $\mathbf{z}_1, \mathbf{z}_2, \hat{\mathbf{y}}$, and $\mathbf{y}^*(\tilde{\mathbf{x}})$ all lie on

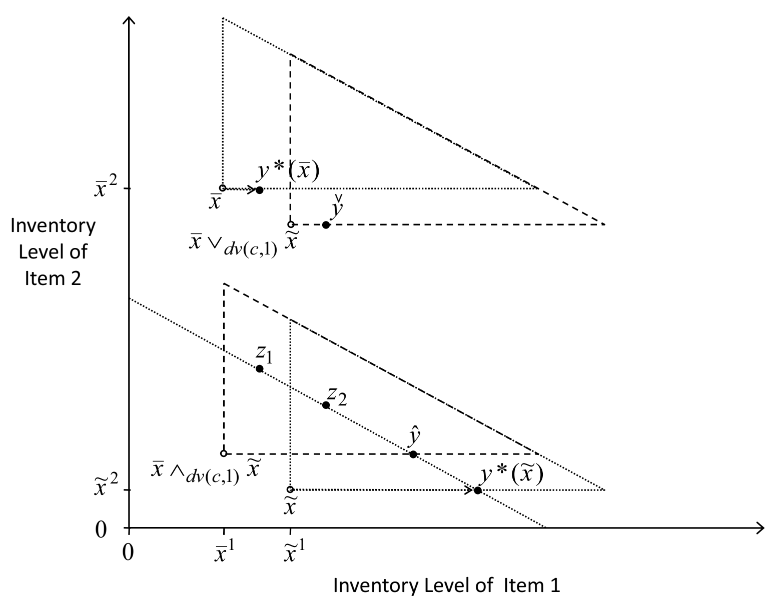


Figure E.1. Diagram of the points referred to in Case 1 of the proof of $dv(\mathbf{c},1)$-submodularity.

the line:

$$\left\{\mathbf{y} \in I\!R^2 : \mathbf{c}^{\mathrm{T}}\mathbf{y} = \mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}})\right\}. \tag{E.5}$$

Furthermore, we have:

$$z_2^1 - z_1^1 = \tilde{x}^1 - \bar{x}^1 = y^{1^*}(\tilde{\mathbf{x}}) - \hat{y}^1, \tag{E.6}$$

$$z_1^1 = y^{1^*}(\bar{\mathbf{x}}) \le y^{1^*}(\bar{\mathbf{x}}) + \tilde{x}^1 - \bar{x}^1 = z_2^1 \le y^{1^*}(\tilde{\mathbf{x}}), \text{ and} \tag{E.7}$$

$$z_1^1 = y^{1^*}(\bar{\mathbf{x}}) \le y^{1^*}(\tilde{\mathbf{x}}) + \bar{x}^1 - \tilde{x}^1 = \hat{y}^1 \le y^{1^*}(\tilde{\mathbf{x}}). \tag{E.8}$$

Equations (E.6)-(E.8) and the convexity of $\hat{G}_{l-1}(\cdot)$ along the line defined in (E.5) imply:

$$\hat{G}_{l-1}\big(\mathbf{y}^*(\tilde{\mathbf{x}})\big) \ge \hat{G}_{l-1}(\mathbf{z}_2) - \hat{G}_{l-1}(\mathbf{z}_1) + \hat{G}_{l-1}(\hat{\mathbf{y}}). \tag{E.9}$$

Additionally, by the $dv(\mathbf{c},1)$-submodularity of $\hat{G}_{l-1}(\cdot)$, we have:

$$\begin{aligned}\hat{G}_{l-1}\big(\mathbf{y}^*(\bar{\mathbf{x}})\big) &\ge \hat{G}_{l-1}\left(\mathbf{y}^*(\bar{\mathbf{x}}) \wedge_{dv(\mathbf{c},1)} \mathbf{z}_2\right) + \hat{G}_{l-1}\left(\mathbf{y}^*(\bar{\mathbf{x}}) \vee_{dv(\mathbf{c},1)} \mathbf{z}_2\right) - \hat{G}_{l-1}(\mathbf{z}_2)\\ &= \hat{G}_{l-1}(\mathbf{z}_1) + \hat{G}_{l-1}(\check{\mathbf{y}}) - \hat{G}_{l-1}(\mathbf{z}_2).\end{aligned} \tag{E.10}$$

Summing (E.9) and (E.10) yields the desired result, (E.4).

<u>Case 2</u>: $y^{1^*}(\bar{\mathbf{x}}) \le y^{1^*}(\tilde{\mathbf{x}}) - (\tilde{x}^1 - \bar{x}^1)$ and $\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\bar{\mathbf{x}}) \le \mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}})$

If $y^{2^*}(\bar{\mathbf{x}}) < y^{2^*}(\tilde{\mathbf{x}}) + \frac{c^1}{c^2} \cdot (\tilde{x}^1 - \bar{x}^1)$, then $\mathbf{y}^*(\bar{\mathbf{x}}) \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$ and $\mathbf{y}^*(\tilde{\mathbf{x}}) \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$. So in this case, we just let $\hat{\mathbf{y}} = \mathbf{y}^*(\bar{\mathbf{x}})$ and $\check{\mathbf{y}} = \mathbf{y}^*(\tilde{\mathbf{x}})$.

Otherwise, we have:

$$y^{2^*}(\bar{\mathbf{x}}) \ge y^{2^*}(\tilde{\mathbf{x}}) + \frac{c^1}{c^2} \cdot (\tilde{x}^1 - \bar{x}^1). \tag{E.11}$$

Define $\hat{\mathbf{y}}$ and $\check{\mathbf{y}}$ the same way as in Case 1, and define:

$$\mathbf{z}_1 := \mathbf{y}^*(\bar{\mathbf{x}}) \vee_{dv(\mathbf{c},2)} \hat{\mathbf{y}} = \left(\frac{\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}}) - c^2 \cdot y^{2^*}(\bar{\mathbf{x}})}{c^1}, y^{2^*}(\bar{\mathbf{x}})\right), \text{ and}$$

$$\mathbf{z}_2 := \mathbf{y}^*(\tilde{\mathbf{x}}) \vee_{dv(\mathbf{c},2)} \check{\mathbf{y}}$$
$$= \left( \frac{\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\tilde{\mathbf{x}}) - c^2 \cdot y^{2^*}(\bar{\mathbf{x}}) + c^1 \cdot (\tilde{x}^1 - \bar{x}^1)}{c^1}, y^{2^*}(\bar{\mathbf{x}}) - \frac{c^1}{c^2} \cdot \left(\tilde{x}^1 - \bar{x}^1\right) \right).$$

See Figure E.2 for a diagram of these points. The rest of the argument to show (E.4)

