STRUCTURAL RESULTS FOR CONSTRAINED MARKOV DECISION PROCESSES

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Cory Jay Girard

August 2018

STRUCTURAL RESULTS FOR CONSTRAINED MARKOV DECISION PROCESSES

Cory Jay Girard, Ph.D.

Cornell University 2018

In the existing literature on the dynamic control of service systems, a decision-maker seeks to optimize a single performance metric over a given time-horizon. However, in many settings, the decision-maker may be interested in multiple performance metrics. Take, for instance, the problem of assigning cross-trained hospital staff to two classes of patients: low-priority and high-priority. In the typical framework, this problem could be modelled as a Markov Decision Process (MDP), in which the performance metric to be minimized is a weighted combination of expected waiting times for each class. However, we argue that a more natural approach is to consider the constrained problem: minimizing the expected waiting time for lower priority patients, while keeping that of higher priority patients under a given target, V. In particular, we concern ourselves with uncovering structural properties of this problem. These properties imply the existence of simple optimal policies that are easy to implement in practice.

We formulate this problem (the parallel setting), as well as a related problem in which customers undergo two phases of service in series (the tandem setting), as Constrained Markov Decision Processes (CMDPs). We present a general framework for solving two-class CMDPs, showing that they can be solved by using the Lagrangian dual to specify a particular unconstrained problem. If an appropriate Lagrange multiplier can be discerned, structural results from the resulting La-

grangian relaxation can be used to exploit structure in the original CMDP. We show that for both the parallel and tandem settings, the framework leads to simple threshold-like optimal policies. The results in each case are used to develop heuristics for analogous problems with abandonments with applications to healthcare, call centers and manufacturing systems. The efficacy of the heuristics are verified in each case via a detailed numerical study. We then extend the results in the parallel case to handle multiple classes and constraints.

Lastly, we consider a controlled, truncated birth-death chain motivated by optimal treatment prescription in the context of personalized medicine. In this model, states represent the patient's state of health, and treatments can be prescribed to influence improvement and/or deterioration of health. The problem of dynamically prescribing treatments at minimal cost while maintaining a given level of health is modelled as a two-cost CMDP. Rather than employing the more general methods developed earlier, we decompose the state space and consider an alternative Lagrangian relaxation involving two simpler subproblems. We obtain structural results for this problem by showing that optimal solutions to these subproblems can be combined into a constrained-optimal policy. We then attempt to find conditions under which a monotone optimal policy exists under more general transition rates between states of health.

BIOGRAPHICAL SKETCH

Cory was born on May 7, 1991 in Atlanta, Georgia, and grew up in the suburbs of Doraville and Dunwoody. He stayed in his hometown for his undergraduate career at Georgia Tech, completing his B.S. in Industrial and Systems Engineering in 2013. After realizing that he was not ready to be thrown into the real world just yet, and that 22 years in Atlanta was far too much time to spend in one place, he began work on his Ph.D. at Cornell in August 2013. In his spare time, he enjoys dunking on people during pickup basketball games, playing ultimate frisbee, discovering new music, writing about himself in the third person, and hopelessly following Atlanta sports teams. He ranks dogs and turtles among his favorite animals. After completing his Ph.D. he will begin working at Wayfair in Boston, Massachusetts.

This thesis is dedicated to my parents, my friends, and my dog.

ACKNOWLEDGEMENTS

First, I would like to thank my advisor, Mark Lewis, for the advice and time he has given me over the course of my graduate career. He has introduced me to interesting problems, and his open-mindedness has fostered my growth as a researcher, allowing me the autonomy and freedom to approach these problems from different angles. I would also like to thank Shane Henderson and Adrian Lewis for serving on my committee and for providing useful feedback on my work. I must also express my gratitude to Alisha Waller, an ORIE Ph.D alumna who taught my undergraduate optimization course and who played an instrumental role in my decision to pursue a doctoral degree, and to Jim Dai, who I had the privilege of learning from in both my undergraduate and graduate careers. Additionally, I would like to thank Jamol Pender for his enthusiastic presence in the department and for being the second-best player on our intramural basketball team.

I also am extremely grateful for the other ORIE Ph.D students. In addition to providing thoughtful comments about research or other technical subject matter, they also have provided great company and support. In particular, I would like to acknowledge:

- James Dong, for being an amazing roommate, endless hours of basketball and cooking show videos, and, in the words of the legendary Kenneth C. Chong, "knowing literally everything"
- Chamsi Hssaine, for being a fellow basketball fan, turtle enthusiast, and for her "glowing personality" and "excellent texting etiquette"
- Julian Sun, for being Julian
- Kenneth "K-money" Chong, for never releasing his mixtape, his baking skills,

his openness to discuss research questions and ideas, and because his mannerisms serve as an endless source of entertainment

- Andrew Daw, for being open in discussing research, and for being my primary source for music discovery and discussion
- Steve Pallone, for being a linear algebra machine, and for his (amusing) Tommy Wiseau impersonation and (very poor, but also amusing) Desiigner one
- Pamela Badian-Pessot, for being good at trivia and giving me someone to talk about MDPs with
- Venus Lo, for her efforts to place Ken on a never-ending baking fellowship
- Weici Hu, for making me think my jokes are funnier than they actually are
- Ravi Kumar, for being someone I could bounce ideas off of during the early stages of my research career
- Chaoxu Tong, for his inspiring, head-scratching series of ORIE Ph.D talks
- Angela Zhou, for making me appreciate Flying Lotus and Nujabes
- Matthew Zalesak, for assuming I (and the entire department) know far more about Japanese culture than I actually do
- Dave Lingenbrink, for making "Peanut Office" somehow stick
- Sam Gutekunst, for realizing that sending page-long climbing emails every week was not sustainable
- David Eckman, for being the nicest person in the department
- Amy Zhang, whose kindness may lead her into becoming the "next David Eckman"

- Woo-Hyung Cho, for occasionally dropping advice on how to be an adult, which I usually (unwisely) ignore
- Emily Fischer, for not making me feel guilty when I distract her office
- Patrick Steele, for his awe-inspiring beard
- Ben Grimmer, for his never-ending stream of ideas for questionable businesses
- Sander Aarts, for his magnificent drawings of turtles
- Anyone that I accidentally left out, for not being upset with me.

Finally, I must thank my parents, John and Mei-Kuan, for their love and support every step of the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Markov Decision Processes (MDPs) have proven to be a useful tool in modelling the dynamic control of service systems. In particular, formulating control problems as MDPs can be beneficial in two primary ways. First, practitioners can turn to solution methods from the existing literature, such as value iteration, policy iteration and linear programming, to solve their decision problems. Second, for more complicated problems or settings which call for more robust solutions, the optimality equations of an MDP can be leveraged to uncover structural properties. Often, these structural properties imply the existence of simple optimal policies for the control problem (e.g. threshold policies). This is commonly utilized in one of two ways. First, the existence of such policies can lead to more efficient solution methods by reducing the size of the search space. Second, these policies can provide intuition and inform decision-makers of the class of policies they should consider. Since these types of polices are often easily parametrized (e.g. by the location of a threshold), the exact policy to use could then be determined by a practitioner with domain expertise. The search for structural results in dynamic control problems is a well-studied, active area of research, and serves as our primary focus.

In certain applications, however, formulating control problems as MDPs is inadequate. This arises in situations when the decision-maker is interested in multiple performance measures, but is unsure of how to weigh the importance of each one. As a motivating example, consider the problem of assigning cross-trained emergency department (ED) staff to two classes of patients: lower and higher priority. By viewing each class of customers as having its own queue, and by assigning each class a (per unit) holding cost, we can frame this as a classical queueing control

problem: assigning servers to queues in order to minimize the expected holding cost. When the interarrival and processing times are exponentially distributed, this can be formulated as an MDP. Furthermore, in the single-server case, a structural result known as the $c\mu$-rule holds: it is optimal to serve a customer with the highest $c\mu$ index, calculated as the product of holding cost and service rate [16]. In the ED setting, however, it is unclear how to weigh the importance (via holding costs) of higher priority patients v. lower priority patients. While one could justify a choice of holding costs via cost analysis (e.g. how much it costs a hospital to treat each class of patient), we argue that a constrained optimization formulation is more appropriate: minimizing the lower priority cost while keeping the higher priority cost under a target value, say V.

We consider control problems that can be formulated as Constrained Markov Decision Processes (CMDPs), and aim to prove structural results. In particular, we study how structure for the constrained problem can be extended from the unconstrained setting by considering the Lagrangian. In Chapter 2, we present a general framework for extracting structure from CMDPs in which there are two costs and a single constraint placed upon one of them. We consider the previously described problem motivated by the ED setting (the parallel case), and a similar problem in which customers undergo two phases of service (the tandem case). We formulate each as a CMDP and cast them inside this framework, utilizing the structural results of their unconstrained counterparts to prove the optimality of a broad class of randomized-threshold policies in each case. Motivated by these results, we select particularly simple randomized-threshold policies as heuristics when abandonments are introduced in each case. For each problem, the efficacy of these policies is verified via a detailed numerical study. It should be noted that while these problems are motivated by healthcare, the results can be applied to

various systems, such as call centers, manufacturing systems, and web chat support.

In Chapter 3, we extend the results in the parallel case to handle multiple classes and constraints. This offers decision-makers the ability to further stratify customers into more refined groups, and allows for more flexibility in how service level constraints are placed upon these groups. To illustrate this, reconsider the ED setting, originally modelled as having two classes of patients. Suppose that the higher priority class is actually composed of two extremely different types of patients: type 1 patients take a long time to treat, while type 2 patients are treated very quickly. When framed as a two-class CMDP, it is possible that an optimal policy may severely violate the higher priority service requirement among type 1 patients, while still satisfying the service requirement among all higher priority patients. However, if formulated as a three-class CMDP, we could specify target service levels for each subclass, ensuring that both types of higher priority patients receive adequate service. When there is a single constraint, we are able to fully extend the results from the two-class setting. However, in the case where multiple classes each have a service requirement constraint, the results can only be partially extended. In particular, a specific subclass of randomized-threshold policies is shown to be optimal.

In Chapter 4, we revisit the two-cost setting, this time considering a problem motivated by optimal treatment prescription in the context of personalized medicine. We first model the problem as a controlled, truncated birth-death chain. In this model, states represent the patient's state of health, and treatments can be prescribed to influence the improvement (or deterioration) of health. The problem

of dynamically prescribing treatments at minimal cost while maintaining a given level of health is modelled as a two-cost CMDP. However, in this setting, the Lagrangian method developed in Chapter 2 is of limited use: we do not know of any structural results for the Lagrangian relaxation, and finding them may be difficult. Rather than seeking to find such structure, we instead employ a decomposition-based approach, resulting in an alternative Lagrangian relaxation involving two simpler subproblems for which uncovering structural properties is easier. We then obtain structural results for the constrained case by showing that optimal solutions to these subproblems can be "stitched" together to form a constrained-optimal policy. We then attempt to find conditions under which a monotone optimal policy exists under more general transition rates between states of health.

CHAPTER 2

A FRAMEWORK FOR FINDING STRUCTURAL RESULTS FOR TWO COST CONSTRAINED MDPS

2.1 Introduction

An interesting dilemma in modeling resource allocation in service systems is how to model preferences of the decision-maker with regard to prioritization of service requests. Suppose there are multiple queues and flexible resources. When there is only one customer class, it is reasonable to seek the allocation that maximizes the departure rate (throughput). Alternatively, we might search for the allocation that minimizes the average waiting time per customer. In the case of multiple customer classes, the analogue to throughput is to attach a distinct reward for the completion of service of each class and to pursue the maximum average reward rate. Similarly, rather than the average waiting time in the single class case, the waiting time of customers in each class are weighted by their relative importance. That is to say, if the cost per unit time for a class 1 customer waiting is h_1, then one way to capture a customer twice as important as class 1 is to set a second class cost of waiting to be $h_2 = 2h_1$. In the parlance of scheduling literature h_1 and h_2 are called holding cost rates per customer per unit time. Here again, the policy that achieves the minimum long-run average holding cost rate is considered optimal. Once the model is developed, the appropriate criterion set, and the parameters estimated, the next task is to find an optimal control strategy. This is where the search for simple structure in the optimal control has lead to a vast literature quite often using Markov decision processes (MDPs).

In this chapter we seek an alternative to the methodology of modeling pref-

erences described above. Constrained Markov decision processes (CMDPs) is a methodology that has not seen wide applications in the literature, but is a more natural specification for modern service systems. The reasons for this we contend are twofold. On the one hand, MDPs can be written as linear programs (LPs) so for direct applications one could incorporate constraints into an LP and use the vast literature on linear programming to solve specific problem instances. This has the down side of not taking advantage of any efficiencies gained by dynamic programming, and also does not provide general guidelines for the problem space (like structural results provide). Secondly, it is quite difficult to use constrained MDPs to find structural results that can be applied to large classes of problems. It is on this second front that we make a contribution. We explain how we can use the structural results in related unconstrained MDPs to obtain optimal or near optimal controls in the constrained formulation.

The idea is actually simple. Suppose there is a single constraint. Consider the related Lagrangized (unconstrained) problem. A priori policies that optimize the unconstrained problem may not be optimal in the original objective, and may have costs that violate the constraints. It turns out that (under conditions we develop here) if the unconstrained problem can be solved for a particular Lagrange multiplier, and a policy that achieves the optimal value for this unconstrained problem also has cost that meets the constraint at equality, that policy is optimal for the constrained problem. Suppose then we can characterize a class of policies that are optimal in the unconstrained problem for a particular multiplier with enough structure so that they can be ordered **and** at the extremes straddle the constraint. We can obtain a policy (within this class of policies) that meets the constraint at equality by randomizing between policies closest to the constraint on either side. Of course, this is a very much simplified description since finding the "right"

Lagrange multiplier, the structure within a class of policies, and a reasonable sequence along which to search requires particular care. In addition there needs to include some monotonicity within the class and some form of continuity so that the randomization ensures the constraint is met at equality.

In terms of the applicability of our methodological contributions, one need not look far for detailed examples in the literature for applications of parallel [25, 13, 24, 2] and tandem queues (also called queues in series) [30, 21, 19, 51]. Here we focus on patient flow in health care as this is what motivated much of our recent work. Consider the fact that according to The National Ambulatory Medical Care Survey in 2012 there were approxiamtely 130 million emergency department (ED) visits, according to the CDC [17]. Aside from those that arrive via ambulance patients in an emergency department are unscheduled. As EDs become overcrowded, they seek methods by which lower acuity patients (in particular those without life-threatening injuries) can be cleared more quickly. Immediately upon arriving, patients are triaged to decide the level of the severity of their injuries using the Emergency Severity Index (ESI). The ESI has 5 levels of severity with levels 1 and 2 signifying patients requiring prompt attention and in need of hospitalization, while levels 3-5 have lower acuity injuries and may be treated and released. The issue is that if we performed service on each patient on a first come, first served basis it is quite possible that a lower acuity patient is caught behind a higher acuity patient thereby increasing their waiting time significantly. Instead some hospitals like the Lutheran Medical Center in Brooklyn, NY (a full service, 450+ bed hospital) have decided to provide separate treatment areas for lower acuity patients. This program is called a "Triage-Treatment-Release" (TTR) program and has been studied in detail by some of the authors in the context of maximizing long-run average rewards [53, 54]. The TTR program represents operations (triage and

treatment) that are performed in series. Similarly, consider those patients that are classified with indices 1 and 2. These patients are too severely injured to be seen in the TTR, but can also be stratified into two classes. Patients arriving with index 1 are of the highest priority, but we should still ensure that index 2 receive high quality care. This is a question of which of two parallel queues to prioritize. Both scenarios make the assumption that the medical service providers (doctors and/or physician assistants, etc.) can perform all of the tasks, but this is becoming more common in the EDs as a way to improve patient flow (cf. Soremekun et al. [45] or Subash et al. [46]). The plan then is to model both scenarios using constrained MDPs, obtain structural results in the Lagrangian of the constrained problem and to show in which cases optimality can be obtained. Finally, we use the structural results to develop heuristics that perform well in practice.