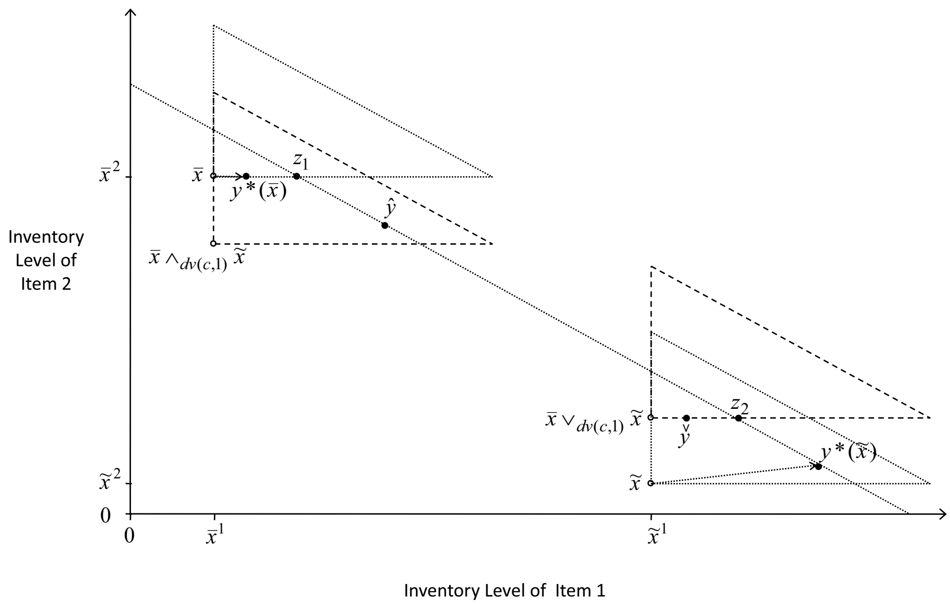


Figure E.2. Diagram of the points referred to in Case 2 of the proof of $dv(\mathbf{c}, 1)$-submodularity.

is the same as Case 1, except that (E.10) follows from the $dv(\mathbf{c}, 2)$-submodularity of $\hat{G}_{l-1}(\cdot)$ rather than the $dv(\mathbf{c}, 1)$-submodularity of $\hat{G}_{l-1}(\cdot)$.

<u>Case 3</u>: $y^{1^*}(\tilde{\mathbf{x}}) - (\tilde{x}^1 - \bar{x}^1) \leq y^{1^*}(\bar{\mathbf{x}}) \leq y^{1^*}(\tilde{\mathbf{x}})$

If $\mathbf{c}^{\mathrm{T}} [\mathbf{y}^*(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}] \geq c^1 \cdot \left(y^{1^*}(\bar{\mathbf{x}}) - \bar{x}^1\right)$, let

$$\hat{\mathbf{y}} := \mathbf{y}^*(\bar{\mathbf{x}}) \wedge_{dv(\mathbf{c},1)} \mathbf{y}^*(\tilde{\mathbf{x}}), \text{ and}$$
$$\check{\mathbf{y}} := \mathbf{y}^*(\bar{\mathbf{x}}) \vee_{dv(\mathbf{c},1)} \mathbf{y}^*(\tilde{\mathbf{x}}) \ .$$

Then (E.4) follows directly from the $dv(\mathbf{c}, 1)$-submodularity of $\hat{G}_{l-1}(\cdot)$. A fair bit of algebra shows that $\hat{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$ and $\check{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$, as desired.

Otherwise, we have:

$$\mathbf{c}^{\mathrm{T}} [\mathbf{y}^*(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}] \leq c^1 \cdot \left(y^{1^*}(\bar{\mathbf{x}}) - \bar{x}^1\right) \ .$$

Define:

$$\hat{\mathbf{y}} := \bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}} + \mathbf{y}^*(\bar{\mathbf{x}}) - \bar{\mathbf{x}},$$

$$\check{\mathbf{y}} := \bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}} + \mathbf{y}^*(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}},$$

$$\mathbf{z}_1 := \mathbf{y}^*(\bar{\mathbf{x}}) \wedge_{dv(\mathbf{c},1)} \check{\mathbf{y}} = \left( y^{1^*}(\tilde{\mathbf{x}}), \frac{\mathbf{c}^{\mathrm{T}}\mathbf{y}^*(\bar{\mathbf{x}}) - c^1 \cdot y^{1^*}(\tilde{\mathbf{x}})}{c^2} \right), \text{ and}$$

$$\mathbf{z}_2 := \mathbf{y}^*(\tilde{\mathbf{x}}) \vee_{dv(\mathbf{c},1)} \hat{\mathbf{y}} = \left( y^{1^*}(\tilde{\mathbf{x}}), \frac{\mathbf{c}^{\mathrm{T}}\hat{\mathbf{y}} - c^1 \cdot y^{1^*}(\tilde{\mathbf{x}})}{c^2} \right) .$$

See Figure E.3 for a diagram of these points. Then, by the convexity of $\hat{G}_{l-1}(\cdot)$ along

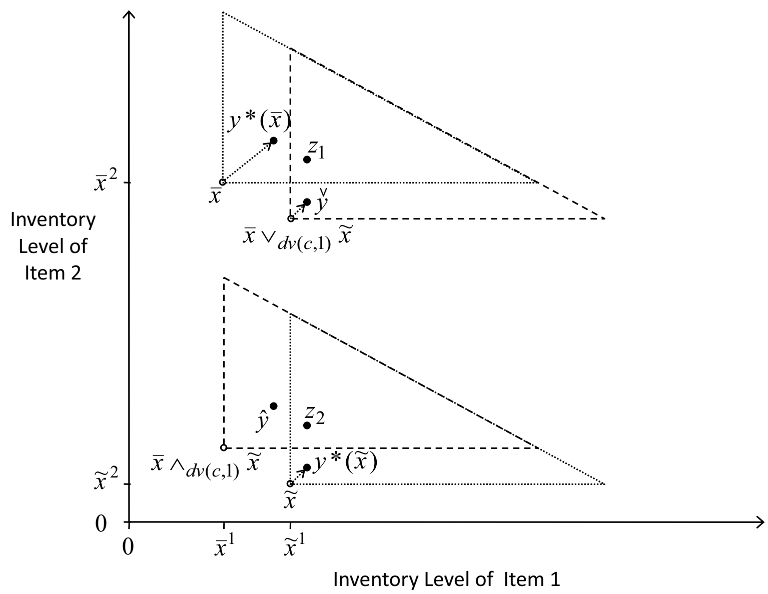


Figure E.3. Diagram of the points referred to in Case 3 of the proof of $dv(\mathbf{c}, 1)$-submodularity.

the line $y^1 = y^{1^*}(\tilde{\mathbf{x}})$, we have:

$$\hat{G}_{l-1}\left(\mathbf{y}^*(\tilde{\mathbf{x}})\right) \geq \hat{G}_{l-1}\left(\check{\mathbf{y}}\right) + \hat{G}_{l-1}\left(\mathbf{z}_2\right) - \hat{G}_{l-1}\left(\mathbf{z}_1\right) \ , \tag{E.12}$$

and by the $dv(\mathbf{c}, 1)$-submodularity of $\hat{G}_{l-1}(\cdot)$, we have:

$$\hat{G}_{l-1}\left(\mathbf{y}^*(\bar{\mathbf{x}})\right) \geq \hat{G}_{l-1}\left(\hat{\mathbf{y}}\right) + \hat{G}_{l-1}\left(\mathbf{z}_1\right) - \hat{G}_{l-1}\left(\mathbf{z}_2\right) \ . \tag{E.13}$$

Summing (E.12) and (E.13) yields the desired result, (E.4).

Case 4: $y^{1^*}(\tilde{\mathbf{x}}) \leq y^{1^*}(\bar{\mathbf{x}})$

If $y^{2^*}(\tilde{\mathbf{x}}) \geq (\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}})^2 = \tilde{x}^2 + \frac{c^1}{c^2} \cdot (\tilde{x}^1 - \bar{x}^1)$, then $\mathbf{y}^*(\tilde{\mathbf{x}}) \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$ and $\mathbf{y}^*(\bar{\mathbf{x}}) \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$. So in this case, we just let $\hat{\mathbf{y}} = \mathbf{y}^*(\tilde{\mathbf{x}})$ and $\check{\mathbf{y}} = \mathbf{y}^*(\bar{\mathbf{x}})$.