A classic result for CMDPs with one constraint states that there exists an optimal policy that randomizes between actions in at most one state (a 1-randomized policy) [3]. For our problem, this means that in at most one state, the policy will flip a (potentially) biased coin to determine which class of customers to serve. In the rest of the states, the policy chooses which class to serve deterministically. This result works well when the state space is single-dimensional (or multi-dimensional with a single infinite dimension), since the search in one dimension is quite often a search for a threshold value (at which point we randomize). Our formulation requires two dimensions, each of which is infinite, making the search for a 1-randomized policy more difficult. Second, a practical challenge. The 2-dimensional state space usually means the decision-maker (a medical service provider in the health care example) needs visibility into both dimensions. The heuristics we develop are more consistent with the single dimensional search and require that the decision-maker monitor only one dimension and then randomize at that threshold

(no matter the value of the other dimension). This simplifies both the search and implementation. In most of the cases we consider, the heuristics lead to policies that are within one percent of optimal. Moreover, it turns out that we are also able to show that our heuristic is optimal in the model without abandonments; which may be of independent interest.

The remainder of the chapter is arranged as follows. We further discuss related literature in Section 2.2. Preliminaries are covered in Section 2.3. Section 2.4 conceptualizes a general framework for considering two-cost CMDPs with a single constraint. The next two sections apply this framework to specific problems involving dynamic server allocation in two different queueing systems. In particular, Section 2.5 focuses on server allocation in a two-class, parallel queueing system, whereas Section 2.6 focuses on a similar problem for a two-class, tandem system.

2.2 Related Literature

Much of the framework we develop for CMDPs with two costs and a single constraint rely on results from Altman [3], which provides fundamental results in CMDPs spanning a broad class of problems and cost criteria.

We consider two applications for CMDPs with two costs and a single constraint. These two problems are rooted in queueing theory, and in particular address the allocation of a server (or pool of servers) to customers in a multiclass queueing network. A similar problem to the ones we consider is analyzed in Huang et al. [29], in which server allocation in a more complex multiclass queueing network is studied in a heavy-traffic regime. In this problem, the authors consider assigning servers to classes of triage patients or in-process (IP) patients in an emergency

department. Triage patients have a patience time, viewed as a stochastic due date. Two constrained problems are considered, in which a metric related to the IP patients is minimized subject to due date constraints placed on the triage patients, and asymptotically optimal policies are obtained. By considering due date constraints and analyzing both problems within the heavy-traffic regime, the methodology employed by the authors is much different from the one we use. Much of the groundwork for our approach is laid out in the first problem, and the insights gained from solving this problem are then used to solve the second.

The first application is server allocation in a two class, parallel queueing system. This problem is closely related to the unconstrained problem (in the same setting) of minimizing the expected weighted total cost across both classes. In the absence of abandonments, structural results are given by the $c\mu$-rule in Buyukkoc et al. [16]. However, when abandonments are introduced, obtaining structural results becomes considerably more difficult. Two reasons for this are that interchange arguments as found in Nain [36] and the uniformization technique in Lippman [35] and Serfozo [43] are no longer applicable. There are two common approaches to deal with the difficulty of proving structural results in the presence of abandonments: (1) consider a subset of the parameter space in which structural results can be obtained, and (2) consider asymptotic performance of policies. In Down, Koole and Lewis [20], the former approach is taken. The authors consider parallel queues where both classes of customers may abandon. When the service rates are equal, they provide conditions under which a priority policy is optimal. These conditions mimic that of the $c\mu$-rule, with an additional condition on the abandonment rates of each class. The second approach (asymptotic analysis) can be further divided into two categories. When focusing on the overloaded regime, the fluid model approach is taken. In Atar et al. [11], the authors show that a generalization of the

$c\mu$-rule is asymptotically optimal under a many-server fluid scaling for a general parallel queueing system with (possibly) more than two classes of customers. In the critically loaded regime a diffusion model is used in Ghamami and Ward [26],[27], Harrison and Zeevi [28], and Tezcan and Dai [47]. Arapostathis et al. [6] consider a limiting problem that is similar to the unconstrained problem we consider. Ward and Glynn [49] [50] are concerned with approximating single-class systems with abandonments by a regulated Ornstein-Uhlenbeck process in heavy traffic. Surveys of fluid and diffusion approximations for queues with abandonments can be found in Dai and He [18] and Ward [48].

Other similar unconstrained problems have been considered. Salch et al. [41] consider a stochastic scheduling problem (which can be seen as a multi-class queueing system with no arrivals) with abandonments (stochastic due dates) under two cases. In the first, jobs may abandon during service, while in the second, jobs may only abandon before service commences. Under distributional assumptions on service and patience times, structural results are obtained for each case. Argon et al. [7] considers the unconstrained problem for a related clearing system with abandonments from both classes. Ayesta et al. [12] further extends work in this direction, proving the optimality of an index rule for the scheduling problem with 1 or 2 customers in the system. For more customers, the authors derive a nearly-optimal index rule which recovers the $c\mu$-rule and coincides with the $\frac{c\mu}{\theta}$-rule under certain conditions. Argon et al. [8] and Armony and Maglaras [9] [10] consider the unconstrained problem, but where customers are given a call-back option to influence behavior. In work more directly motivated by the emergency department setting, Saghafian et al. [40] use analytic and simulation-based methods to study the effects of stratifying patients into classes with respect to various performance metrics commonly used in different phases of treatment in the ED.

A constrained optimization problem applied to call centers is considered by Gans and Zhou in [25]. The general setting they consider is similar to that considered here with a few simplifying assumptions like an infinite backlog of class 2 customers and without abandonments. In the single server case, their assumptions allow for a single-dimensional state space as discussed above (with multiple servers, there is a finite search along that dimension as well). Furthermore, their objective is to maximize the rate at which class 2 customers receive service, rather than minimizing the number of class 2 customers in system. Their problem, like ours, places a constraint on the number of class 1 customers in system. In addition to considering a different system, Gans and Zhou [25] obtain structural results only in the case where the service rates of the two classes are equal.

The work of Bhulai [15] independently obtains results for the same problem as considered by Gans and Zhou [25], although the approach taken by the latter is closer to the one we consider. Berman et al. [14] considers a system similar to that studied in Gans and Zhou [25], focusing on a class of "switching point" policies. By focusing on this class of policies, they were able to obtain explicitly the long-run average number of class 2 customers in the system. They then considered two constrained optimization problems: (1) for a fixed number of workers, minimizing the expected waiting time of front-room jobs while meeting a service level requirement for back-room, and (2) minimizing the number of workers while meeting service level constraints for both classes of jobs. Yang et al. [52] considers general multiclass systems with no abandonments in which each class is differentiated by its arrival rate and cost function that varies in the number of workers assigned to work on class each class. For the problem of allocating workers to customers in such a way so as to minimize cost subject to quality of service (waiting time) constraints for each class of customer, the authors use an MDP value function approach to

prove the following structural result for the optimal policy: if the number of customers for a particular class increases, then assign more servers to work on that class. Our work is the first (to our knowledge) to solve the constrained problem for this system, while employing an approach that largely ignores the standard value function based approach typically used to prove structural results. Furthermore, ours is the first work to attempt to generalize this approach for the broader class of two cost CMDPs.

The same approach of linking the constrained problem to a related unconstrained problem is used to attack the second problem, a server allocation problem in a two-class, tandem queueing system. Here, we look at the related unconstrained problem of minimizing the expected long-run average holding cost across two queues in tandem, as considered in Ahn et al.[1]. The structural result for that problem, a modified version of the $c\mu$-rule, is used to derive structural results for the constrained case. The problem considered in Ahn et al. [1] is similar to problems considered in Kaufman et al. [19] and Zayas-Cabán et al. [53] [54]. The former generalizes the problem by making the available servers dynamic over time, while the latter considers introducing abandonments in the downstream queue and aims to maximize the revenue collected from service completions across both queues. Duenyas et al. [22] and Irivani [30],[31] independently considered the dynamic allocation of a single flexible server in a tandem queueing system, finding an optimal policy characterized by a monotone switching curve. Earlier results for optimal service disciplines for tandem queueing systems can be found in [37],[33]. A strong collection of work on server allocation in tandem queueing systems with respect to different objectives can be found. In particular, Van Oyen et al. [38] aimed to minimize the cycle time of each job, Andradottir et al. [5],[4] considered the problem of minimizing throughput, and Javidi et al. [32] considered minimiz-

ing a dynamic version of makespan. None of these papers cover server allocation in a tandem queueing system subject to a constraint.

2.3 Preliminaries

We consider a constrained Markov decision problem (CMDP) defined by the set of objects $< \mathbb{X}, \mathbb{A}, \mathbb{P}, c_1, c_2 >$, where we focus on the infinite-horizon average cost criterion. Here

- $\mathbb{X}$ is the discrete state space,
- $\mathbb{A} = \bigcup_{x \in \mathbb{X}} \mathbb{A}(x)$ is the action space, where $\mathbb{A}(x)$ is finite for all states x,
- $\mathbb{P}(\cdot|x, a)$ describes the transition dynamics of the system, and
- $c_k(\cdot),\ k = 1, 2$ are cost functions mapping state-action pairs to the positive reals.

We should note that sometimes, when dealing with CMDPs in continuous-time, the generator function $G(\cdot|x, a)$ is used to describe the transition dynamics in place of $\mathbb{P}(\cdot|x, a)$. A stationary policy σ is a sequence $(\sigma_x)_{x \in \mathbb{X}}$ of probability distributions such that $\sigma_x(A)$ denotes the probability that an action in $A \subseteq \mathbb{A}(x)$ is chosen in state x under policy σ. We denote the class of stationary polcies by Π^S and the class of stationary deterministic policies by Π^D. With some abuse of notation, we let $\sigma_x \in \mathbb{A}(x)$ denote the action chosen by policy σ in state x under a stationary deterministic policy ($\sigma \in \Pi^D$). Every stationary policy σ induces a Markov process $X^\sigma = \{X^\sigma(t) : t \geq 0\}$. Define the long-run average expected costs

$$C_k(\sigma) := \limsup_{T \to \infty} \mathbb{E}\left[\frac{1}{T} \int_0^T c_k(X^\sigma(t)) dt\right],\ \ k = 1, 2. \tag{2.1}$$

We consider the constrained problem, B(V)

$$\min_{\sigma\in\Pi}\{C_2(\sigma) : C_1(\sigma) \leq V\}. \tag{B(V))}$$

In this chapter, we focus on how to extend structural results from related unconstrained MDPs in order to find optimal policies for B(V). In what follows, we make the assumption that every stationary policy yields a Markov process that has a single positive recurrent class (and possibly some transient states). This assumption is a technicality that makes our analysis easier and our problem well-defined. Under this assumption, we can rewrite the expected long-run average costs under a given stationary policy, σ, in terms of the induced stationary distribution, π^σ

$$C_k(\sigma) = \sum_{x\in\mathbb{X}} \pi^\sigma(x) \int_{\mathbb{A}(x)} c_k(x,a) d\sigma_x(a) \tag{2.2}$$

$$= \sum_{x\in\mathbb{X}} \sum_{a\in\mathbb{A}(x)} c_k(x,a)\pi^\sigma(x)\sigma_x(\{a\}), \quad k=1,2, \tag{2.3}$$

where the last line follows since $\mathbb{A}$ is discrete. We can relate a stationary policy and its corresponding stationary distribution by introducing the concept of an *occupation measure*, which captures the long-run average fraction of time spent in each state-action pair under a given policy. Formally, define the occupation measure to be $\phi : \mathbb{X}\times\mathbb{A} \mapsto [0,1]$ by

$$\phi(x,a) := \pi^\sigma(x)\sigma_x(\{a\}).$$

The expected long-run average costs can then be rewritten

$$C_k(\sigma) = \sum_{x\in\mathbb{X}} \sum_{a\in\mathbb{A}(x)} c_k(x,a)\phi(x,a), \quad k=1,2.$$

Under light conditions on the cost functions $c_k(\cdot), k=1,2$ (see Appendix A), the following theorem from Altman holds, which allows us to relate the constrained problem, B(V), to its Lagrangian dual problem.

Theorem 2.3.1 (adapted from Thereom 12.7 in Altman)

1. *The optimal value C_V of the problem B(V) can be computed as,*

$$C_V = \inf_{\sigma} \sup_{\gamma \geq 0} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\}. \tag{a}$$

2. *A policy σ^* is optimal for B(V) if and only if*

$$C_V = \sup_{\gamma \geq 0} \{C_2(\sigma^*) + \gamma(C_1(\sigma^*) - V)\}$$

   That is to say, σ^* attains the infimum in (a)*.*

3. *For any class of policies Π such that $\Pi^D \subseteq \Pi$,*

$$C_V = \sup_{\gamma \geq 0} \min_{\sigma \in \Pi} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\},$$

   where we can take $\Pi = \Pi^S$, the set of all stationary policies.

Theorem 2.3.1 provides conditions under which a stationary policy is optimal for B(V) via Statements 2 and 3 by showing an equivalence between B(V) and its Lagrangian dual:

$$\sup_{\gamma \geq 0} \min_{\sigma \in \Pi^S} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\}. \tag{LD(V)}$$

Unfortunately, these conditions are not particularly useful in practice, as verifying them requires knowledge of C_V. In the next section, we develop a more practical general framework for solving two cost CMDPs.

2.4 General Framework

In this section, we develop a general procedure for exploiting structure in two cost CMDPs. We first develop sufficient optimality conditions that are easier to verify

than those introduced in Theorem 2.3.1. Rewrite

$$\begin{aligned} C_V &= \sup_{\gamma \geq 0} \min_{\sigma \in \Pi^S} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\} \\ &= \sup_{\gamma \geq 0} \{ \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V\}, \end{aligned}$$

and define the function $g(\gamma) := \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V$. Note that the minimization in $g(\gamma)$,

$$\min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\}, \qquad \text{(LR}(\gamma)\text{)}$$

is an unconstrained MDP with cost function $c_\gamma(x, a) = \gamma c_1(x, a) + c_2(x, a)$. Denote this problem by LR(γ) and let O_γ denote the set of optimal stationary policies that achieve the minimum in LR(γ). The following proposition provides properties of g that help us find a stationary policy satisfying the equation in Statement 2 of Theorem 2.3.1.

Proposition 2.4.1 *The following hold for all* $\gamma \in \mathbb{R}$

1. $g(\gamma)$ *is concave in* γ.
2. *For any* $\sigma_\gamma \in O_\gamma$, $V - C_1(\sigma_\gamma) \in \partial(-g)(\gamma)$, *where* ∂f *is the subdifferential (set of all subgradients) of the function* f.
3. *If* $\gamma < \hat{\gamma}$, *and* $\sigma_\gamma \in O_\gamma$ *and* $\sigma_{\hat{\gamma}} \in O_{\hat{\gamma}}$, *then* $C_1(\sigma_\gamma) \geq C_1(\sigma_{\hat{\gamma}})$.

Proof. Note that since sums of concave functions are concave, and the minimum of concave functions is concave the first result holds. To show the second result we need to show that for any $\gamma \in \mathbb{R}, \sigma_\gamma \in O_\gamma$,

$$-g(\gamma_0) \geq -g(\gamma) + (V - C_1(\sigma_\gamma))(\gamma_0 - \gamma) \qquad \forall \gamma_0 \in \mathbb{R}.$$

Fix $\gamma \in \mathbb{R}$. We have, for any $\gamma_0 \in \mathbb{R}$,

$$\begin{aligned} g(\gamma_0) - g(\gamma) &= \min_{\sigma \in \Pi^S} \{\gamma_0 C_1(\sigma) + C_2(\sigma)\} - \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - V(\gamma_0 - \gamma) \\ &= \min_{\sigma \in \Pi^S} \{\gamma_0 C_1(\sigma) + C_2(\sigma)\} - \gamma C_1(\sigma_\gamma) - C_2(\sigma_\gamma) - V(\gamma_0 - \gamma) \\ &\leq \gamma_0 C_1(\sigma_\gamma) + C_2(\sigma_\gamma) - \gamma C_1(\sigma_\gamma) - C_2(\sigma_\gamma) - V(\gamma_0 - \gamma) \\ &= -(V - C_1(\sigma_\gamma))(\gamma_0 - \gamma). \end{aligned}$$

Hence

$$-g(\gamma_0) \geq -g(\gamma) + (V - C_1(\sigma_\gamma))(\gamma_0 - \gamma),$$

as desired.