Otherwise, we have:

$$y^{2^*}(\bar{\mathbf{x}}) \geq \bar{x}^2 \geq (\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}})^2 = \tilde{x}^2 + \frac{c^1}{c^2} \cdot (\tilde{x}^1 - \bar{x}^1) > y^{2^*}(\tilde{\mathbf{x}}) \ .$$

Let $\beta \in (0,1)$ be such that:

$$\beta \cdot y^{2^*}(\bar{\mathbf{x}}) + (1-\beta) \cdot y^{2^*}(\tilde{\mathbf{x}}) = (\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}})^2 \ .$$

Then define:

$$\hat{\mathbf{y}} := \beta \mathbf{y}^*(\tilde{\mathbf{x}}) + (1-\beta)\mathbf{y}^*(\bar{\mathbf{x}}), \text{ and}$$

$$\check{\mathbf{y}} := (1-\beta)\mathbf{y}^*(\tilde{\mathbf{x}}) + \beta \mathbf{y}^*(\bar{\mathbf{x}}) \ .$$

See Figure E.4 for a diagram of these points. It is straightforward to show $\hat{\mathbf{y}} \in$

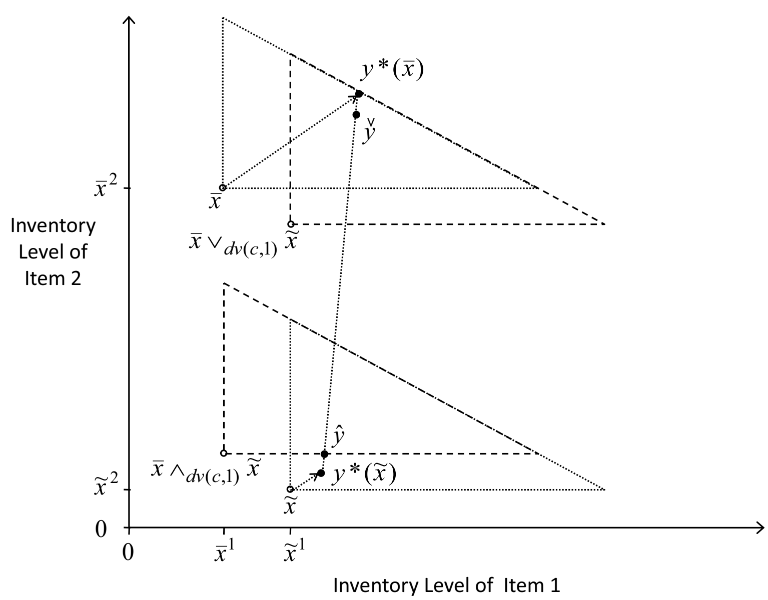


Figure E.4. Diagram of the points referred to in Case 4 of the proof of $dv(\mathbf{c},1)$-submodularity.

$\hat{\mathcal{A}}\left(\bar{\mathbf{x}} \wedge_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$ and $\check{\mathbf{y}} \in \hat{\mathcal{A}}\left(\bar{\mathbf{x}} \vee_{dv(\mathbf{c},1)} \tilde{\mathbf{x}}\right)$, and (E.4) follows directly from the convexity of $\hat{G}_{l-1}(\cdot)$. This completes the induction step for (iii).

Statement (vii), the $dv(\mathbf{c}, i)$-submodularity of $G_l(\mathbf{y})$, follows from the same series of arguments as (iii) and (iv) in Theorem 5.8, because, like convexity and supermodularity, $dv(\mathbf{c}, i)$-submodularity is preserved under addition and positive scalar multiplication. $\square$

## E.2 Proof of Theorem 7.4

The structure of the optimal policy when the vector of starting inventories is in regions $\hat{\mathcal{R}}_I(n), \hat{\mathcal{R}}_{II}(n), \hat{\mathcal{R}}_{III-A}(n)$, and $\hat{\mathcal{R}}_{III-B}(n)$ follows from essentially the same arguments as Theorem 5.9. Let $\mathbf{x} \in \hat{\mathcal{R}}_{IV-A}(n)$ be arbitrary. Also let $\tilde{\mathbf{y}} \in \hat{\mathcal{A}}(\mathbf{x})$ be arbitrary, and define:

$$\mathbf{z}_1 := \tilde{\mathbf{y}} \wedge_{dv(\mathbf{c},2)} \left(\hat{f}_n^1(x^2), x^2\right) = \left(\frac{\mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^2 \cdot x^2}{c^1}, x^2\right), \text{ and}$$
$$\mathbf{z}_2 := \tilde{\mathbf{y}} \vee_{dv(\mathbf{c},2)} \left(\hat{f}_n^1(x^2), x^2\right) = \left(\frac{c^1 \cdot \hat{f}_n^1(x^2) + c^2 \cdot x^2 - c^2 \cdot \tilde{y}^2}{c^1}, \tilde{y}^2\right) .$$

See Figure E.5 for a diagram of these points. By the $dv(\mathbf{c}, 2)$-submodularity of $\hat{G}_n(\cdot)$,

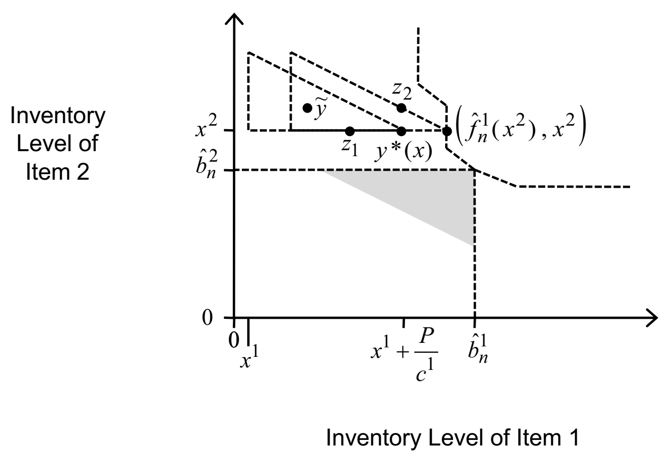


Figure E.5. Diagram of the points referred to in the proof of the optimal control action in region $\hat{\mathcal{R}}_{IV-A}(n)$.

we have:

$$\hat{G}_n(\mathbf{z}_1) + \hat{G}_n(\mathbf{z}_2) \leq \hat{G}_n(\tilde{\mathbf{y}}) + \hat{G}_n\Big(\big(\hat{f}_n^1(x^2), x^2\big)\Big) . \tag{E.14}$$

Furthermore, by the definition of $\hat{f}_n^1(\cdot)$ and the convexity of $\hat{G}_n(\cdot)$ along the curve $\Big(\hat{f}_n^1(x^2), x^2\Big)$, we have:

$$\hat{G}_n(\mathbf{z}_2) \geq \hat{G}_n\Big(\big(\hat{f}_n^1(\tilde{y}^2), \tilde{y}^2\big)\Big) \geq \hat{G}_n\Big(\big(\hat{f}_n^1(x^2), x^2\big)\Big) \ . \tag{E.15}$$

Equations (E.14) and (E.15) imply:

$$\hat{G}_n(\mathbf{z}_1) \leq \hat{G}_n(\tilde{\mathbf{y}}) \ , \tag{E.16}$$

and the convexity of $\hat{G}_n(\cdot)$ along the line $y^2 = x^2$ implies:

$$\hat{G}_n\Big(\big(\hat{f}_n^1(x^2), x^2\big)\Big) \leq \hat{G}_n\big(x^1 + \frac{P}{c^1}, x^2\big) \leq \hat{G}_n(\mathbf{z}_1) \ . \tag{E.17}$$

Equations (E.16) and (E.17) imply:

$$\hat{G}_n\big(x^1 + \frac{P}{c^1}, x^2\big) \leq \hat{G}_n(\tilde{\mathbf{y}}) \ ,$$

and since $\tilde{\mathbf{y}} \in \hat{\mathcal{A}}(\mathbf{x})$ was arbitrary, we have:

$$\hat{G}_n\big(x^1 + \frac{P}{c^1}, x^2\big) = \min_{\mathbf{y} \in \hat{\mathcal{A}}(\mathbf{x})} \Big\{\hat{G}_n(\mathbf{y})\Big\} \ .$$

The optimality of $\mathbf{y}_n^*(\mathbf{x}) = \big(x^1, x^2 + \frac{P}{c^2}\big)$ for $\mathbf{x} \in \hat{\mathcal{R}}_{IV-C}(n)$ follows from a symmetric argument.

Next, let $\mathbf{x} \in \hat{\mathcal{R}}_{IV-B}(n)$ be arbitrary. The fact that $\mathbf{c}^{\mathrm{T}}\left[\mathbf{y}_n^*(\mathbf{x}) - \mathbf{x}\right] = P$ follows from essentially the same argument as Theorem 5.9, so we want to show here that there exists a $\mathbf{y}_n^*(\mathbf{x}) \preceq \hat{\mathbf{b}}_n$. First, let $\tilde{\mathbf{y}} \in \hat{\mathcal{A}}(\mathbf{x})$ be arbitrary with $\tilde{y}^2 > \hat{b}_n^2$. Define the points:

$$\mathbf{z}_1 := \tilde{\mathbf{y}} \wedge_{dv(\mathbf{c},2)} \hat{\mathbf{b}}_n = \left(\frac{\mathbf{c}^{\mathrm{T}}\tilde{\mathbf{y}} - c^2 \cdot b^2}{c^1}, b^2\right), \text{ and}$$
$$\mathbf{z}_2 := \tilde{\mathbf{y}} \vee_{dv(\mathbf{c},2)} \hat{\mathbf{b}}_n = \left(\frac{\mathbf{c}^{\mathrm{T}}\hat{\mathbf{b}}_n - c^2 \cdot \tilde{y}^2}{c^1}, \tilde{y}^2\right) \ .$$

See Figure E.6 for a diagram of these points. By the $dv(\mathbf{c}, 2)$-submodularity of $\hat{G}_n(\cdot)$,

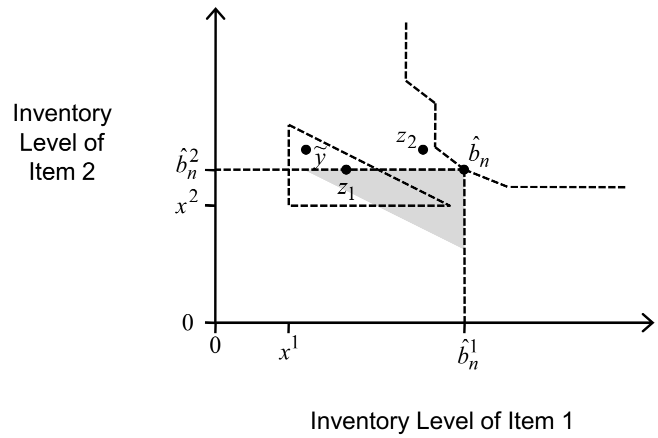


Figure E.6. Diagram of the points referred to in the proof of the optimal control action in region $\hat{\mathcal{R}}_{IV-B}(n)$.

we have:

$$\hat{G}_n(\mathbf{z}_1) + \hat{G}_n(\mathbf{z}_2) \leq \hat{G}_n(\tilde{\mathbf{y}}) + \hat{G}_n\Big(\big(\hat{f}_n^1(x^2), x^2\big)\Big) \ . \tag{E.18}$$

Furthermore, by the global optimality of $\hat{\mathbf{b}}_n$, we have:

$$\hat{G}_n(\mathbf{z}_2) \geq \hat{G}_n(\hat{\mathbf{b}}_n) \ . \tag{E.19}$$

Equations (E.18) and (E.19) imply:

$$\hat{G}_n(\mathbf{z}_1) \leq \hat{G}_n(\tilde{\mathbf{y}}) \ .$$

Note that, by construction, $\mathbf{z}_1 \in \hat{\mathcal{A}}(\mathbf{x})$ and $\mathbf{z}_1 \preceq \hat{\mathbf{b}}_n$. By a symmetric argument, for every $\tilde{\mathbf{y}} \in \hat{\mathcal{A}}(\mathbf{x})$ with $\tilde{y}^1 > \hat{b}_n^1$, there exists a $\bar{\mathbf{y}} \in \hat{\mathcal{A}}(\mathbf{x})$ with $\bar{\mathbf{y}} \preceq \hat{\mathbf{b}}_n$ such that $\hat{G}_n(\bar{\mathbf{y}}) \leq \hat{G}_n(\tilde{\mathbf{y}})$. Thus, there exists a choice of $\mathbf{y}_n^*(\mathbf{x})$ such that $\mathbf{y}_n^*(\mathbf{x}) \preceq \hat{\mathbf{b}}_n$.

$\square$

# References


[1] Adelman, D., and Mersereau, A. J. Relaxations of weakly coupled stochastic dynamic programs. *Operations Research 56*, 3 (May-June 2008), 712–727.

[2] Agarwal, M., Borkar, V. S., and Karandikar, A. Structural properties of optimal transmission policies over a randomly varying channel. *IEEE Transactions on Automatic Control 53*, 6 (July 2008), 1476–1491.

[3] Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. Wireless sensor networks: A survey. *Computer Networks 38*, 4 (March 2002), 393–422.

[4] Alfa, A. S. Vacation models in discrete time. *Queueing Systems 44* (2003), 5–30.

[5] Altman, E., and Jr., S. S. Optimality of monotonic policies for two-action Markovian decision processes with applications to control of queues with delayed information. *QUESTA 21* (1995), 267–291.

[6] Andrews, M., Kumaran, K., Ramanan, K., Stolyar, A., Vijayakumar, R., and Whiting, P. Scheduling in a queueing system with asynchronously varying service rates. *Probability in the Engineering and Informational Sciences 18* (2004), 191–217.

[7] Andrews, M., Kumaran, K., Ramanan, K., Stolyar, A., Whiting, P., and Vijayakumar, R. Providing quality of service over a shared wireless link. *IEEE Communications Magazine 39*, 2 (February 2001), 150–154.

[8] Antoniadou, E. *Lattice Programming and Economic Optimization.* PhD thesis, Stanford University, 1996.

[9] Antoniadou, E. Latting programming and consumer theory. Part I: Method and applications to two goods. *SSRN Working Paper* (2004).

[10] Antoniadou, E. Latting programming and consumer theory. Part II: Comparative statics with many goods. *SSRN Working Paper* (2004).

[11] Antoniadou, E. Comparative statics for the consumer problem. *Economic Theory 31*, 1 (2007), 189–203.

[12] Arapostathis, A., Borkar, V. S., Fernandez-Gaucherand, E., Ghosh, M. K., and Marcus, S. I. Discrete-time controlled Markov processes with average cost criterion: A survey. *SIAM J. Control and Optimization 31*, 2 (March 1993), 282–344.

[13] Ata, B. Dynamnic power control in a wireless static channel subject to a quality-of-service constraint. *Operations Research 53*, 5 (September-October 2005), 842–851.

[14] Bambos, N. Toward power-sensitive network architectures in wireless communications: Concepts, issues, and design aspects. *IEEE Personal Communications 5*, 3 (June 1998), 50–59.

[15] Bellman, R., Glicksberg, I., and Gross, O. On the optimal inventory equation. *Management Science 2*, 1 (October 1955), 83–104.

[16] Bensoussan, A., Crouhy, M., and Proth, J.-M. *Mathematical Theory of Production Planning*. Elsevier Science, 1983.