For the remaining result, fix $\gamma \in \mathbb{R}$ and let $\delta > 0$. Let $\nu \in O_\gamma$ and $\hat{\nu} \in O_{\gamma+\delta}$. This implies

$$\begin{aligned} \gamma C_1(\nu) + C_2(\nu) &\leq \gamma C_1(\hat{\nu}) + C_2(\hat{\nu}) \\ (\gamma + \delta) C_1(\hat{\nu}) + C_2(\hat{\nu}) &\leq (\gamma + \delta) C_1(\nu) + C_2(\nu). \end{aligned}$$

Using the fact that $A \leq B$ and $C \leq D$ implies $C - B \leq D - A$ yields

$$\delta(C_1(\hat{\nu}) - C_1(\nu)) \leq 0,$$

so that $C_1(\hat{\nu}) \leq C_1(\nu)$. ∎

Suppose we find $\gamma^* \geq 0$ such that there exists an optimal policy $\sigma^* \in O_{\gamma^*}$ for $\mathrm{LR}(\gamma)$ satisfying the constraint at equality: $C_1(\sigma^*) = V$. This implies $0 \in \partial(-g)(\gamma^*)$ by way of the second statement of Proposition 2.4.1. Since $g(\gamma)$ is concave, this implies

that γ^* attains the supremum of $g(\gamma)$. Observe that

$$\begin{aligned} C_V &= \sup_{\gamma \geq 0} \{ \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V \} \\ &= \min_{\sigma \in \Pi^S} \{\gamma^* C_1(\sigma) + C_2(\sigma)\} - \gamma^* V \\ &= C_2(\sigma^*) + \gamma^*(C_1(\sigma^*) - V) \\ &= \sup_{\gamma \geq 0} \{C_2(\sigma^*) + \gamma(C_1(\sigma^*) - V)\}, \end{aligned}$$

where the last equality follows since $C_1(\sigma^*) - V = 0$. From Statement 2 of Theorem 2.3.1, σ^* is optimal for B(V). This implies sufficient optimality conditions, summarized in the following proposition.

Proposition 2.4.2 *(Sufficient optimality conditions) Suppose that* $(\sigma^*, \gamma^*) \in \Pi^S \times \mathbb{R}_+$ *satisfies*

$$\sigma^* \in O_{\gamma^*} \tag{2.4}$$

$$C_1(\sigma^*) = V. \tag{2.5}$$

The policy σ^* *is optimal for B(V).*

Given these optimality conditions, we have converted the problem of directly finding a constrained-optimal stationary policy to that of finding the optimal policy for the *appropriate* unconstrained MDP. In fact, the results in Propositions 2.4.1 and 2.4.2, combined with algorithms such as subgradient descent for convex optimization problems, yield algorithms capable of solving a CMDP by instead solving a sequence of unconstrained MDPs. Of course, such algorithms are of little practical use: one could just solve the constrained MDP directly via linear programming (possibly with some truncation methods if the state space is countably infinite), which is much faster. However, this perspective allows us to look at structural

results for unconstrained MDPs (which is well-understood) and see how they may be extended to the constrained problem. More precisely, it turns the problem of finding structured constrained-optimal policies into that of finding structured policies that are optimal for the unconstrained problem with the *correct* Lagrange multiplier. The rest of the section is dedicated to making this process more exact. Doing so involves answering some important questions:

1. Do we need the existence of a single optimal policy with the desired structural properties or do we need to show something stronger?
2. For what values of γ do these results need to hold? Clearly, the results must hold for γ^* attaining the supremum in the Lagrangian dual, but what if it is not obvious what γ^* is?
3. Assuming we are able to show structural results for the *correct* unconstrained MDP, how do the results extend? Is the same structure maintained or does it change slightly?

To answer these questions, we consider the range of class 1 costs obtainable by policies optimal for the Lagrangian relaxation $\mathrm{LR}(\gamma)$:

$$\mathcal{C}_1(\gamma) := \{C_1(\sigma) : \sigma \in O_\gamma\}.$$

Note that, for every $\gamma \geq 0$, $\mathcal{C}_1(\gamma)$ is an interval of costs (where the singleton interval is a possibility). To see this, note that for any two policies $\sigma, \sigma' \in O_\gamma$ with corresponding occupation measures ϕ, ϕ', we can create a randomized policy σ_p for $p \in [0, 1]$ corresponding to the occupation measure $\phi_p = p\phi + (1-p)\phi$, that has the same objective value as σ and σ', but whose class 1 cost is a convex combination of $C_1(\sigma)$ and $C_1(\sigma')$:

$$C_1(\sigma_p) = pC_1(\sigma) + (1-p)C_1(\sigma').$$

Hence, if $c_1, c_1' \in \mathcal{C}_1(\gamma)$ for $c_1 < c_1'$, then $[c_1, c_1'] \subseteq \mathcal{C}_1(\gamma)$.

Now suppose that we have found the optimal multiplier γ^* described in Proposition 2.4.2, and thus, the correct Lagrangian relaxation problem for which to prove structural results. We find ourselves in one of two cases. In the first case, we have that $|\mathcal{C}_1(\gamma^*)| = 1$. This means that every policy in the argmin, O_{γ^*} has the same class 1 cost. Hence, if we are able to show the *existence* of an optimal policy with the desired structural properties for this Lagrangian relaxation, then we have shown the exact same structural results hold for the constrained problem. We suspect that these types of problems are uncommon: both of the applications we consider do not fall into this case, and indeed have not found an example (other than the trivial example of a CMDP with a singleton action space in every state). However, this case is still considered for completeness.

On the other hand, if $|\mathcal{C}_1(\gamma^*)| \neq 1$, then it is uncountably infinite (since it is a continuous interval), and, thus, it is not necessary that every policy in O_{γ^*} is constrained-optimal, since not all of these policies are binding. This is indeed the case in both of the applications we study. Thus, stronger structural properties must be shown in order to extend the results to the constrained case. We seek to develop a general procedure that allows us to extend structural properties for the unconstrained problem to the constrained problem in this case. In doing so, we define the notion of "extreme" policies in a given structured class, Π^{Str}:

$$P_1^{Str} := \text{argmin}_{\Pi^{Str}} C_1(\sigma)$$
$$P_2^{Str} := \text{argmax}_{\Pi^{Str}} C_1(\sigma).$$

It should be noted that the structured class of policies, Π^{Str}, needs to be chosen carefully so that $\Pi^{Str} \subseteq O_{\gamma^*}$. In this case, if $V \in [C_1(P_1^{Str}), C_1(P_2^{Str})]$, then one of these structured policies, say $\sigma^* \in \Pi^{Str}$, satisfies the constraint at equality, and is

constrained-optimal by way of Proposition 2.4.2.

One method for constructing a class of policies with these properties is to pick the extreme policies, P_1^{Str} and P_2^{Str}, first, and then construct a sequence of policies $(\sigma_n)_{n=0}^{\infty} \subseteq O_{\gamma^*}$ conforming to some desired structural property (e.g. threshold policies) such that $\sigma_1 = P_1^{Str}$ and $\sigma_n \to P_2^{Str}$ as $n \to \infty$. Intuitively, since we start with a policy whose class 1 cost is below the constraint, and converge to a policy whose class 1 cost exceeds the constraint, we should find two policies along the sequence that "straddle" the constraint: one policy has class 1 cost below the constraint, the other above. We should then be able to find a binding policy by randomizing between these two policies, thus producing a constrained-optimal policy. Doing this requires cost continuity with respect to the mode of convergence in which $\sigma_n \to P_2^{Str}$. We find that it is most intuitive to consider *pointwise* convergence of policies: for every $x \in \mathbb{X}$, $\lim_{n\to\infty}(\sigma_n)_x(A) = (P_2^{Str})_x(A)$ for every $A \subseteq \mathbb{A}(x)$. In this context, cost continuity means that $\sigma_n \to P_2^{Str}$ pointwise implies that $\lim_{n\to\infty} C_1(\sigma_n) = C_1(P_2^{Str})$. This allows for the use of the intermediate value theorem to find an optimal policy for B(V).

In the sections that follow, we introduce specific problems in which $|\mathcal{C}_1(\gamma^*)| \neq 1$, and show how to find a constrained-optimal policy within a particular structured class.

2.5 Server Allocation in a Parallel Queueing System

Typically, in multi-class service systems, customer classes are differentiated by arrival rates, service requirements, patience times and either rewards or holding costs. When analyzing these systems, each class is modeled as having its own queue

and an optimal scheduling policy for the server(s) is sought. Take for example, the classical problem of allocating servers to queues to minimize the sum of the long-run expected average holding costs. When the patience times of customers are infinite, if the holding costs and service rates at the k^{th} station are denoted by c_k and μ_k, respectively, the well-known $c\mu$-rule is optimal. In short, one need only create an index $c_k\mu_k$ for each queue and choose the next non-empty queue with the highest index to allocate servers. The proof technique (an interchange argument) has been cross-applied to a wide range of scheduling problems (see e.g. Buyukkoc et al. [16]).

One difficulty with this formulation is that the relative importance captured by the holding costs of each customer class is not always easily quantified. As previously noted, an example is a hospital emergency department (ED) where both urgent and non-urgent patients seek treatment. In this setting, it is crucial to assure that urgent patients are served within a specified amount of time in order to avoid adverse consequences. At the same time it is also important to minimize waiting times for non-urgent patients, especially if they leave the system before being treated if wait times are too long. In this case, it is prudent to consider a constrained version of the control problem, rather than choosing holding costs for the unconstrained problem so as to drive down the average holding cost of the urgent class. This constrained problem is closer to the way practitioners approach the trade-off between prioritized classes of customers/patients. Of course, the model is also relevant in any service system (such as call centers) which places importance on the timeliness of service for a particular type of customer.

Motivated by the ED example above, we consider a two-class service system with a constraint on the holding costs for one class and the possibility of aban-

donments from the other. We formulate this problem as a constrained Markov decision process (CMDP) and use the model without abandonments to develop implementable heuristics for the more general problem.

2.5.1 System Dynamics

Suppose there are two classes of customers that arrive to a service system staffed by a single server. Class k ($k = 1, 2$) customers arrive to the system according to independent Poisson processes with rate $\lambda_k > 0$. Customer service requirements are exponentially distributed with rate 1, and are also independent of all else. The server can work on class k customers at rate $\mu_k > 0$. For each class k customer the system incurs a cost of h_k per unit time. Class 2 customers have a patience time that is exponentially distributed with rate $\beta_2 \geq 0$, after which they leave the system charging a penalty of P_2. See Figure 2.1. Our objective is to create a schedule for the server that minimizes the long-run weighted (by $h_2 > 0$) average number of customers at station 2 plus the cost incurred from abandonments, while keeping the long-run weighted (by $h_1 > 0$) average number of customers at station 1 under a given threshold denoted by V.

Note that we assume customers currently in service may abandon if they run out of patience before service is complete. There are several instances in which it is reasonable to allow abandonments during service. For example, in web chats and call centers this is common (especially if customers need to be put on hold). In the context of an emergency department it depends on what constitutes a service. Suppose a patient requires multiple procedures (x-rays, blood work, etc.). If "service" is defined from the provider's perspective (service is only completed if the patient undergoes the treatment (s)he prescribes), then it is reasonable to

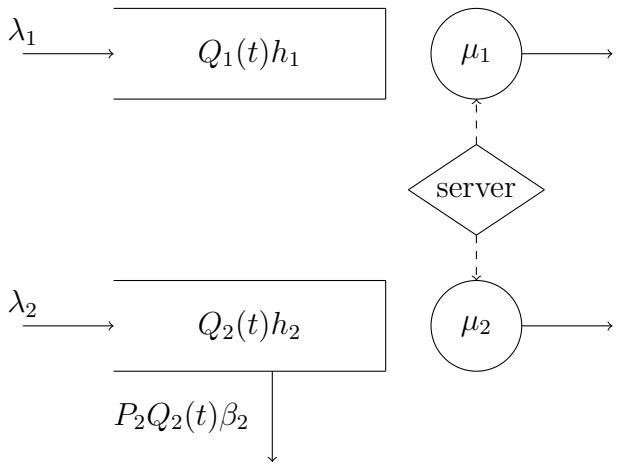


Figure 2.1: A two class queueing system with a single server. $Q_k(t)$ denotes the number of class k customers at time t.

model an abandonment during treatment if all procedures are not completed.

2.5.2 CMDP Formulation

We model this problem as a CMDP, with countable state space $\mathbb{X} := \mathbb{Z}^+ \times \mathbb{Z}^+$, where the k^{th} component corresponds to the number of class k customers in the system. The action sets for $(i, j) \in \mathbb{X}$ are

$$A(i,j) = \begin{cases} \{0,1\} & i,j > 0, \\ \{1\} & i > 0, j = 0, \\ \{0\} & i = 0, j > 0, \\ \{-1\} & \text{otherwise}, \end{cases}$$

where $a = -1$ is a "dummy" action to denote idling when the system is empty, and action $a \in \{0, 1\}$ represents serving class 1 with probability a and serving class 2 with probability $1 - a$. This interpretation, while trivial (note $\{0, 1\}$ is not the interval $[0, 1]$) given the current definition of the action space, will be helpful later

when we consider randomized policies. The transition rates are, for $i, j > 0$:

$$G((k,\ell)|(i,j),a) = \begin{cases} \lambda_1 & (k,\ell) = (i+1,j), \\ \lambda_2 & (k,\ell) = (i,j+1), \\ (1-a)\mu_2 + j\beta_2 & (k,\ell) = (i,j-1), \\ a\mu_1 & (k,\ell) = (i-1,j), \\ -\lambda_1 - \lambda_2 - j\beta_2 - a\mu_1 - (1-a)\mu_2 & (k,\ell) = (i,j), \\ 0 & \text{otherwise.} \end{cases}$$

We note that we do not allow unforced idling, making the transition rates for $i > 0, j = 0$:

$$G((k,\ell)|(i,j),a) = \begin{cases} \lambda_1 & (k,\ell) = (i+1,0), \\ \lambda_2 & (k,\ell) = (i,1), \\ \mu_1 & (k,\ell) = (i-1,0), \\ -\lambda_1 - \lambda_2 - \mu_1 & (k,\ell) = (i,j), \\ 0 & \text{otherwise.} \end{cases}$$

Similarly for $i = 0$, $j > 0$ and $i = j = 0$. We make several observations. First, the instantaneous cost rate where there are i customers at station 1 and j customers at station 2 is $ih_1 + j(h_2 + \beta_2 P_2)$. In terms of computing optimal controls we may define $\widehat{h_2} = h_2 + \beta_2 P_2$. Define the cost functions recording the number of each type of customer in the system by

$$c_1(i,j) = i \qquad\qquad c_2(i,j) = j.$$

Since the holding costs are directly proportional to the number in system, we can write our constrained problem as

$$\inf_{\sigma \in \Pi^S} \{\hat{h}_2 C_2(\sigma) : h_1 C_1(\sigma) \leq V\},$$

where, for $k = 1, 2$, $C_k(\sigma)$ is as defined in (2.1), and represents the expected long-run average number of class k customers in the system under policy σ. Note that the coefficient $\hat{h}_2$ does not affect the optimization (it can be taken outside of the infimum). Similarly, an equivalent problem is to replace the quality of service level with $V' = \frac{V}{h_1}$. Thus, we assume (without loss of generality) that the per-unit holding costs are $h_1 = \widehat{h_2} = 1$. The problem of finding a policy with minimal class 2 cost while maintaining class 1 cost of at most V is then

$$\inf_{\sigma \in \Pi^S} \{C_2(\sigma) : C_1(\sigma) \leq V\}.$$

Note that this problem, which we refer to as B(V), is the same (modulo the search over stationary policies) as that defined in Section 2.3.

2.5.3 Assumptions

In our analysis, we consider the single-server model for tractability (as opposed to an a model with N-servers). This simplification is justified in two cases.

1. Suppose instead of a single server we have multiple servers that are allowed to collaborate on a single job with an additive rate. That is, if N servers work at station k, the service rate is $N\mu_k$, $k = 1, 2$. In this case, the single server formulation is equivalent to the N service case by replacing μ_k with $\widehat{\mu_k} = N\mu_k$. Take for example the average cost optimality equations for the unconstrained scheduling problem implied for a fixed state (i, j) with $i, j \geq 1$

(cf. Corollary 7.5.10 of Sennott [42]),

$$\begin{aligned} J + h(i,j) = i + j + \lambda_1 h(i+1,j) + \lambda_2 h(i,j+1) + (1-\lambda_1-\lambda_2)h(i,j) \\ + \min_{k\in\{0,1,\ldots,N\}} \{k\mu_1(h(i-1,j) - h(i,j)) \\ + (N-k)\mu_2(h(i,j-1) - h(i,j))\}, \end{aligned} \tag{2.6}$$

where J is the optimal cost and h is the relative value function. An optimal control policy is obtained by choosing an action that achieves the minimum in (2.6). Note that the minimization (over k) is linear in k so that the optimal control is at the extremes; for $k = 0$ or N.