[17] Berggren, F., and Jäntti, R. Asymptotically fair transmission scheduling over fading channels. *IEEE Transactions on Wireless Communications 3*, 1 (January 2004), 326–336.

[18] Berry, R. A., and Gallager, R. G. Communication over fading channels with delay constraints. *IEEE Transactions on Information Theory 48*, 5 (May 2002), 1135–1149.

[19] Berry, R. A., and Yeh, E. M. Cross-layer wireless resource allocation. *IEEE Signal Processing Magazine 21*, 5 (September 2004), 59–69.

[20] Bertsekas, D., and Shreve, S. E. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific, 1996.

[21] Bhorkar, A., Karandikar, A., and Borkar, V. S. Power optimal opportunistic scheduling. In *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM)* (San Francisco, CA, November 2006).

[22] Borst, S., and Whiting, P. Dynamic channel-sensitive scheduling algorithms for wireless data throughput optimization. *IEEE Transactions on Vehicular Technology 52*, 3 (May 2003), 569–586.

[23] Boyd, S., and Vandenberghe, L. *Convex Optimization*. Cambridge University Press, 2004.

[24] Brémaud, P. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag, 1991.

[25] Brown, D. B., Smith, J. E., and Sun, P. Information relaxations and duality in stochastic dynamic programs. *Operations Research* (to appear).

[26] Cerpa, A., and Estrin, D. ASCENT: Adaptive Self-Configuring sEnsor Networks Topologies. *IEEE Transactions on Mobile Computing 3*, 3 (July 2004), 272–285.

[27] Chang, J., and Tassiulas, L. Energy conserving routing in wireless ad hoc networks. In *Proceedings of IEEE INFOCOM 2000* (Tel Aviv, Israel, March 2000).

[28] Chang, N., and Liu, M. Optimal channel probing and transmission scheduling for opportunistic spectrum access. In *Proceedings of the ACM International Conference on Mobile Computing and Networking (MobiCom'07)* (Montreal, Canada, September 2007), pp. 27–38.

[29] Chang, N. B., and Liu, M. Optimal channel probing and transmission scheduling for opportunistic spectrum access. *IEEE/ACM Transactions on Networking 17*, 6 (December 2009), 1805–1818.

[30] Chaporkar, P., and Proutiere, A. Optimal joint probing and transmission strategy for maximizing throughput in wireless systems. *IEEE Journal on Selected Areas in Communications 26*, 8 (October 2008), 1546–1555.

[31] Chen, B., Jamieson, K., Balakrishnan, H., and Morris, R. Span: An energy-efficient coordination algorithm for topology maintenance in ad hoc wireless networks. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking* (Rome, Italy, July 2001).

[32] Chen, S. X. The optimality of hedging point policies for stochastic two-product flexible manufacturing systems. *Operations Research 52*, 2 (March-April 2004), 312–322.

[33] Chen, W., Mitra, U., and Neely, M. J. Energy-efficient scheduling with individual delay constraints over a fading channel. In *Proceedings of the International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (Limassol, Cyprus, April 2007).

[34] Chen, W., Mitra, U., and Neely, M. J. Energy-efficient scheduling with individual delay constraints over a fading channel. *Wireless Networks 15*, 5 (July 2009), 601–618.

[35] Chow, C.-S., and Tsitsiklis, J. N. The complexity of dynamic programming. *Journal of Complexity 5*, 4 (December 1989), 466–488.

[36] Collins, B. E., and Cruz, R. L. Transmission policies for time varying channels with average delay constraints. In *Proceedings of the Allerton Conference on Communication, Control, and Computing* (Monticello, IL, 1999).

[37] Culler, D., Estrin, D., and Srivastava, M. Guest editors' introduction: Overview of sensor networks. *Computer 37*, 8 (August 2004), 41–49.

[38] de Farias, D. P., and Van Roy, B. The linear programming approach to approximate dynamic programming. *Operations Research 51*, 6 (November-December 2003), 850–865.

[39] de Vericourt, F., Karaesmen, F., and Dallery, Y. Dynamic scheduling in a make-to-stock system: A partial characterization of optimal policies. *Operations Research 48*, 5 (September-October 2000), 811–819.

[40] DeCroix, G. A., and Arreola-Risa, A. Optimal production and inventory policy for multiple products under resource constraints. *Management Science 44*, 7 (July 1998), 950–961.

[41] Djonin, D. V., and Krishnamurthy, V. Structural results on the optimal transmission scheduling policies and costs for correlated sources and channels. In *Proceedings of the IEEE Conference on Decision and Control* (Seville, Spain, December 2005), pp. 3231–3236.

[42] Dynkin, E. B., and Yushkevich, A. A. *Controlled Markov Processes*. Springer-Verlag, 1979.

[43] Elson, J., Girod, L., and Estrin, D. Fine-grained network time synchronization using reference broadcasts. In *Proceedings of the Fifth Symposium on Operating Systems Design and Implementation (OSDI 2002)* (Boston, MA, December 2002), vol. 36, pp. 147–163.

[44] Elson, J., and Römer, K. Wireless sensor networks: A new regime for time synchronization. *ACM SIGCOMM Computer Communications Review 33*, 1 (January 2003), 149–154.

[45] Evans, R. Inventory control of a multiproduct system with a limited production resource. *Naval Research Logistics Quarterly 14*, 2 (1967), 173–184.

[46] Fabian, T., Fisher, J. L., Sasieni, M. W., and Yardeni, A. Purchasing raw material on a fluctuating market. *Operations Research 7*, 1 (January-February 1959), 107–122.

[47] Federgruen, A., and So, K. C. Optimality of threshold policies in single-server queueing systems with server vacations. *Adv. Appl. Prob. 23*, 2 (June 1991), 388–405.

[48] Federgruen, A., and Zipkin, P. An inventory model with limited production capacity and uncertain demands II. The discounted-cost criterion. *Mathematics of Operations Research 11*, 2 (May 1986), 208–215.

[49] Feeney, L. M., and Nilsson, M. Investigating the energy consumption of a wireless network interface in an ad hoc networking environment. In *Proceedings of IEEE INFOCOM 2001* (Anchorage, Alaska, April 2001).

[50] Fernández-Gaucherand, E., Marcus, S. I., and Arapostathis, A. Structured solutions for stochastic control problems. In *Stochastic Theory and Adaptive Control* (1992), Springer, pp. 172–184.

[51] Fu, A., Modiano, E., and Tsitsiklis, J. Optimal energy allocation for delay-constrained data transmission over a time-varying channel. In *Proceedings of the IEEE Infocom* (San Francisco, CA, April 2003), vol. 2, pp. 1095–1105.

[52] Fu, A., Modiano, E., and Tsitsiklis, J. N. Optimal transmission scheduling over a fading channel with energy and deadline constraints. *IEEE Transactions on Wireless Communications 5*, 3 (March 2006), 630–641.

[53] Gesbert, D., and Alouini, M.-S. How much feedback is multi-user diversity really worth? In *Proceedings of the IEEE International Conference on Communications* (Paris, France, June 2004), vol. 1, pp. 234–238.

[54] Girod, B., Kalman, M., Liang, Y. J., and Zhang, R. Advances in channel-adaptive video streaming. In *Proceedings of the International Conference on Image Processing* (Rochester, NY, September 2002), vol. 1, pp. 9–12.

[55] Golabi, K. A single-item inventory model with stochastic prices. In *Proceedings of the Second International Symposium on Inventories* (Budapest, Hungary, 1982), pp. 687–697.

[56] Golabi, K. Optimal inventory policies when ordering prices are random. *Operations Research 33*, 3 (May-June 1985), 575–588.

[57] Gopalan, A., Caramanis, C., and Shakkottai, S. On wireless scheduling with partial channel-state information. In *Proceedings of the 45th Allerton Conference on Communication, Control, and Computing* (Urbana, IL, September 2007).

[58] Goyal, M., Kumar, A., and Sharma, V. Power constrained and delay optimal policies for scheduling transmission over a fading channel. In *Proceedings of the IEEE INFOCOM* (San Francisco, CA, March-April 2003), pp. 311–320.

[59] Goyal, M., Kumar, A., and Sharma, V. Optimal cross-layer scheduling of transmissions over a fading multiaccess channel. *IEEE Transactions on Information Theory 54*, 8 (August 2008), 3518–3537.

[60] Gravey, A., Louvion, J., and Boyer, P. On the Geo/D/1 and Geo/D/1/n queues. *Performance Evaluation 11* (1990), 117–125.

[61] Guha, S., Munagala, K., and Sarkar, S. Jointly optimal transmission and probing strategies for multichannel wireless systems. In *Proceedings of Conference on Information Sciences and Systems* (Princeton, NJ, March 2006).

[62] Guha, S., Munagala, K., and Sarkar, S. Optimizing transmission rate in wireless channels using adaptive probes. In *Proceedings of the ACM Sigmetrics/Performance Conference* (Saint-Malo, France, June 2006).

[63] Ha, A. Y. Optimal dynamic scheduling policy for a make-to-stock production system. *Operations Research 45*, 1 (January-February 1997), 42–53.

[64] Hanson, S., Seok, M., Lin, Y., Foo, Z., Kim, D., Lee, Y., Liu, N., Sylvester, D., and Blaauw, D. A low-voltage processor for sensing applications with picowatt standby mode. *IEEE Journal of Solid-State Circuits 44*, 4 (April 2009), 1145–1155.

[65] Hawkins, J. T. *A Lagrangian Decomposition Approach to Weakly Coupled Dynamic Optimization Problems and its Applications*. PhD thesis, Massachusetts Institute of Technology, 2003.

[66] Heinzelman, W. R., Chandrakasan, A., and Balakrishnan, H. Energy-efficient communication protocol for wireless microsensor networks. In *Proceedings of the Hawaii International Conference on Systems Sciences* (Maui, Hawaii, January 2000).

[67] Hernández-Lerma, O. *Adaptive Markov Control Processes*. Springer-Verlag, 1989.

[68] Hernández-Lerma, O., and Lasserre, J. B. *Discrete-Time Markov Control Processes*. Springer-Verlag, 1996.

[69] Heyman, D. P., and Sobel, M. J. *Stochastic Models in Operations Research*, vol. II. McGraw-Hil, 1984.

[70] Hinderer, K. Estimates for finite-stage dynamic programs. *Journal of Mathematical Analysis and Applications 55* (1976), 207–238.

[71] Hinderer, K., and Hübner, G. An improvement of J. F. Shapiro's turnpike theorem for the horizon of finite stage discrete dynamic programs. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes* (1974), pp. 245–255.

[72] Hinderer, K. F. On the structure of solutions of stochastic dynamic programs. In *Proceedings of the Seventh Conference on Probability Theory* (Bucharest, 1984), M. Iosifescu, Ed., Editura Academiei Republicii Socialiste România, pp. 173–182.

[73] Hoel, P. G., Port, S. C., and Stone, C. J. *Introduction to Stochastic Processes*. Waveland Press, 1987.

[74] Holtzman, J. M. CDMA forward link waterfilling power control. In *Proceedings of the IEEE Vehicular Technology Conference* (Tokyo, Japan, May 2000), vol. 3, pp. 1663–1667.

[75] Hübner, G. Improved procedures for eliminating suboptimal actions in Markov programming by the use of contraction properties. In *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, and Random Processes* (1974), pp. 257–263.

[76] Hübner, G. On the fixed points of the optimal reward operator in stochastic dynamic programming with discount factor greater than one. *ZAMM - Journal of Applied Mathematics and Mechanics 57*, 8 (1977), 477–480.

[77] Iglehart, D. L. Optimality of (s,S) policies in the infinite horizon dynamic inventory problem. *Management Science 9*, 2 (January 1963), 259–267.

[78] Intanagonwiwat, C., Govindan, R., Estrin, D., Heidemann, J., and Silva, F. Directed diffusion for wireless sensor networking. *IEEE/ACM Transactions on Networking 11*, 1 (February 2003), 2–16.

[79] Janakiraman, G., Nagarajan, M., and Veeraraghavan, S. Simple policies for managing flexible capacity. *Submitted to Management Science* (June 2009).

[80] Ji, Z., Yang, Y., Zhou, J., Takai, M., and Bagrodia, R. Exploiting medium access diversity in rate adaptive wireless lans. In *Proceedings of MOBICOM* (Philadelphia, PA, September 2004), pp. 345–359.

[81] Kalymon, B. Stochastic prices in a single-item inventory purchasing model. *Operations Research 19*, 6 (October 1971), 1434–1458.

[82] Karlin, S. Optimal inventory policy for the Arrow-Harris-Marschak dyanmic model. In *Studies in the Mathematical Theory of Inventory and Production* (1958), K. J. Arrow, S. Karlin, and H. Scarf, Eds., Stanford University Press, pp. 135–154.

[83] Karush, W. A theorem in convex programming. *Naval Research Logistics Quarterly 6*, 3 (September 1959), 245–260.

[84] Kingsman, B. G. Commodity purchasing. *Operational Research Quarterly 20* (1969), 59–80.

[85] Kingsman, B. G. *Commodity Purchasing in Uncertain Fluctuating Price Markets*. PhD thesis, University of Lancaster, 1969.

[86] Kittipiyakul, S., and Javidi, T. Resource allocation in OFDMA with time-varying channel and bursty arrivals. *IEEE Communications Letters 11*, 9 (September 2007), 708–710.

[87] Kittipiyakul, S., and Javidi, T. Delay-optimal server allocation in multi-queue multi-server systems with time-varying connectivities. *IEEE Transactions on Information Theory* (to appear).

[88] Knopp, R., and Humblet, P. A. Information capacity and power control in single-cell multiuser communications. In *Proceedings of the International Conference on Communications* (Seattle, WA, June 1995), vol. 1, pp. 331–335.

[89] Kumar, P. R., and Varaiya, P. *Stochastic Systems: Estimation, Identification, and Adaptive Control*. Prentice-Hall, 1986.

[90] Lee, J., and Jindal, N. Energy-efficient scheduling of delay constrained traffic over fading channels. In *Proceedings of the IEEE International International Symposium on Information Theory* (Toronto, Canada, July 2008).

[91] Lee, J., and Jindal, N. Asymptotically optimal policies for hard-deadline scheduling over fading channels. *Submitted to IEEE Transactions on Information Theory* (2009).

[92] Lee, J., and Jindal, N. Delay constrained scheduling over fading channels: Optimal policies for monomial energy-cost functions. In *Proceedings of the IEEE International Conference on Communications* (Dresden, Germany, June 2009).

[93] Lee, J., and Jindal, N. Energy-efficient scheduling of delay constrained traffic over fading channels. *IEEE Transactions on Wirless Communications 8*, 4 (April 2009), 1866–1875.

[94] Li, Y., and Bambos, N. Power-controlled wireless links for media streaming applications. In *Proceedings of the Wireless Telecommunications Symposium* (Pomona, CA, May 2004), pp. 102–111.

[95] Li, Y., Markopoulou, A., Bambos, N., and Apostolopoulos, J. Joint power-playout control for media streaming over wireless links. *IEEE Transactions on Multimedia 8*, 4 (August 2006), 830–843.