2. Similarly, if the servers cannot collaborate on a single job, but the workload is high, then a large proportion of time is spent in states with the number of customers in each queue is greater than N. In this case the single-server proxy is reasonable. The difficulty in this case is in deciding what to do when the number of customers for a particular queue is less than N, so that servers might need to be split between queues. This becomes increasingly important as the system load decreases.

Next we state the traffic assumption:

$$\textbf{Traffic Assumption: } \rho := \frac{\lambda_1 + \lambda_2}{\min(\mu_1, \mu_2)} < 1. \tag{T1}$$

With Assumption (T1), we verify that any stationary Markov policy induces a positive-recurrent Markov process with a unique stationary distribution.

Proposition 2.5.1 *Under Assumption* (T1)*, every stationary policy σ induces a positive-recurrent Markov chain $X^\sigma = \{(I^\sigma(t), J^\sigma(t)) : t \geq 0\}$.*

Proof. It suffices to show that the claim holds for the case without abandonments. In this setting, the transition rates are bounded, so uniformization can be applied to obtain an equivalent discrete-time Markov chain, $\tilde{X}^\sigma = \{(I_n^\sigma, J_n^\sigma) : n \in \mathbb{Z}_+\}$. Without loss of generality, assume that the uniformization factor is 1, so that the transition rates of the CTMC, X^σ, coincide with the transition probabilities of the uniformized DTMC, $\tilde{X}^\sigma$. We proceed to show positive recurrence by applying *Foster's Criterion* (cf. Theorem 4.10 of Kulkarni [34]). Define the finite subset of the state space

$$\mathbb{X}_1 := \{(i,j) \in \mathbb{X} : i + j \leq 1\}$$

and the function $f : \mathbb{X} \mapsto \mathbb{R}_+$

$$f(i,j) := i + j.$$

Pick any $(i,j) \notin \mathbb{X}_1$, and let a denote the probability of serving class 1 under policy σ in state (i,j). Since σ is non-idling,

$$\begin{aligned}
&\mathbb{E}[f(I_{n+1}^\sigma, J_{n+1}^\sigma) - f(I_n^\sigma, J_n^\sigma) | (I_n^\sigma), J_n^\sigma) = (i,j)] \\
&= \lambda_1(1) + \lambda_2(1) + a\mu_1(-1) + (1-a)\mu_2(-1) \\
&\leq \lambda_1 + \lambda_2 - \min(\mu_1, \mu_2) < 0.
\end{aligned}$$

Thus, $\tilde{X}^\sigma$ is positive-recurrent. ∎

Hence, the costs $C_k(\sigma)$ for any stationary policy σ can be expressed in terms of its induced stationary distribution, as defined in (2.2). Since the cost functions $c_1(\cdot)$ and $c_2(\cdot)$ are independent of the action taken, this representation can be simplified as follows:

$$C_k(\sigma) := \sum_{i,j \in \mathbb{X}} c_k(i,j)\pi^\sigma(i,j).$$

In addition to giving us a more convenient way to express $C_1(\cdot)$ and $C_2(\cdot)$, Assumption (T1) also implies that $C_1(\cdot)$ and $C_2(\cdot)$ are bounded over the class of stationary policies. That is,

$$\max\{\sup_{\sigma\in\Pi^S} C_1(\sigma), \sup_{\sigma\in\Pi^S} C_2(\sigma)\} \leq B$$

for some finite $B > 0$. This can be seen by considering an $M/M/1$ queueing system with birth rate $\lambda_1 + \lambda_2$ and death rate $\min\{\mu_1, \mu_2\}$, and comparing this system to the total number of customers in the two station system under any stationary policy. Note that the $M/M/1$ system has more customers in the system than the original process since it sees no abandonments and serves one of the classes of customers at a (potentially) slower rate than it would in the original system. By Assumption (T1), the $M/M/1$ system has a finite expected long-run average number of total customers in the system, and hence a finite number for each class. In fact, by using a coupling argument, we can strengthen this statement (see Proposition 2.5.2 below).

Proposition 2.5.2 *For any stationary policy $\sigma \in \Pi^S$, let $X^\sigma(\infty) = (I^\sigma(\infty), J^\sigma(\infty))$ be distributed according to π^σ. There exists a non-negative random variable $\hat{X}$ such that,*

$$\mathbb{P}(I^\sigma(\infty) + J^\sigma(\infty) \leq \hat{X}) = 1 \tag{2.7}$$

$$\mathbb{E}[\hat{X}] = \frac{\lambda_1 + \lambda_2}{\min(\mu_1, \mu_2) - (\lambda_1 + \lambda_2)} < \infty. \tag{2.8}$$

Proof. We use a coupling argument. Without loss of generality, assume $\mu_1 \leq \mu_2$. Define, for a fixed, arbitrary stationary policy σ, the Markov process $\{X^\sigma(t) = (I^\sigma(t), J^\sigma(t)) : t \geq 0\}$; Process 1. In Process 1, services at station 1 (2) are completed at rate μ_1 (μ_2). Similarly define $\{X(t) = (I(t), J(t)) : t \geq 0\}$ (called

Process 2) on the same probability space to be the queue-length process of a two station parallel queueing model that completes service at rate μ_1 regardless of the station it is serving. Assume that the Process 2 assigns the server to the same station σ does if possible and avoids idling otherwise. The probability space is defined in the following manner: the arrival times of each class of customer are the same for both processes. The service time for the m^{th} customer served in class 1 is $S_m^1 \sim \text{Exp}(1)$. The service time for the m^{th} class 2 customer is $S_m^2 \sim \text{Exp}(1)$. Note that, since $\mu_1 \leq \mu_2$, the time to complete each service (of each customer) in Process 1 are almost surely less than or equal to that in the Process 2 (for all m). This implies $I^\sigma(t) \leq I(t)$ and $J^\sigma(t) \leq J(t)$ for all t (almost surely). Hence for all $t \geq 0$

$$I(t) + J(t) \geq I^\sigma(t) + J^\sigma(t) \quad \text{a.s.} \tag{2.9}$$

Observe that the process $\{I(t) + J(t) : t \geq 0\}$ behaves as a birth-death process with birth rate $\lambda_1 + \lambda_2$ and death rate $\mu_1 = \min(\mu_1, \mu_2)$. Assumption (T1) implies that a limiting random variable, say $\hat{X}$, exists. Taking limits above (in (2.9)) yields the first result

$$\hat{X} \geq I^\sigma(\infty) + J^\sigma(\infty) \quad \text{a.s.}$$

Using Markov process theory (in particular that regarding the birth-death process) yields,

$$\mathbb{E}[\hat{X}] = \frac{\lambda_1 + \lambda_2}{\min(\mu_1, \mu_2) - (\lambda_1 + \lambda_2)} < \infty.$$

This completes the proof. ∎

Next, consider the two extremal policies that prioritize station 1 and 2 except to avoid idling; denoted P_1 and P_2, respectively. We make the following assumption

on V:

$$\textbf{RHS Assumption: } V \in (C_1(P_1), C_1(P_2)). \tag{RHS}$$

Notice that P_2 yields the highest possible expected long-run average class 1 cost among all stationary policies (since it delays serving class 1 customers as long as possible along every sample path). Thus, in the case that $V \geq C_1(P_2)$, the problem B(V) is unconstrained and it is optimal to prioritize class 2 customers. Alternatively, if $V \leq C_1(P_1)$, the problem is infeasible unless $V = C_1(P_1)$, in which case P_1 is optimal. Any V in this range we refer to as feasible and non-trivial. Under Assumptions (RHS) and (T1), we can show (see Appendix A) that Theorem 2.3.1 holds, allowing us to instead consider the equivalent Lagrangian dual problem, LD(V) (as defined in Section 2.3).

2.5.4 Cost Continuity

The goal of this section is to prove that the long-run average costs for each class are pointwise continuous over the set of stationary policies. This is used to prove some of the structural results for B(V). The result is stated more precisely in the following theorem.

Theorem 2.5.3 *For parallel queueing setting, suppose that* $(\sigma_n)_{n=0}^{\infty} \subseteq \Pi^S$ *is a sequence of policies such that* $\sigma_n(x) \to \sigma(x)$ *as* $n \to \infty$ *for each* $x = (i, j) \in \mathbb{X}$, *where* $\sigma \in \Pi^S$. *For* $k = 1, 2$, $\lim_{n\to\infty} C_k(\sigma_n) = C_k(\sigma)$.

Theorem 2.5.3 can be proved by applying the (probabilistic) dominated convergence theorem. Providing the environment to do so involves proving stationary distribution convergence and constructing a random variable that (almost surely)

bounds above the (long-run) number class 1 and class 2 customres obtained by any stationary policy. The latter is covered by Proposition 2.5.2, while the former is covered in Lemma 2.5.4 appearing below.

Lemma 2.5.4 *Let σ be a stationary policy, and let $(\sigma_n)_{n=0}^{\infty}$ be a sequence of stationary policies. Suppose that $\sigma_n \to \sigma$ pointwise. Let π^{σ_n} and π^{σ} denote the stationary distributions of the respective induced (Markov) processes. Then $\pi^{\sigma_n} \to \pi^{\sigma}$ pointwise.*

Proof. See Section A.3 of the appendix. ∎

The proof of Theorem 2.5.3 can be found in Section A.4 of the appendix. Its implications, as discussed in Section 2.4, allow us to find policies that satisfy the constraint at equality. We show that this condition is sufficient for optimality in the case without class 2 abandonments. While we have not been able to extend these results to the case with abandonments, we use this result as motivation to construct highly-structured heuristic policies that perform well numerically.

2.5.5 No Abandonments: Simplification of Optimality Conditions

We simplify the optimality conditions (2.4) and (2.5) in the absence of abandonments ($\beta_2 = 0$). To do this, we leverage the optimality of the well-known $c\mu$-rule [16] for the Lagrangian relaxation, LR(γ), along with Assumptions (RHS) and (T1), to find γ^* attaining the supremum of $g(\gamma)$ as defined in Section 2.4. Once this is done, we make the sufficient condition (2.4) extraneous by showing that

every stationary policy is optimal for LR(γ) at $\gamma = \gamma^*$: that is, $O_{\gamma^*} = \Pi^S$. Hence, to solve the constrained problem B(V), it suffices to find a stationary policy that satisfies the constraint at equality. This makes it easier to find constrained-optimal policies.

Finding the Optimal Lagrange Multiplier

Recall the priority policies P_k for $k = 1, 2$, where P_k denotes the policy that prioritizes class k. That is, policy P_k serves exhaustively at station k, switching stations only if it is empty to avoid idling. Note that the Lagrangian dual problem, LD(V), can be rewritten as

$$\sup_{\gamma \geq 0} \{ \min_{\sigma} \{ \gamma C_1(\sigma) + C_2(\sigma) \} - \gamma V \},$$

and that the minimization includes LR(γ), an unconstrained MDP with cost function $\gamma i + j$ when in state (i, j). That is to say, the problem within the minimum is the classic scheduling problem that we know has an optimal control, the $c\mu$ rule: if $\gamma \geq \frac{\mu_2}{\mu_1}$, then P_1 is optimal, and if $\gamma \leq \frac{\mu_2}{\mu_1}$, then P_2 is optimal. Hence we have

$$\begin{aligned}
C_V &= \max \left(\sup_{\gamma \in \left[0, \frac{\mu_2}{\mu_1}\right]} g(\gamma), \sup_{\gamma \geq \frac{\mu_2}{\mu_1}} g(\gamma) \right) \\
&= \max \left(\sup_{\gamma \in \left[0, \frac{\mu_2}{\mu_1}\right]} \{ C_2(P_2) + \gamma (C_1(P_2) - V) \}, \sup_{\gamma \geq \frac{\mu_2}{\mu_1}} \{ C_2(P_1) + \gamma (C_1(P_1) - V) \} \right) \\
&= \max \left(C_2(P_2) + \frac{\mu_2}{\mu_1} (C_1(P_2) - V), C_2(P_1) + \frac{\mu_2}{\mu_1} (C_1(P_1) - V) \right) \\
&= \max \left(\frac{\mu_2}{\mu_1} C_1(P_2) + C_2(P_2), \frac{\mu_2}{\mu_1} C_1(P_1) + C_2(P_1) \right) - \frac{\mu_2}{\mu_1} V,
\end{aligned}$$

where the third equality follows from the assumption that $V \in (C_1(P_1), C_1(P_2))$. Since the supremum is attained at $\gamma = \frac{\mu_2}{\mu_1}$, the previous discussion implies that both P_1 an P_2 are optimal; the two terms in the maximum are equal. Hence

we have shown that in the case without abandonments, condition (2.4) becomes $\sigma^* \in O_{\frac{\mu_2}{\mu_1}}$.

Elimination of the optimality condition (2.4)

We proceed to show that we need not consider condition (2.4) by proving that $O_{\frac{\mu_2}{\mu_1}} = \Pi^S$. In doing so, we show that to find an optimal policy for B(V) in the case with no abandonments, it suffices to find a binding policy. To show that $O_{\frac{\mu_2}{\mu_1}} = \Pi^S$, consider the Lagrangian relaxation LR(γ) at $\gamma^* = \frac{\mu_2}{\mu_1}$. Define the interior of the state space $\mathbb{X}$,

$$\widehat{\mathbb{X}} := \{(i,j) \in \mathbb{X} : i, j \geq 1\}.$$

Consider the average cost optimality equations for the scheduling problem implied by $O_{\frac{\mu_2}{\mu_1}}$ for a fixed state $(i,j) \in \widehat{\mathbb{X}}$,

$$\begin{aligned} J + h(i,j) = \frac{\mu_2}{\mu_1} i + j + \lambda_1 h(i+1,j) + \lambda_2 h(i,j+1) + (1-\lambda_1-\lambda_2)h(i,j) \\ + \min_{a \in [0,1]} \{a\mu_1(h(i-1,j) - h(i,j)) + (1-a)\mu_2(h(i,j-1) - h(i,j))\}. \end{aligned} \tag{2.10}$$

Since the holding cost in the first station for this problem is $\frac{\mu_2}{\mu_1}$ the $c\mu$ index is $\mu_1 \frac{\mu_2}{\mu_1} = \mu_2$. Similarly, the index for the second station is $\mu_2(1) = \mu_2$. Thus, the $c\mu$-rule result from Buyukkoc et al. [16] states that for this problem, both P_1 and P_2 are optimal. This implies that the minimum on the right-hand side of (2.10) is attained at both $a = 0$ and $a = 1$. Noting that the quantity inside the minimization term is linear in a, we conclude that any action in the augmented action space $\tilde{A} := [0,1]$ is optimal. Since the action space is a singleton in states (i,j) where $i = 0$ or $j = 0$, it follows that every stationary policy is average-cost optimal (see Theorem 7.2.3 and Corollary 7.5.10 of [42]). Hence, applying

Proposition 2.4.2 to find an optimal policy for B(V), it suffices to find a stationary policy σ^* with $C_1(\sigma^*) = V$.

2.5.6 No Abandonments: Constructing Optimal Control Policies

We define a class of threshold policies that we prove contains an optimal policy for any $V \in (C_1(P_1), C_1(P_2))$. This class of policies has the special property of containing optimal policies that randomize on general subsets of the state space. As alluded to earlier, this differs from the classic theory of constrained MDPs (see [3],[23]), that explains the existence of optimal policies that randomize in one state. When the state space is multidimensional (as in the present study) finding such a state may be difficult. Most importantly, in the hospital application, implementing said policy is impractical. The existence of optimal policies in the more general class simplifies our search from multidimensional (finding a single state to randomize in) to single-dimensional (finding a subset of the state space to randomize in).

Definition 2.5.5 *Let $G = (G_n)_{n=0}^{\infty}$ be a sequence of sets satisfying $G_0 = \emptyset$ and $G_n \uparrow \widehat{\mathbb{X}}$. That is, for every $x = (i,j) \in \widehat{\mathbb{X}}$, there exists N such that $x \in G_n$ for all $n \geq N$. For given $n \in \mathbb{Z}_+, p \in [0,1]$ define $\sigma_{n,p}^G$ to be a (non-idling) stationary policy satisfying, for every $x \in \hat{\mathbb{X}}$,*

$$(\sigma_{n,p}^G)_x(\{1\}) = \begin{cases} 1 & x \notin G_{n+1} \\ p & x \in G_{n+1} \setminus G_n \\ 0 & x \in G_n. \end{cases}$$

That is, $\sigma^G_{n,p}$ serves class 1 when in states not in G_{n+1}, serves class 1 with probability p when in states in G_{n+1} but not in G_n, and serves class 2 when in states not in G_n. We define the class of ***randomized-threshold policies with respect to*** *G by*

$$\Pi^G := \bigcup_{n\in\mathbb{Z}_+, p\in[0,1]} \{\sigma^G_{n,p}\}.$$

Letting $\mathcal{G}$ denote the set of all sequences $G = (G_n)_{n=0}^{\infty}$ satisfying the conditions above, define the class of ***randomized-threshold policies*** *by*

$$\Pi^{RT} := \bigcup_{G\in\mathcal{G}} \Pi^G.$$

We refer to any seqeuence of sets $G \in \mathcal{G}$ as *suitable*. It is worth noting that, for any suitable sequence G, $\sigma^G_{0,1} = P_1$ and $\sigma^G_{n,1}$ converges pointwise to P_2 as $n \to \infty$. In addition, note that $\sigma^G_{n+1,1} = \sigma^G_{n,0}$ for any n.

Remark 2.5.6 *A simple example of such a sequence G is the (half open) rectangles $G_n = \{(i,j) \in \widehat{\mathbb{X}} : j \leq n\}$ so that the decision-maker works at station 1 as long as there are less than n people at station 2 and at station 2 otherwise. The $p-$randomized threshold policy defined using this G is deemed the* ***horizontal*** *heuristic. Similarly, $G_n = \{(i,j) \in \widehat{\mathbb{X}} : i \leq n\}$ is called the* ***vertical*** *heuristic and $G_n = \{(i,j) \in \widehat{\mathbb{X}} : i + j \leq n\}$ is called the* ***total*** *heuristic.*

The main result is summarized in the theorem below.

Theorem 2.5.7 *For the parallel queue constrained server allocation problem with quality of service $V \in (C_1(P_1), C_1(P_2))$ and any suitable sequence of sets $G = (G_n)_{n=0}^{\infty}$, there exists $n^* \in \mathbb{Z}_+, p^* \in [0,1]$ so that $\sigma^G_{n^*,p^*} \in \Pi^G$ is optimal for B(V).*

Before we proceed, we need one small result.

Lemma 2.5.8 *Let $(x_n)_{n=0}^{\infty}$ be a real sequence of numbers such that $x_n \to x \in \mathbb{R}$ as $n \to \infty$. For any $v \in (x_0, x)$, there exists $m \in \mathbb{Z}_+$ such that $v \in (x_m, x_{m+1}]$.*

Proof. Since $x_n \to x > v$ and $x_0 < v$, there exists m so that $x_m \geq v > x_0$. If $x_m = v$, the result holds trivially. Otherwise, decrease m by one until $x_m \leq v$. The algorithm is guaranteed to terminate since $x_0 < v$, and results in $v \in (x_m, x_{m+1}]$, as desired. ∎

Proof of Theorem 2.5.7. Fix arbitrary $V \in (C_1(P_1), C_1(P_2))$ and a suitable sequence of sets $G = (G_n)_{n=0}^{\infty}$. Since $\sigma_{0,1}^G = P_1$ and $\sigma_{n,1}^G \to P_2$ pointwise as $n \to \infty$, Theorem 2.5.3 yields that $C_1(\sigma_{n,1}^G) \to C_1(P_2)$ as $n \to \infty$. Since $V \in (C_1(\sigma_{0,1}^G), C_1(P_2))$, by Lemma 2.5.8 there exists $n^* \in \mathbb{Z}^+$ so that $V \in (C_1(\sigma_{n^*,1}^G), C_1(\sigma_{n^*+1,1}^G)]$. Now note that $\sigma_{n^*+1,1}^G = \sigma_{n^*,0}^G$ and that $\sigma_{n^*,p}^G \to \sigma_{n^*,1}^G$ pointwise as $p \to 1$. Thus, again by Theorem 2.5.3, $C_1(\sigma_{n^*,p})$ is continuous in p on $[0, 1]$, with $C_1(\sigma_{n^*,1}) < V \leq C_1(\sigma_{n^*,0}^G)$. Applying the intermediate value theorem yields the existence of $p^* \in (0, 1)$ such that $C_1(\sigma_{n^*,p^*}) = V$. By Proposition 2.4.2, $\sigma_{n^*,p^*}^G \in \Pi^G$ is optimal for B(V). ∎

This leads us to the following algorithm for constructing binding randomized-threshold policies.

1. Choose a suitable sequence of sets $G = (G_n)_{n=0}^{\infty}$.
2. Initialize $n = 0, p = 1$. Increase n until $C_1(\sigma_{n,1}^G) < V \leq C_1(\sigma_{n+1,1})$, or equivalently

$$C_1(\sigma_{n,1}^G) < V \leq C_1(\sigma_{n,0}).$$

3. Find $p \in [0,1]$ so that $C_1(\sigma^G_{n,p}) = V$.

2.5.7 Numerical Experiments

Numerical experiments are performed on three parameter sets to test the effectiveness of the horizontal, vertical, and total classes of heuristic policies (see Remark 2.5.6) against that of the priority policies (those most likely to be implemented in a hospital setting). All experiments used the truncated state space $\mathbb{X}_{100} := \{0, 1, ..., 100\}^2$, and with the abandonment rate varying in the range $[0, 0.1]$ in increments of 0.002. The truncation of the state space allows us to calculate costs for a given policy by solving a sparse linear system. For each parameter set, we choose three values of V as follows: we first calculate the class 1 costs for P_1 ($C_1(P_1, \beta_2)$) and P_2 ($C_1(P_1, \beta_2)$) for each value of β_2. Note the added dimension to the nomenclature for the dependence on the class 2 abandonment rate. Letting $a = \max_{\beta_2} C_1(P_1, \beta_2)$ and $b = \min_{\beta_2} C_1(P_2, \beta_2)$, we use the values:

$$V_{low} = 0.75a + 0.25b$$

$$V_{med} = 0.5a + 0.5b$$

$$V_{high} = 0.25a + 0.75b$$

as our choices for V. This methodology ensures that each V is feasible and non-trivial for the entire range of abandonment rates. For each of these values, we find the optimal objective value for the (truncated) constrained problem by solving the dual LP. We then perform the algorithm described in Section 2.5.6, with a left bisection search to find p so that the class 1 cost of each heuristic policy is larger than $V - 0.0001$. The feasibility gaps for the heuristic policies are compared to that of P_2, and their optimality gaps are compared to that of P_1. Feasibility

gaps are summarized in Tables 2.2, 2.3, and 2.4 for the first, second, and third parameter sets, respectively. The minimums and maximums in these tables are taken with respect to the abandonment rate. Note that a negative number implies that the policy is feasible, and the minimum gap means that it is the furthest from the bound V. A positive gap implies that the policy is infeasible with the maximum (minimum) being the furthest (closest) from the upper bound V. The class (horizontal (h), vertical (v), or total (t)) of heuristic policy that attains the minimum feasibility gap (over all abandonment rates) is noted in parentheses. The optimality gaps are presented graphically: the optimality gap percentage for each policy is obtained, and its base-10 log is plotted as a function of abandonment rate. The first parameter set serves as the baseline case, and subsequent parameter sets are chosen to more closely mimic the dynamics expected in an emergency department. In particular, the second parameter set sees an increase in the non-urgent (class 2) patient arrival rate, and the third parameter set continues this intuition further by increasing the processing rate of less urgent patients. The parameters for the numerical experiments are summarized in the Table 2.1.

Parameter Set (Setting)	λ_1	λ_2	μ_1	μ_2
1 (baseline)	0.2	0.1	1	1
2 (ED)	0.1	0.7	1	1
3 (ED2)	0.1	0.7	1	2

Table 2.1: Parameter sets for the parallel queueing problem

The feasibility gaps from Tables 2.2, 2.3, and 2.4 show that the priority policy, P_2, violates the quality of service constraint at all levels across all three parameter sets, while the heuristic policies are all feasible and close to binding (by construction). The feasibility gap for P_2 is particularly high in the second parameter set, as a result of the increased class 2 workload (0.7 compared to 0.1 in parameter set

1). As a result, P_2 spends a lot more time serving class 2 patients, allowing class 1 patients to build up in the queue.

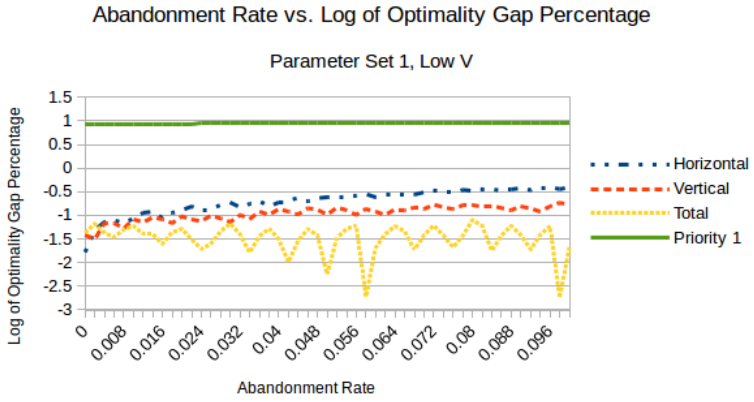
Regarding objective value performance, all three heuristic policies significantly outperform P_1 across all parameter sets and quality of service levels. For the first parameter set with a low value of V (restrictive quality of service), the optimality gap for P_1 varies (over the range of abandonments) from 8.60% to 9.10%, while the worst and best heuristic policies (among the three classes and among all abandonment rates) achieved optimality gaps of 0.394% and 0.002%, respectively. For the medium and high values of V, the optimality gap range for P_1 is $[18.81\%, 20.01\%]$ and $[31.14\%, 32.42\%]$, respectively, compared to $[0.007\%, 0.864\%]$ and $[0.017\%, 0.749\%]$ for the heuristic policies. For the second parameter set, P_1 performs closer to optimal, perhaps because of the increased class 2 workload. The optimality gap is as low as 3.29% (for low V) and is as high as 16.48% (for high V) over the range of abandonment rates. The heuristic policies still perform far better, with optimality gaps ranging from 0.020% (low V) to 0.710% (medium V). For the third and final parameter set, the optimality gaps for P_1 ranged from 6.96% (low V) to 24.23% (high V), far higher than those of the heuristic policies, which ranged from under 0.01% (low V) to as high as 0.52% (high V). Across all parameter sets, quality of service levels, and abandonment rates, every heuristic policy performed within less than 1% of optimal.

A theme to notice in all three parameter sets is that the vertical heuristic class appears to produce stronger (relative to the other heuristic policies) policies as the abandonment rate increases, while the horizontal class appears to perform better for smaller abandonment rates. A similar trend appears with respect to the quality of service level, as the vertical class performs well for high values of V, while the

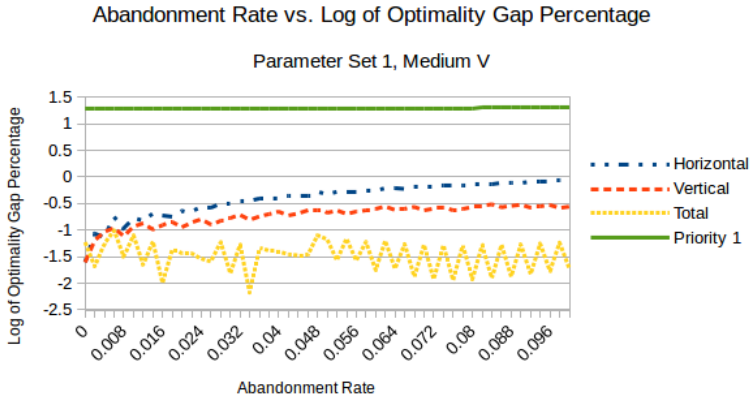
horizontal class seems to perform better for lower values of V. The total class appears to have the steadiest performance as a function of the abandonment rate. For all tested abandonment rates, and across the three parameter sets and quality of service levels, the total class of policies dominates the horizontal class. The total class also largely dominates the vertical class, except in cases where the abandonment rate is relatively large or V is high. In particular, the vertical class appears to see its most significant boost in performance in the second parameter set, where the class 2 traffic intensity is at its highest. One last observation to make is that the difference between the performance of the three classes of heuristic policies tends to increase as the abandonment rate increases. This is reasonable as they all have the same optimality gap (of 0%) when there are no abandonments, but as the abandonment rate grows, performance becomes more dependent on how often and when each policy serves class 2 patients. These numerical studies suggest that, although (slightly) more structurally complex than the priority policies, the efficiencies gained by the heuristic policies with regards to both feasibility (compared to P_2) and optimality (compared to P_1) are significant enough to consider them preferable to the priority policies.

V	Min Feas. Gap	Min Priority 2 Feas. Gap	Max Priority 2 Feas. Gap
0.2641	-0.0374 % (h)	16.05 %	20.19 %
0.2783	-0.0357 % (h)	10.16 %	14.09 %
0.2924	-0.0339 % (h)	4.83 %	8.57 %

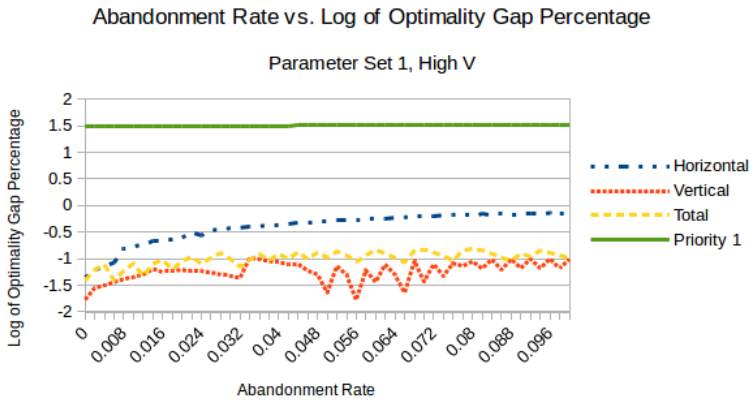
Table 2.2: Baseline case.



(a) Low bound



(b) Medium bound



(c) High bound

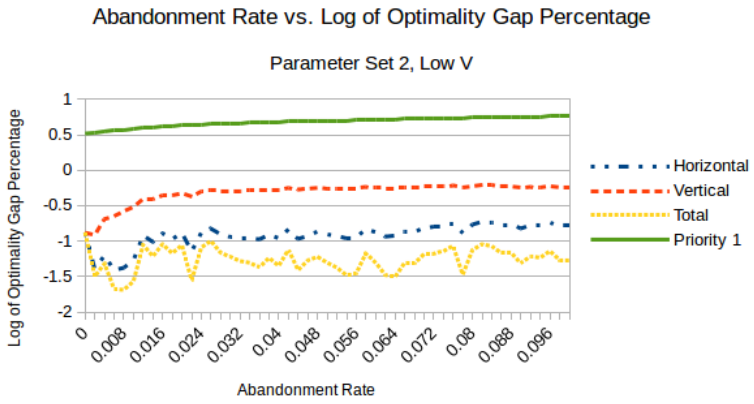
Figure 2.2: Parameter set 1 log optimality gap

V	Min Feas. Gap	Min Priority 2 Feas. Gap	Max Priority 2 Feas. Gap
0.2299	-0.0435 % (h)	155.03 %	624.84 %
0.3488	-0.0284 % (v)	68.14 %	377.89 %
0.4676	-0.0213 % (h)	25.41 %	256.44 %

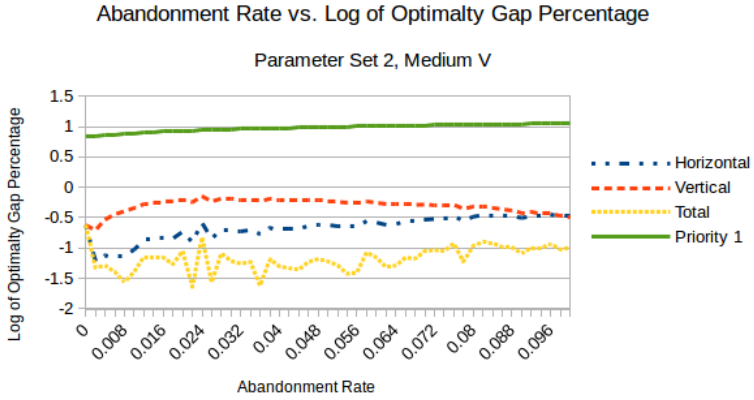
Table 2.3: Feasibility gaps for the ED example

V	Min Feas. Gap	Min Priority 2 Feas. Gap	Max Priority 2 Feas. Gap
0.1362	-0.0734 % (v)	55.34 %	69.38 %
0.1614	-0.0617 % (v)	31.15 %	43.00 %
0.1865	-0.0534 % (t)	13.48 %	23.73 %

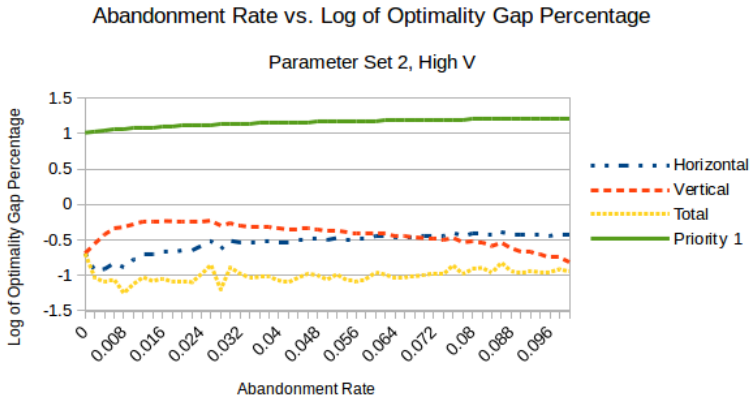
Table 2.4: Feasibility gaps for ED example with lower service level requirement.