[96] Lin, Y., Sylvester, D., and Blaauw, D. A sub-pW timer using gate leakage for ultra low-power sub-Hz monitoring systems. In *Proceedings of the Custom Integrated Circuits Conference* (San Jose, CA, September 2007), pp. 397–400.

[97] Liu, X., Chong, E. K. P., and Shroff, N. B. A framework for opportunistic scheduling in wireless networks. *Computer Networks 41*, 4 (March 2003), 451–474.

[98] Liu, X., Chong, E. K. P., and Shroff, N. B. Optimal opportunistic scheduling in wireless networks. In *Proceedings of the Vehicular Technology Conference* (Orlando, FL, October 2003), vol. 3, pp. 1417–1421.

[99] Liu, X., Shroff, N. B., and Chong, E. K. P. Opportunistic scheduling: An illustration of cross-layer design. *Telecommunications Review 14*, 6 (December 2004), 947–959.

[100] Liu, Z., Nain, P., and Towsley, D. Sample path methods in the control of queues. *QUESTA 21* (1995), 293–335.

[101] Love, D. J., R. W. Heath, J., Lau, V. K. N., Gesbert, D., Rao, B. D., and Andrews, M. An overview of limited feedback in wireless communication systems. *IEEE Journal on Selected Areas in Communications 26*, 8 (October 2008), 1341–1365.

[102] Luna, C. E., Eisenberg, Y., Berry, R., Pappas, T. N., and Katsaggelos, A. K. Joint source coding and data rate adaptation for energy efficient wireless video streaming. *IEEE Journal on Selected Areas in Communications 21*, 10 (December 2003), 1710–1720.

[103] Magirou, V. F. Stockpiling under price uncertainty and storage capacity constraints. *European Journal of Operational Research 11* (1982), 233–246.

[104] Magirou, V. F. Comments on 'On Optimal Inventory Policies when Ordering Prices are Random' by Kamal Golabi. *Operations Research 35*, 6 (November-December 1987), 930–931.

[105] Milgrom, P., and Shannon, C. Monotone comparative statics. *Econometrica 62*, 1 (January 1994), 157–180.

[106] Mirman, L. J., Morand, O. F., and Reffett, K. A qualitative approach to monotone Markovian equilibrium in infinite horizon economies with capital. *SSRN Working Paper* (2004).

[107] Mirman, L. J., and Ruble, R. Lattice programming and the consumer's problem. *SSRN Working Paper* (2003).

[108] Moghaddam, M., Entekhabi, D., Farhadi, L., Goykhman, Y., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D. Initial analyses and demonstration of a soil moisture smart sensor web. In *Proceedings of the Earth Science Technology Conference* (East Adelphi, MD, June 2008).

[109] Moghaddam, M., Entekhabi, D., Goykhman, Y., Li, K., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D. A wireless soil moisture smart sensor web using physics-based optimal control: Concept and initial demonstrations. *Submitted to IEEE Journal of Special Topics in Applied Earth Observations and Remote Sensing*.

[110] Moghaddam, M., Entekhabi, D., Goykhman, Y., Liu, M., Mahajan, A., Nayyar, A., Shuman, D., and Teneketzis, D. A soil moisture smart sensor web using data assimilation and optimal control: Formulation and first laboratory demonstration. In *Proceedings of the International Geoscience and Remote Sensing Symposium* (Boston, MA, July 2008).

[111] Montes-de-Oca, R. The average cost optimality equation for Markov control processes on Borel spaces. *Systems and Control Letters 22* (1994), 351–357.

[112] Nain, P., Tsoucas, P., and Walrand, J. Interchange arguments in stochastic scheduling. *Journal of Applied Probability 27* (1989), 815–826.

[113] Neely, M. J. Max weight learning algorithms with application to scheduling in unknown environments. In *Proceedings of the Information Theory and Applications Workshop* (La Jolla, CA, February 2009).

[114] Neely, M. J., Modiano, E., and Rohrs, C. E. Dynamic power allocation and routing for time varying wireless networks. In *Proceedings of the IEEE INFOCOM* (San Francisco, CA, April 2003), vol. 1, pp. 745–755.

[115] Peña Perez, A., and Zipkin, P. Dynamic scheduling rules for a multiproduct make-to-stock queue. *Operations Research 45*, 6 (November-December 1997), 919–930.

[116] Porteus, E. L. Stochastic inventory theory. In *Stochastic Models* (1990), D. P. Heyman and M. J. Sobel, Eds., Elsevier Science, pp. 605–652.

[117] Porteus, E. L. *Foundations of Stochastic Inventory Theory*. Stanford University Press, 2002.

[118] Puterman, M. L. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley and Sons, 1994.

[119] Quah, J. K.-H. The comparative statics of constrained optimization problems. *Econometrica 75*, 2 (March 2007), 401–431.

[120] Rajan, D., Sabharwal, A., and Aazhang, B. Delay-bounded packet scheduling of bursty traffic over wireless channels. *IEEE Transactions on Information Theory 50*, 1 (January 2004), 125–144.

[121] Robinson, D. R. Markov decision chains with unbounded costs and applications to the control of queues. *Adv. Appl. Prob. 8* (1976), 159–176.

[122] Rockafellar, R. T. *Convex Analysis*. Princeton University Press, 1970.

[123] Römer, K. Time synchronization in ad hoc networks. In *Proceedings of the ACM Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc 01)* (Long Beach, CA, October 2001), pp. 173–182.

[124] Rust, J. Using randomization to break the curse of dimensionality. *Econometrica 65*, 3 (May 1997), 487–516.

[125] Sabharwal, A., Khoshnevis, A., and Knightly, E. Opportunistic spectral usage: Bounds and a multi-band CSMA/CA protocol. *IEEE/ACM Transactions on Networking 15*, 3 (June 2007), 533–545.

[126] Sadeghi, P., Kennedy, R. A., Rapajic, P. B., and Shams, R. Finite-state Markov modeling of fading channels - a survey of principles and applications. *IEEE Signal Processing Magazine 25*, 5 (September 2008), 57–80.

[127] Sanayei, S., and Nosratinia, A. Opportunistic beamforming with limited feedback. *IEEE Transactions on Wireless Communications 6*, 8 (August 2007), 2765–2771.

[128] Sarkar, M., and Cruz, R. L. Analysis of power management for energy and delay trade-off in a WLAN. In *Proceedings of the Conference on Information Sciences and Systems* (Princeton, New Jersey, March 2004).

[129] Sarkar, M., and Cruz, R. L. An adaptive sleep algorithm for efficient power management in WLANs. In *Proceedings of the Vehicular Technology Conference* (Stockholm, Sweden, May 2005).

[130] Schäl, M. Average optimality in dynamic programming with general state space. *Mathematics of Operations Research 18*, 1 (February 1993), 163–172.

[131] Schormans, J. A., Scharf, E. M., and Pitts, J. M. Prioritised Geo/D/1 telecommunications switch model. *Electronics Letters 28*, 6 (March 1992), 597–598.

[132] Schweitzer, P. J., and Seidmann, A. Generalized polynomial approximatins in Markovian decision processes. *Journal of Mathematical Analysis and Applications 110*, 2 (September 1985), 568–582.

[133] Sennott, L. I. *Stochastic Dynamic Programming and the Control of Queueing Systems*. John Wiley and Sons, 1999.

[134] Shakkottai, S., Srikant, R., and Stolyar, A. Pathwise optimality of the exponential scheduling rule for wireless channels. *Advances in Applied Probability 36*, 4 (2004), 1021–1045.

[135] Shapiro, J. F. Turnpike planning horizons for a Markovian decision model. *Management Science 14*, 5 (January 1968), 292–300.