(a) Low bound

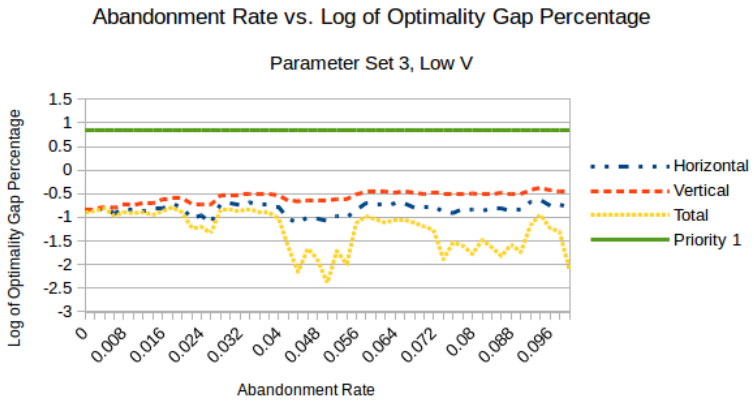


(b) Medium bound

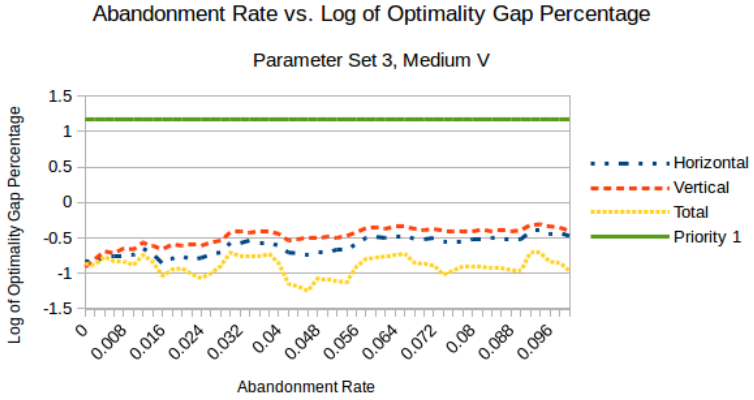


(c) High bound

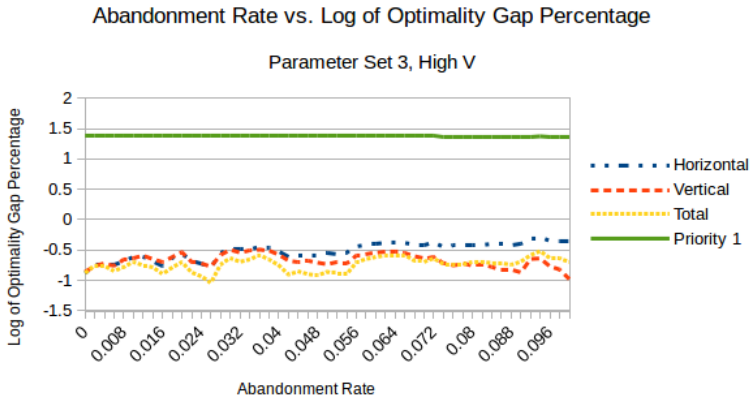
Figure 2.3: Parameter set 2 log optimality gap



(a) Low bound



(b) Medium bound



(c) High bound

Figure 2.4: Parameter set 3 log optimality gap

2.6 Server Allocation in a Tandem Queueing System

Consider a system that receives jobs that must undergo two stages of processing. This is common in many service systems, such as manufacturing processes where parts may be milled sequentially. The main application we consider is the allocation of cross-trained medical providers to different stages of care in hospital emergency departments. In many emergency departments, low-acuity (less urgent) patients must first be triaged before receiving treatment. Cross-trained providers (e.g. physicians) are capable of handling both phases of service, and it is of interest to determine how to allocate these physicians in order to balance initial delays and timely discharges. Further complicating the problem is the issue of patient abandonments: patients may choose to leave the system before receiving treatment. Zayas-Cabán et al. study this problem in [53] and [54] under the reward setting, in which a phase-dependent reward is associated with each service completion. In the former work, structural properties of the optimal policy are studied, and conditions under which priority policies are optimal are presented. The latter introduces a class of heuristic policies called K-level threshold policies, which prioritize the second phase (treatment) unless there are at least K patients in the first phase (triage). These policies are studied numerically on parameter sets generated using data from the Lutheran Medical Center (LMC) ED in New York and are shown to perform well. We consider a slightly modified version of this problem. First, rather than analyzing the reward model we consider the cost model, in which holding costs are incurred for each patient in the system at a phase-dependent rate. Second, we consider the constrained version of this problem: rather than balancing the trade-off between triage and treatment holding costs by adjusting the holding cost rates for each phase of service, we instead fix a given quality of service for the

triaged patients and aim to provide the highest quality treatment possible while meeting this service level. By Little's Law, this is equivalent to minimizing the average time spent waiting for treatment while ensuring a "sufficiently short" waiting time for triaged patients. The intuition is that when balancing such a trade-off, it is more natural to estimate desirable target waiting times for both treatment and triage than it is to assign weights to them. From a practical perspective, this approach also lends itself more easily to data-driven approaches: a decision-maker could, for example, use historical waiting time data to determine an appropriate quality of service level. As in the parallel queueing setting described in Section 2.5, we aim to prove that the broad class of randomized-threshold policies (which contains the previously mentioned K-level threshold policies) performs well for the constrained problem. In particular, we show that this class is optimal in the absence of abandonments, and performs well in numerical experiments based on those conducted in [54].

2.6.1 System Dynamics

Suppose customers arrive to a system in which they receive service in two consecutive stages. These customers are referred to as class 1 and class 2, depending on if they are waiting to be served in the first or second phase, respectively. The system is staffed by a single flexible server capable of performing both phases of service. We justify considering the single-server proxy with the same reasoning as in the parallel case. Customers arrive into the system according to a Poisson process with rate λ, and it takes the server an exponentially distributed (with rate μ_k) amount of time to process customers in the k^{th} $(k = 1, 2)$ stage of service. For each stage, there is an infinite-capacity queue for customers to wait that have yet

to be processed. In the k^{th} queue, each job incurs a holding cost of h_k per unit time. Each customer awaiting service (class 2) has an exponentially distributed patience time with rate β_2, after which the customer will leave the system. See Figure 2.5. The arrival, service, and patience times are assumed to be independent of each other. Our objective is to create a schedule for the server to minimize the expected long-run average holding cost incurred at the second station, while keeping that of the first station below the target level, V.

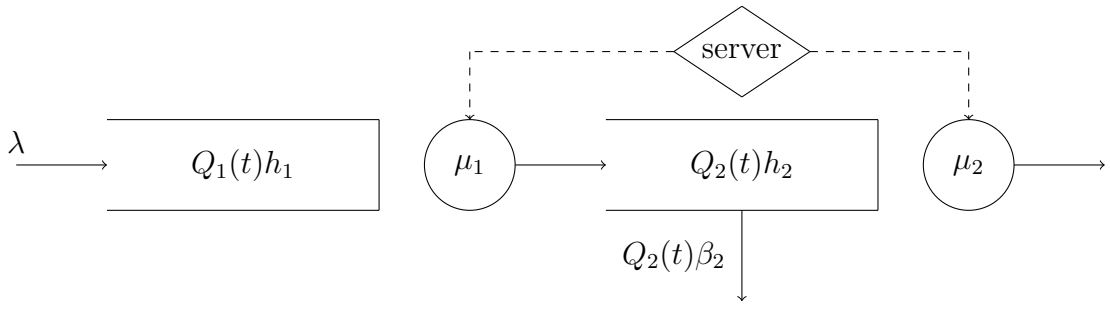


Figure 2.5: A two-class tandem queueing system with a single server. $Q_k(t)$ denotes the number of class k customers at time t.

2.6.2 CMDP Formulation

We formulate the server allocation problem as a CMDP with countable state space $\mathbb{X} := \mathbb{Z}^+ \times \mathbb{Z}^+$, where the k^{th} component corresponds to the number of class k customers in the system. The action space is

$$A(i,j) = \begin{cases} \{0,1\} & i,j > 0, \\ \{1\} & i > 0, j = 0, \\ \{0\} & i = 0, j > 0, \\ \{-1\} & i = j = 0, \end{cases}$$

where $a = -1$ is a dummy action to denote idling when the system is empty, and action a otherwise denotes the number of servers to allocate to class 1 customers (hence allocating $1 - a$ servers to class 2 customers) to avoid unnecessary idling. The transition dynamics can be captured via the following generator function (for $i, j > 0$)

$$G((k,\ell)|(i,j),a) = \begin{cases} \lambda & (k,\ell) = (i+1, j), \\ a\mu_1 & (k,\ell) = (i-1, j+1), \\ (1-a)\mu_2 - \ell\beta_2 & (k,\ell) = (i, j-1), \\ -\lambda - a\mu_1 - (1-a)\mu_2 - \ell\beta_2 & (k,\ell) = (i,j), \\ 0 & \text{otherwise.} \end{cases}$$

For $i > 0, j = 0$, we have

$$G((k,\ell)|(i,j),a) = \begin{cases} \lambda & (k,\ell) = (i+1, 0), \\ \mu_1 & (k,\ell) = (i-1, 1), \\ -\lambda - \mu_1 & (k,\ell) = (i, 0), \\ 0 & \text{otherwise,} \end{cases}$$

and similarly for $i = 0, j > 0$ and $i = 0, j = 0$. The cost functions representing the number of type k customers in the system ($k = 1, 2$) are (recall this is equivalent to the case with station dependent holding costs)

$$c_1(i, j) = i \qquad c_2(i, j) = j.$$

We define the class of stationary policies, Π^S as in Section 2.3. For a given stationary policy $\sigma \in \Pi^S$, the expected long-run average number of customers in each queue can be described by (2.1).

To simplify our analysis, we use the cost representation in (2.2). To do so, we need to ensure that every stationary policy induces a positive recurrent Markov chain with a unique stationary distribution. Proposition 2.6.1 shows that this is indeed the case, under the following traffic assumption.

$$\lambda \left(\frac{1}{\mu_1} + \frac{1}{\mu_2 + \beta_2} \right) < 1. \tag{T2}$$

Proposition 2.6.1 *Pick any stationary policy $\sigma \in \Pi^S$ and let $X^\sigma = \{X^\sigma(t) : t \geq 0\}$ denote the induced Markov chain. Under Assumption* (T2)*, X^σ is positive recurrent, and hence admits a unique stationary distribution, π^σ.*

The proof follows almost precisely (replacing 2 servers with 1) as Proposition 2.2 of Ahn et al. [1] and is omitted for brevity. We consider the constrained problem, B(V), as introduced in Section 2.3, shown below.

$$\inf_{\sigma \in \Pi^S} \{C_2(\sigma) : C_1(\sigma) \leq V\}.$$

In addition to the traffic assumption, we make the additional assumption that the quality of service level, V, is both *feasible* and *non-trivial*. To make this assumption more formal, define the priority policies P_1 and P_2 which serve exhaustively job

types 1 and 2, respectively. Then, the assumption can be rewritten

$$V \in (C_1(P_1), C_1(P_2)) .$$

The intuition of this assumption is the same as in Section 2.5: If $V \leq C_1(P_1)$, then no policy can attain the desired quality of service (or P_1 is optimal), and if $V \geq C_1(P_2)$, then it is optimal to serve class 2 customers exhaustively, and so the problem is effectively unconstrained.

2.6.3 Cost Continuity

In this section we prove that the cost continuity result from Section 2.5.4 holds in the tandem setting. Observe the following regarding the parallel setting:

1. The proof of pointwise convergence of the stationary distribution in Lemma 2.5.4 only depends on the tightness of the set of occupation measures associated with stationary policies, which in turn only relies on cost boundedness.
2. Assuming the costs are bounded, the only step needed to apply the dominated convergence theorem in the proof of Theorem 2.5.3 is to construct a (non-negative) random variable with finite expectation which bounds above (almost surely) the limiting number of class 1 and class 2 customers in the system under any stationary policy.

Hence, we can show cost continuity in the tandem setting by

1. Proving that the class 1 and class 2 costs are bounded under any stationary policy.

2. Finding an appropriate random variable to bound above the limiting number of class 1 and class 2 customers in the system under any stationary policy.

The first point is addressed by the following lemma.

Lemma 2.6.2 *Under the traffic assumption* (T2)*,*

$$\sup_{\sigma\in\Pi^S} C_k(\sigma) < \infty, \qquad k = 1, 2.$$

Proof. First note that since P_1 delays working at station 2 as long as possible it maximizes the class 2 cost. Similarly, P_2 maximizes the class 1 cost. Thus,

$$\sup_{\sigma\in\Pi^S} C_2(\sigma) = C_2(P_1),$$
$$\sup_{\sigma\in\Pi^S} C_1(\sigma) = C_1(P_2).$$

Hence, it suffices to show that $C_2(P_1)$ and $C_1(P_2)$ are finite. Using Theorem 3.2 of Ahn et al. [1] (a variant of the $c\mu$-rule), we find that for $\gamma = \gamma^* = \frac{\mu_2}{\mu_1} + 1$, both priority policies P_1 and P_2 are optimal. Thus,

$$(\frac{\mu_2}{\mu_1} + 1)C_1(P_1) + C_2(P_1) = (\frac{\mu_2}{\mu_1} + 1)C_1(P_2) + C_2(P_2).$$

We proceed to show that $C_1(P_2) + C_2(P_2) \leq B$ for some $B < \infty$. Upon showing this result, we conclude that

$$\begin{aligned} C_2(P_1) &< (\frac{\mu_2}{\mu_1} + 1)C_1(P_1) + C_2(P_1) \\ &= (\frac{\mu_2}{\mu_1} + 1)C_1(P_2) + C_2(P_2) \\ &\leq (\frac{\mu_2}{\mu_1} + 1)B. \end{aligned}$$

Similarly,

$$C_1(P_2) \leq (\frac{\mu_2}{\mu_1} + 1)(C_1(P_2) + C_2(P_2)) = (\frac{\mu_2}{\mu_1} + 1)B,$$

completing the proof. To show that $C_1(P_2) + C_2(P_2) \leq B$, note that the policy P_2, initialized from an empty system, follows a customer throughout each phase of service: each customer is served continuously with no wait time between the first and second phases of service. Hence, $C_1(P_2) + C_2(P_2)$ is the expected long-run average number of customers in a queueing system with arrivals generated according to a Poisson process with rate λ, and service times that are distributed as the sum of two independent exponential random variables, one of rate μ_1, and the other of rate μ_2. Thus, the mean and variance of the service times are given by $\frac{1}{\mu_1} + \frac{1}{\mu_2}$ and $\frac{1}{\mu_1^2} + \frac{1}{\mu_2^2}$, respectively. Using the Pollaczek-Khinchine formula yields

$$C_1(P_2) + C_2(P_2) = \lambda \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) + \frac{\lambda^2 \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right)^2 + \lambda^2 \left(\frac{1}{\mu_1^2} + \frac{1}{\mu_2^2} \right)}{2 \left(1 - \lambda \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) \right)}$$

which is finite by the traffic assumption (T2). ∎

The second point is addressed by Lemma 2.6.3.

Lemma 2.6.3 *For each stationary policy $\sigma \in \Pi^S$, define the random vectors $X^\sigma(\infty) = (I^\sigma(\infty), J^\sigma(\infty))$ to be the limiting processes of the induced Markov chain. There exists a random variable $\hat{X} \geq 0$ with finite expectation such that*

$$I^\sigma(\infty) + J^\sigma(\infty) \leq \hat{X} \quad a.s.$$

Proof. For the policy that prioritizes station 1 (P_1), define

$$\hat{X} = \lim_{t \to \infty} I^{P_1}(t) + J^{P_1}(t) = I^{P_1}(\infty) + J^{P_1}(\infty).$$

Recall from Lemma 2.6.2 we know that P_1 has finite average queuelength (so that $\mathbb{E}(\hat{X}) < \infty$). We show that P_1 yields the highest number of (total) customers in the system among all stationary policies. That is, using a coupling argument, we

show for arbitrary $\sigma \in \Pi^S$,

$$I^{P_1}(t) + J^{P_1}(t) \geq I^{\sigma}(t) + J^{\sigma}(t), \tag{2.11}$$

for all $t \geq 0$. On the same probability space (so that arrival and service requirements are common), start two Markov processes in the same state. Process 1 uses P_1 and Process 2 uses σ. We show that if a customer is in the system at time t for Process 2, then that customer must also be in the system at time t in Process 1. This implies (2.11). To do so, we show that the delay until beginning phase 2 service (for the first time) for each customer is maximized under policy P_1. Consider a particular customer, say customer x. Customer x's delay until beginning phase 2 service can be divided into two parts, the time spent on customers ahead of x and that spent on customers behind x. Since customers are served in the order that they arrived, the time spent serving customers ahead of x is the same for all non-idling policies. This leaves the time spent on customers that are behind x. Note that customers that arrived after x, can only be served before x if the service (in station 1) occurs after customer x has moved from station 1 to station 2. Since the policy P_1 spends as much time as possible on station 1 customers, this is the policy that maximizes the delay at station 2 for customer x. Taking limits as $t \to \infty$ in (2.11) yields the result. ∎

Thus, we have proved cost continuity in the tandem setting. This result is summarized in Theorem 2.6.4 below.

Theorem 2.6.4 *For tandem queueing setting, suppose that* $(\sigma_n)_{n=0}^{\infty} \subseteq \Pi^S$ *is a sequence of policies such that* $\sigma_n(x) \to \sigma(x)$ *as* $n \to \infty$ *for each* $x = (i, j) \in \mathbb{X}$, *where* $\sigma \in \Pi^S$. *For* $k = 1, 2$, $\lim_{n\to\infty} C_k(\sigma_n) = C_k(\sigma)$.

2.6.4 No Abandonments: Constructing Optimal Policies

In this section, we make use of the Lagrangian dual formulation and its resulting optimality conditions to solve B(V). Combined with results from Ahn et al. [1], we show how to construct optimal policies for B(V) from the *randomized-threshold* class. Recall the Lagrangian dual problem, LD(V):

$$\sup_{\gamma \geq 0} \left\{ \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V \right\}.$$

By the sufficient optimality conditions (2.4) and (2.5), if the supremum in LD(V) is attained by γ^*, then it suffices to find a policy σ^* that is optimal for the *unconstrained* Lagrangian relaxation,

$$\min_{\sigma \in \Pi^S} \{\gamma^* C_1(\sigma) + C_2(\sigma)\}.$$

and additionally satisfies the constraint in B(V) at equality: $C_1(\sigma^*) = V$. Similar to the parallel queueing problem considered in Section 2.5, this is an unconstrained MDP in which there are holding costs of γ^* for class 1 customers, and holding costs of 1 for class 2 customers. For general $\gamma \geq 0$, we denote this unconstrained problem by LR(γ).

Finding the Optimal Lagrange Multiplier

We leverage results from Ahn et al. [1] to find the optimal Lagrange multiplier, γ^*. The following result was proved in the case $m = 2$ in Ahn et al. [1], and the general case for $m = 1$ follows directly.

Theorem 2.6.5 (From Ahn et al. [1]) *Consider the MDP LR(γ), and suppose that*

$$\lambda \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) < 1.$$

The priority policy P_1 (P_2) is optimal if and only if $(\gamma-1)\mu_1 \geq (\leq)\mu_2$.

We leverage this result to find the optimal multiplier, γ^*. Following the notation from Theorem 2.3.1 in Section 2.3, recall that C_V denotes the optimal value of the equivalent problems B(V) and LD(V). In particular, recalling

$$g(\gamma) = \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V,$$

we have

$$C_V = \sup_{\gamma \geq 0} g(\gamma) = \max \left(\sup_{\gamma \in \left[0, \frac{\mu_2}{\mu_1}+1\right]} g(\gamma), \sup_{\gamma \geq \frac{\mu_2}{\mu_1}+1} g(\gamma) \right).$$

Theorem 2.6.5 implies the right-hand side above equals

$$\max \left(\sup_{\gamma \in \left[0, \frac{\mu_2}{\mu_1}+1\right]} \{\gamma(C_1(P_2) - V) + C_2(P_2)\}, \sup_{\gamma \geq \frac{\mu_2}{\mu_1}+1} \{\gamma(C_1(P_1) - V) + C_2(P_1)\} \right).$$

By the assumption that $V < C_1(P_2)$,

$$\begin{aligned}
&\sup_{\gamma \in \left[0, \frac{\mu_2}{\mu_1}+1\right]} \{\gamma(C_1(P_2) - V) + C_2(P_2)\} \\
&= (\frac{\mu_2}{\mu_1} + 1)(C_1(P_2) - V) + C_2(P_2) \\
&= \min_{\sigma \in \Pi^S} \left\{ (\frac{\mu_2}{\mu_1} + 1) C_1(\sigma) + C_2(\sigma) \right\} - (\frac{\mu_2}{\mu_1} + 1) V \\
&= g\left(\frac{\mu_2}{\mu_1} + 1\right).
\end{aligned}$$

Similarly, since $V > C_1(P_1)$,

$$\begin{aligned}
&\sup_{\gamma \geq \frac{\mu_2}{\mu_1}+1} \{\gamma(C_1(P_1) - V) + C_2(P_1)\} \\
&= (\frac{\mu_2}{\mu_1} + 1)(C_1(P_1) - V) + C_2(P_1) \\
&= \min_{\sigma \in \Pi^S} \left\{ (\frac{\mu_2}{\mu_1} + 1) C_1(\sigma) + C_2(\sigma) \right\} - (\frac{\mu_2}{\mu_1} + 1) V \\
&= g\left(\frac{\mu_2}{\mu_1} + 1\right).
\end{aligned}$$

Hence,

$$\begin{aligned} C_V &= \sup_{\gamma \geq 0} \left\{ \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V \right\} \\ &= \min_{\sigma \in \Pi^S} \left\{ (\frac{\mu_2}{\mu_1} + 1) C_1(\sigma) + C_2(\sigma) \right\} - (\frac{\mu_2}{\mu_1} + 1) V, \end{aligned}$$

and thus $\gamma^* = \frac{\mu_2}{\mu_1} + 1$ is the optimal Lagrange multiplier. We leverage this fact to find the set of optimal policies for LR(γ) at $\gamma = \gamma^* = \frac{\mu_2}{\mu_1} + 1$.

Finding O_{γ^*} and Solving the Constrained Problem

Consider LR(γ) with fixed $\gamma = \gamma^* = \frac{\mu_2}{\mu_1} + 1$. We characterize the set of all stationary optimal policies. First, note that for any state (i, j) with at least one of i and j equal to zero, $A(i, j)$ is a singleton. Thus, every stationary policy selects the same action at any of these states. With this in mind, it suffices to focus our attention on states $(i, j) > 0$. Consider the average-cost optimality equations (ACOE) for these states. Note that, by Theorem 2.6.5, both priority policies P_1 and P_2 are optimal. Thus, a solution (J, h) to the ACOE for this unconstrained problem satisfies, for all states $(i, j) > 0$:

$$\begin{aligned} J + h(i,j) &= \left(\frac{\mu_2}{\mu_1} + 1 \right) i + j + \lambda(h(i+1,j) - h(i,j)) \\ &\quad + \mu_1(h(i-1,j+1) - h(i,j)) + h(i,j) \\ &= \left(\frac{\mu_2}{\mu_1} + 1 \right) i + j + \lambda(h(i+1,j) - h(i,j)) \\ &\quad + \mu_2(h(i,j-1) - h(i,j)) + h(i,j). \end{aligned}$$

Hence, for any $p \in [0,1]$, we have

$$\begin{aligned} J + h(i,j) &= p(J + h(i,j)) + (1-p)(J + h(i,j)) \\ &= \lambda(h(i+1,j) - h(i,j)) + p(\mu_1(h(i-1,j+1) - h(i,j))) \\ &\quad + (1-p)\mu_2(h(i,j-1) - h(i,j)) + h(i,j) \\ &= \lambda h(i+1,j) + p\mu_1 h(i-1,j+1) + (1-p)\mu_2 h(i,j-1) \\ &\quad + (1 - \lambda - p\mu_1 - (1-p)\mu_2)h(i,j) \end{aligned}$$

Since the choice of $p \in [0,1]$ is arbitrary, we conclude that *any* $a \in \{0,1\} = A(i,j)$ attains the minimum in

$$\begin{aligned} J + h(i,j) = \min_{a \in A(i,j)} \{&\lambda h(i+1,j) + a\mu_1 h(i-1,j+1) + (1-a)\mu_2 h(i,j-1) \\ &+ (1 - \lambda - a\mu_1 - (1-a)\mu_2)h(i,j)\}. \end{aligned}$$

Hence, any $a \in A(i,j)$ attains the minimum of the ACOE, as does any a in the modified action space, $\tilde{\mathbb{A}} := [0,1]$. Since any $a \in \tilde{\mathbb{A}}$ can be interpreted as the probability of serving class 1, it follows that every stationary policy is optimal for LR(γ) at $\gamma = \gamma^* = \frac{\mu_2}{\mu_1} + 1$. Thus $O_{\gamma^*} = \Pi^S$. As in the parallel queueing system introduced in Section 2.5, this eliminates the need to consider condition (2.4), making it sufficient to find a stationary policy satisfying the constraint at equality.

The definition of the *randomized-threshold* class of stationary policies coincides with Definition 2.5.5. The main result of this section is summarized in the theorem below.

Theorem 2.6.6 *For the two-class series queue server allocation problem with quality of service constraint $V \in (C_1(P_1), C_1(P_2))$, any sequence of sets $G = (G_n)_{n=0}^{\infty}$ satisfying $G_0 = \emptyset$ and $G_n \uparrow \widehat{\mathbb{X}}$, there exists $p \in [0,1]$ and a $p-$randomized threshold policy, $\sigma^* \in \Pi^{G,p}$ that is optimal for B(V).*

It should be mentioned that, since $\Pi^S = O_{\gamma^*}$, the range $\mathcal{C}_1(\gamma^*)$ as defined in Section 2.4 is not a singleton. Thus, as was the case for the parallel queueing problem considered in Section 2.5, we need pointwise cost continuity in order to construct optimal randomized-threshold policies. Since this holds by way of Theorem 2.6.4, Theorem 2.6.6 can be proved in the same manner that Theorem 2.5.7 was proved. Since these proofs are identical, we have omitted repeating the proof for brevity. One important observation to make regarding this proof, however, is that similar analysis can be done on the *flipped* problem, $\tilde{B}(V)$:

$$\min_{\sigma \in \Pi^S} \{C_1(\sigma) : C_2(\sigma) \leq V\}, \qquad (\tilde{B}(V))$$

provided that V is both feasible and non-trivial with respect to the class 2 cost:

$$V \in (C_2(P_2), C_2(P_1)).$$

This problem can be more appropriate in some situations. For instance, in the emergency department setting, it may make more sense to impose a target quality of service on the treatment time while maintaining a short line at the triage station, rather than vice-versa.

2.6.5 Numerical Experiments

Numerical experiments are performed on three parameter sets based on experiments conducted by Zayas-Cabán et al.[54] to test the effectiveness of the horizontal, vertical, and total classes of heuristic policies (recall Remark 2.5.6). The authors determined, based on data from the Lutheran Medical Center, a range of plausible values for λ, μ_1, μ_2, and β_2, summarized in Table 2.5. Unlike our model, their model included potential patient departures from triage (so that not all triaged patients join the treatment queue) and multiple, non-collaborative servers.

Parameter	Value/Value Range
λ	$[4.2, 23.4]$
μ_1	8.57
μ_2	4.62
β_2	$[0.15, 0.8]$

Table 2.5: Parameter ranges from Zayas Cabán et al. [54]

In our numerical experiments, it is always assumed that upon being triaged, all patients move to join the treatment queue. Furthermore, the setting they considered included multiple, non-collaborative servers. We mapped this scenario to the one we consider by multiplying the base service rates for each class by a factor of N, the number of servers. That is, if there are N servers that can each serve class k at rate μ_k $(k = 1, 2)$ in their setting, we map this to our setting by considering a system with a single server which can serve class k customers at a rate of $N\mu_k$. Note that this is equivalent to considering a setting with N servers that can collaborate at an additive rate. We chose these multiplication factors such that the service rates would be high enough to satisfy the traffic condition

$$\lambda \left(\frac{1}{\mu_1} + \frac{1}{\mu_2} \right) < 1$$

for each parameter set. Three parameter sets were chosen, and can be seen in Table 2.6. For each parameter set, the abandonment rate β_2 was varied from 0.15 to 0.8 in increments of 0.013.

Parameter Set	λ	μ_1	μ_2
1	4.2	17.14	9.24
2	9.0	25.71	13.86
3	13.8	42.85	23.10

Table 2.6: Parameter sets for the tandem queueing problem

All numerical experiments were performed using a truncated state space,

$\mathbb{X}_{100} = \{0, 1, ..., 100\}^2$. The costs of the priority policies and the heuristic policies are calculated by solving a sparse linear system, and optimal values are found by solving a linear program. For each parameter set, three values of V, the quality of service requirement, are chosen by the same method used in Section 2.5.7 to ensure that each value of V tested is both feasible and non-trivial for every combination of parameter set and abandonment rate. For each parameter set, quality of service, and abandonment rate, a feasible heuristic policy is found in each class (horizontal, vertical, total) that has class 1 cost within 0.0001 of V. In contrast, the priority policy P_2 is far from feasibility in every parameter set, as shown in Tables 2.7, 2.8, and 2.9. With regards to performance, all three heuristic classes produce policies with significantly lower optimality gaps than the priority policy, P_1. In the first parameter set, for low V, the optimality gap for P_1 varies from $[21.53\%, 26.73\%]$ over the range of abandonment rates. For comparison, the worst and best heuristic policies (across all heuristic classes and abandonment rates) yield optimality gaps of 3.497% and 0.763%, respectively. For the medium and high values of V, the optimality gap ranges for P_1 are $[52.24\%, 62.09\%]$ and $[99.43\%, 108.74\%]$, respectively, compared to $[0.909\%, 12.46\%]$ and $[0.343\%, 21.20\%]$ for the heuristic policies. For the second parameter set, the optimality gaps for the heuristic policies (from 0.020% to 39.62%, both for the high value of V) are far lower than those of P_1 (from 69.29% for low V to 338.52% for high V). This trend remains the same in the third parameter set, where the optimality gaps of the heuristic policies ranges from 0.265% (high V) to 26.45% (high V). These are much lower than those of P_1, which range from 39.09% (for V) to 296.24% (for high V).