[136] Sheu, J., Lai, C., and Chao, C. Power-aware routing for energy conserving and balance in ad hoc networks. In *Proceedings of the 2004 IEEE Conference on Networking, Sensing, and Control* (Taipei, Taiwan, March 2004), pp. 468–473.

[137] Shih, E., Cho, S., Ickes, N., Min, R., Sinha, A., Wang, A., and Chandrakasan, A. Physical layer driven protocol and algorithm design for energy-efficient wireless sensor networks. In *Proceedings of the International Conferece on Mobile Computing and Networking (MobiCom 2001)* (Rome, Italy, July 2001), pp. 272–286.

[138] Shuman, D., and Liu, M. Optimal sleep scheduling for a wireless sensor network node. In *Proceedings of the Fortieth Asilomar Conference on Signals, Systems, and Computers* (Pacific Grove, CA, October 2006), pp. 1337–1341.

[139] Shuman, D., and Liu, M. Optimal sleep scheduling for a wireless sensor network node. EECS Technical Report CSPL 377, University of Michigan, Ann Arbor, 2006.

[140] Shuman, D., and Liu, M. Energy-efficient transmission scheduling for wireless media streaming with strict underflow constraints. In *Proceedings of the International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (Berlin, Germany, April 2008), pp. 354–359.

[141] Shuman, D., Nayyar, A., Mahajan, A., Goykhman, Y., Li, K., Liu, M., Teneketzis, D., Moghaddam, M., and D.Entekhabi. Measurement scheduling for soil moisture sensing: From physical models to optimal control. *Submitted to Proceedings of the IEEE, Special Issue on Sensor Network Applications*.

[142] Shuman, D. I., and Liu, M. Opportunistic scheduling with deadline constraints in wireless networks. In *Performance Models and Risk Management in Communication Systems*, N. Gulpinar, P. Harrison, and B. Rustem, Eds., Springer, to appear.

[143] Shuman, D. I., and Liu, M. Dynamic clock calibration via temperature measurement. In *Proceedings of the Conference on Decision and Control* (Shanghai, China, December 2009), pp. 2082–2087.

[144] Shuman, D. I., Liu, M., and Wu, O. Energy-efficient transmission scheduling with strict underflow constraints. *Submitted to IEEE Transactions on Information Theory* (2010).

[145] Shuman, D. I., Liu, M., and Wu, O. Q. Energy-efficient transmission scheduling with strict underflow constraints. EECS Technical Report CGR 10-01, University of Michigan, Ann Arbor, 2010.

[146] Sichitiu, M. L., and Veerarittiphan, C. Simple, accurate time syncrhonization for wireless sensor networks. In *Proceedings of the IEEE Wireless Communications and Networking Conference (WCNC 2003)* (New Orleans, LA, March 2003), vol. 2, pp. 1266–1273.

[147] Singh, S., Woo, M., and Raghavendra, C. Power-aware routing in mobile ad hoc networks. In *Proceedings of the Fourth Annual ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom 1998)* (Dallas, TX, October 1998), pp. 181–190.

[148] Sivrikaya, F., and Yener, B. Time synchronization in sensor networks: A survey. *IEEE Network 18*, 4 (July/August 2004), 45–50.

[149] Smith, J. E., and McCardle, K. F. Structural properties of stochastic dynamic programs. *Operations Research 50*, 5 (2002), 796–809.

[150] Sobel, M. J. Making short-run changes in production when the employment level is fixed. *Operations Research 18*, 1 (January-February 1970), 35–51.

[151] Sobel, M. J. Production smoothing with stochastic demand II: Infinite horizon case. *Management Science 17*, 11 (July 1971), 724–735.

[152] Srivastava, R., and Koksal, C. E. Energy optimal transmission scheduling in wireless sensor networks. *IEEE Transactions on Wireless Communications* (to appear).

[153] Sundararaman, B., Buy, U., and Kshemkalyani, A. D. Clock synchronization for wireless sensor networks: A survey. *Ad Hoc Networks 3*, 3 (May 2005), 281–323.

[154] Takagi, H. *Queueing Analysis, Vol 3: Discrete-Time Systems*. North-Holland, 1993.

[155] Tarello, A., Sun, J., Zafer, M., and Modiano, E. Minimum energy transmission scheduling subject to deadline constraints. *ACM Wireless Networks 14*, 5 (October 2008), 633–645.

[156] Tassiulas, L., and Ephremides, A. Dynamic server allocation to parallel queues with randomly varying connectivity. *IEEE Transactions on Information Theory 39*, 2 (March 1993), 466–478.

[157] Tayur, S. Computing the Optimal Policy for Capacitated Inventory Models. *Communications in Statistics. Stochastic Models 9*, 4 (1993), 585–598.

[158] Topkis, D. M. Minimizing a submodular function on a lattice. *Operations Research 26*, 2 (March-April 1978), 305–321.

[159] Topkis, D. M. *Supermodularity and Complementarity*. Princeton University Press, 1998.

[160] Tse, D., and Viswanath, P. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[161] Uysal-Biyikoglu, E., and Gamal, A. E. On adaptive transmission for energy efficiency in wireless data networks. *IEEE Transactions on Information Theory 50*, 12 (December 2004), 3081–3094.

[162] Uysal-Biyikoglu, E., Prabhakar, B., and Gamal, A. E. Energy-efficient packet transmission over a wireless link. *IEEE/ACM Transactions on Networking 10*, 4 (August 2002), 487–499.

[163] Veatch, M. H., and Wein, L. M. Scheduling a make-to-stock queue: Index policies and hedging points. *Operations Research 44*, 4 (July-August 1996), 634–647.

[164] Viswanath, P., Tse, D. N. C., and Laroia, R. Opportunistic beamforming using dumb antennas. *IEEE Transactions on Information Theory 48*, 6 (June 2002), 1277–1294.

[165] Wang, H. *Opportunistic Transmission of Wireless Data Over Fading Channels Under Energy and Delay Constraints*. PhD thesis, Rutgers University, 2003.

[166] Wein, L. M. Dynamic scheduling of a multiclass make-to-stock queue. *Operations Research 40*, 4 (July-August 1992), 724–735.

[167] Xu, Y., Heidemann, J., and Estrin, D. Geography-informed energy conservation for ad hoc routing. In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Mobile Computing and Networking* (Rome, Italy, July 2001).

[168] Ye, F., Zhong, G., Lu, S., and Zhang, L. PEAS: A robust energy conserving protocol for long-lived sensor networks. In *Proceedings of the Tenth Annual IEEE International Conference on Network Protocols* (Paris, France, November 2002).

[169] Zafer, M., and Modiano, E. Continuous-time optimal rate control for delay constrained data transmission. In *Proceedings of the 43rd Allerton Conference on Communication, Control, and Computing* (Urbana, IL, September 2005).

[170] Zafer, M., and Modiano, E. Delay-constrained energy efficient data transmission over a wireless fading channel. In *Proceedings of the Information Theory and Applications Workshop* (La Jolla, CA, January 2007).

[171] Zahrn, F. C. *Studies of Inventory Control and Capacity Planning with Multiple Sources.* PhD thesis, Georgia Institute of Technology, August 2009.

[172] Zhang, D., and Wasserman, K. M. Transmission schemes for time-varying wireless channels with partial state observations. In *Proceedings of the IEEE Infocom* (New York, NY, June 2002), vol. 2, pp. 467–476.

[173] Zheng, Y.-S., and Zipkin, P. A queueing model to analyze the value of centralized inventory information. *Operations Research 38*, 2 (March-April 1990), 296–307.

[174] Zorzi, M., and Rao, R. R. Energy-constrained error control for wireless channels. *IEEE Personal Communications 4*, 6 (December 1997), 27–33.