Over all parameter sets, the vertical class of policies generally performs the best, and tends to perform better for larger abandonment rates and less restrictive quality of service targets. In contrast, the horizontal class tends to perform rela-

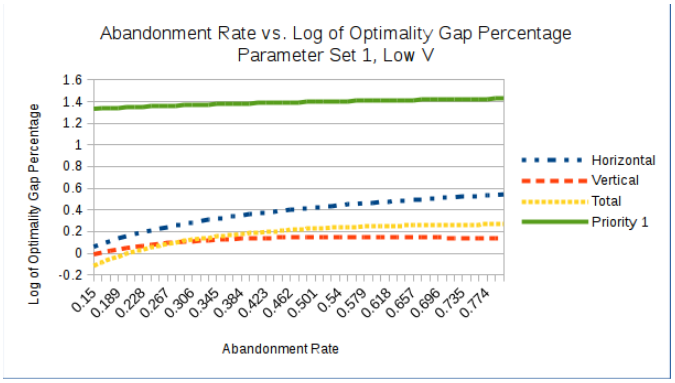
tively poorly, and generally sees an increase in optimality gap as the abandonment rate rises. However, this trend seems to reverse slightly when V is large. The total class of policies is the most variable: for some parameter sets its optimality gap is increasing in the abandonment rate, and in others it is decreasing. A similar pattern appears for the optimality gap with respect to the quality of service level. The effect of the abandonment rate on heuristic policy performance also seems to be affected by traffic intensity. The first parameter set has a traffic intensity range of $[0.66, 0.69]$ across the range of abandonment rates, and a class 1 offered load of 0.245. This is much lighter than the traffic intensities ($[0.96, 0.99]$ and $[0.90, 0.92]$) and slightly lower than the class 1 offered loads (0.35 and 0.32) of the latter two parameter sets. The optimality gap plots in Figures 2.6, 2.7, and 2.8 suggest that, when the traffic intensity is high, the effect of the abandonment rate on heuristic policy optimality gap is magnified. In particular, the optimality gap tends to decrease faster (or increase slower) in the abandonment rate for higher traffic intensities.

V	Priority 2 Feasibility Gap
0.5554	160.49%
0.7862	84.02%
1.0170	42.25%

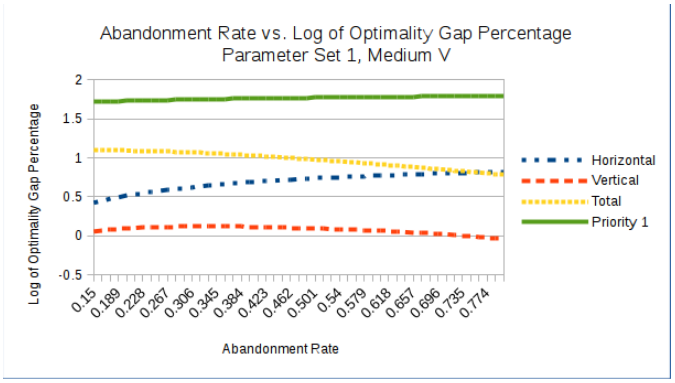
Table 2.7: Parameter Set 1 Feasibility Gaps.

V	Priority 2 Feasibility Gap
5.242	691.54%
9.945	317.20%
14.649	183.25%

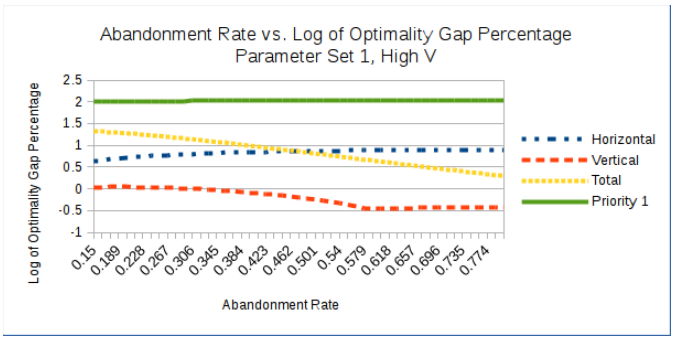
Table 2.8: Parameter Set 2 Feasibility Gaps.



(a) Low bound



(b) Medium bound

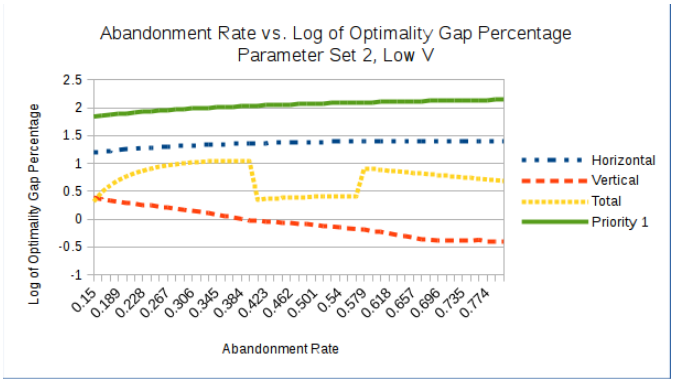


(c) High bound

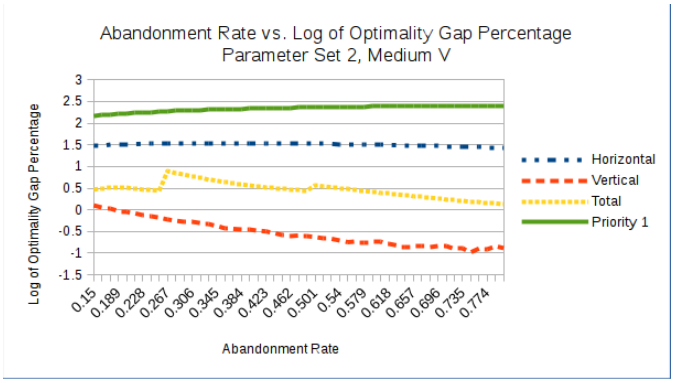
Figure 2.6: Parameter set 1 log optimality gap

V	Priority 2 Feasibility Gap
1.986	302.26%
3.497	128.45%
5.008	59.52%

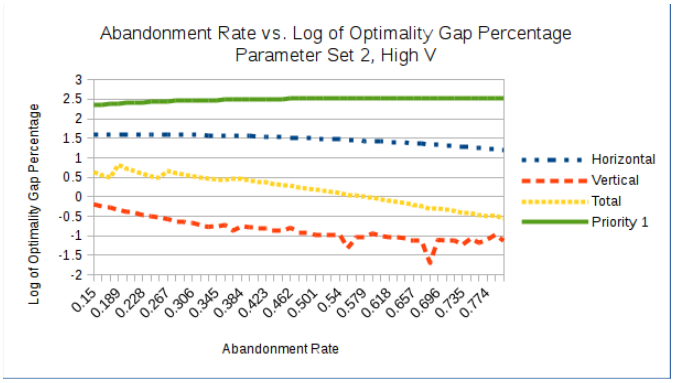
Table 2.9: Parameter Set 3 Feasibility Gaps.



(a) Low bound

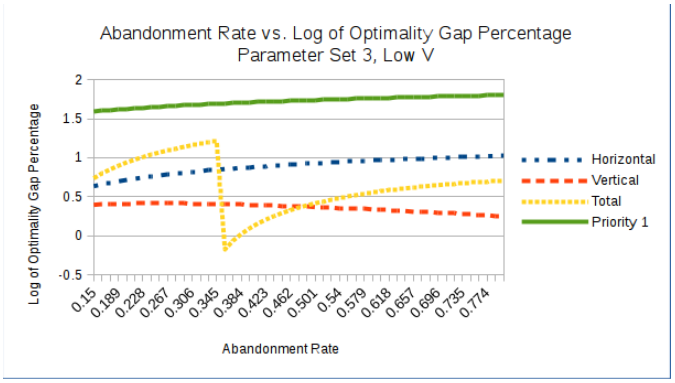


(b) Medium bound

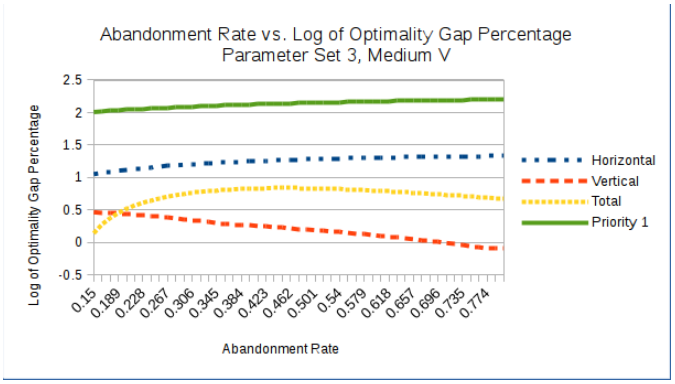


(c) High bound

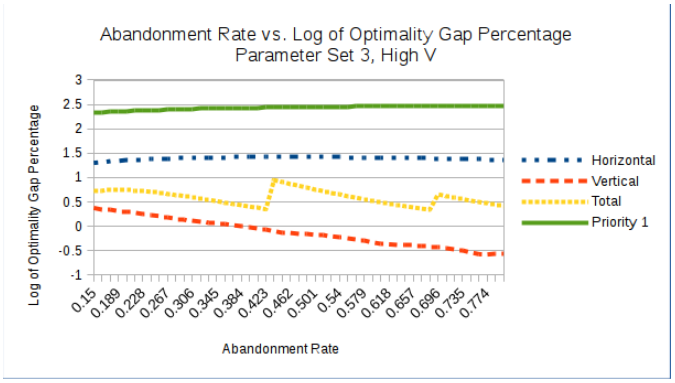
Figure 2.7: Parameter set 2 log optimality gap



(a) Low bound



(b) Medium bound



(c) High bound

Figure 2.8: Parameter set 3 log optimality gap

2.7 Conclusion and Future Work

We considered two problems falling within the general class of constrained Markov decision problems (CMDPs) with two costs and one constraint. We developed a general framework to aid in exploiting and discovering structural properties of such problems by establishing sufficient optimality conditions that lend themselves to a general procedure. At the core of this procedure is the relationship between the constrained problem and its Lagrangian dual problem, which allows us to leverage structural results in related unconstrained problems to their constrained counterparts.

The first problem considered server allocation in a two-class, parallel queueing system, motivated in part by the patient flow in emergency departments. Some of the last three authors' work, make steps toward solving the problem for lower acuity patients. We discuss the significant case of when there are either both low and high acuity patients to consider, or even just high acuity patients with various levels of injury. We model this system as a two-class queueing system, with the objective of minimizing the expected long-run average holding cost of class 2 customers while keeping that of class 1 customers under a given level, V. This is the most natural way to describe the issue to medical service providers and happens to also be a technically challenging problem arising in many other applications, including but not limited to call centers and web chat support services.

Applying a theorem from Altman [3], we used the equivalent Lagrangian dual problem to develop sufficient optimality conditions. In the case where class 2 patients have infinite patience, we leveraged the well-known $c\mu$-rule, simplifying these conditions to finding a stationary policy that satisfies the constraint at equality. We then proved weak continuity results for the expected long-run average holding

costs for both classes. Motivated by these results, we constructed classes of policies with nice structural properties.

While the results outlined above are promising, there are still some aspects of this problem to explore. It is still desirable to prove structural results for the case with class 2 abandonments. However, literature suggests [20] that solving this problem directly may be difficult. A more promising direction may be to consider asymptotic optimality of highly-structured policies, following the approach of Atar et al. [11]. Such a fluid approach may also be useful in approximating the parameters of optimal randomized-threshold policies.

In addition to considering the fluid model approach, we could attempt to solve our problem under a more general setting. Features which could make this problem both more interesting and, in certain cases (such as the ED setting), more realistic are the addition of multiple customer classes, some of which are associated with an additional constraint. It would be of particular interest to develop a set of optimality conditions and heuristics which generalizes those developed in this chapter. Adding multiple servers with the ability to cooperate (in a non-linear fashion) would more accurately describe the ED setting. Lastly, it would be interesting to develop theoretical guarantees for heuristic policies, specifically for those tested in our numerical experiments.

Our second application considered server allocation in a two-class, tandem queueing system as studied by Zayas Cabán et al. [54]. Much like the first problem, this problem is also motivated by an emergency department setting, this time dealing with lower-acuity patients who must be triaged before receiving treatment. In the absence of abandonments, we leverage the structural properties for the unconstrained problem to show the optimality of the same broad class of

randomized-threshold policies as considered in the parallel setting. Three particular subclasses of randomized-threshold policies are considered as heuristic policies for the case when abandonments are reintroduced, and these policies are shown to perform well numerically. It is also noteworthy that this class of policies remains optimal for the "flipped" problem in which the class 1 cost is to be minimized while maintaining a constraint on the class 2 cost. This is not surprising in the setting with parallel queues since the problem is symmetric. However, in the case of tandem queues this is not the case since class 2 arrivals are determined by class 1 service completions. One interesting future direction would be to consider how the performance of these policies changes, depending on which phase of service patients can abandon from, and additionally which phase of service a constraint is placed upon. Another interesting direction is to consider the constrained problem for the similar system considered by Kaufman et al. [19], in which an additional level of control is introduced via server capacity.

CHAPTER 3

EXTENSION TO MULTIPLE CLASSES: PARALLEL QUEUES

3.1 Introduction

We consider the natural extension of the server allocation problem in the parallel queueing system considered in Chapter 2. In particular, we study the case where there are K parallel queues, each representing a class of customers, with a quality of service constraint placed on a subset of higher priority classes. The purpose for this is twofold. First, many service systems (e.g. emergency departments, call centers) with multiple classes of customers have more than two classes. For example, using the motivating emergency department example from Chapter 2, patients are often divided into five "acuity levels" rather than simply being classified as "urgent" or "non-urgent". While the latter classification can be justified, stratifying the patient types further can lead to a more refined, better- performing scheduling policy. Second, a large part of the methodology used to tackle the two-class problem presented in Chapter 2 relies on the optimality of the $c\mu$-rule, a result that holds for an arbitrary number of classes [16]. We focus our attention on cases in which quality of service constraints are placed on individual classes of customers (rather than groups of customers), as placing constraints on groups of classes would complicate our analysis without providing much additional insight. We also assume that there are a relatively small number of constrained classes, since we place constraints on the "more important" customer classes. The rest of the chapter is organized as follows. In Section 3.2, we model the parallel queueing system and cover preliminaries for our analysis. Section 3.2 covers our main results. First, we consider the simple case with K classes and a single constraint,

proving the optimality of two broad classes of highly-structured policies. The first class is a randomized version of the $c\mu$-rule, while the second class is a generalization of the randomized-threshold policies introduced in Chapter 2. Using the intuition developed from this case, we partially extend these results by proving the optimality of a particular class of randomized-threshold policies for the more general case of K classes and L constraints. Finally, Section 3.4 summarizes our contributions.

3.2 System Dynamics and Model Formulation

Consider a single-server system that sees K classes of customers. Customers are differentiated by their arrival rates, processing requirements, and (potentially) holding costs. More specifically, for $k = 1, ..., K$, class k customers arrive to the system according to a Poisson process with rate λ_k, and can be processed by the server in an Exponentially distributed amount of time, with rate μ_k. Each class k customer in the system incurs a cost of h_k per unit time. The holding costs can alternatively be viewed as weights of importance among customer classes: a customer class with a higher weight is more important. We consider multiple settings in which some customer classes are of higher priority than others. For example, in an emergency department, patients are often categorized by the severity of their ailments, and typically an acuity level is assigned to each patient. Typically, higher acuity patients are more urgent than lower acuity patients, and can be seen as higher priority. When this is the case, it is often appropriate to place a constraint the patients in each acuity level deemed severe enough to warrant a target quality of service level. The goal is to devise a scheduling policy for the server to meet these quality of service constraints, while keeping the cost incurred by the less

urgent patients as low as possible. Here, the holding costs among the less urgent patients allows flexibility in modelling the relative importance of the less urgent patients, while maintaining that they are lower priority than the high-priority patients. Motivated by this setting, we consider a constrained version of a classic stochastic scheduling problem: how do we allocate the server to customers to minimize the expected long-run average cost, subject to meeting quality of service targets? We formulate this problem using the constrained Markov decision process (CMDP) framework. The state space is $\mathbb{X} := \mathbb{Z}_+^K$, where the k^{th} component ($k \in [K] := \{1, ..., k\}$) of a state $x \in \mathbb{X}$ represents the number of class k customers in the system. For each state x, the action space is $\mathbb{A}(x) := \{k \in [K] : x_k > 0\}$: the server may be allocated to work on any class that is present in the system. In the special case that $x = 0 \in \mathbb{R}^K$, we set $\mathbb{A}(x) = \{-1\}$, where -1 denotes a "dummy" action to represent (forced) idling. Note that with this definition of the action space $\mathbb{A} := \cup_{x \in \mathbb{X}} \mathbb{A}(x)$, we restrict ourselves to non-idling policies. Letting $e_k \in \mathbb{R}^K$ ($k \in [K]$) denote the vector with k^{th} component equal to 1 and all other components equal to 0, we describe the transition dynamics of the system by the generator

$$G(y|x,a) = \begin{cases} \lambda_k & y = x + e_k \\ \mu_a & y = x - \mathbb{1}\{x_a > 0\} e_a \\ -\sum_{k=1}^K \lambda_k - \mu_a \mathbb{1}\{x_a > 0\} & y = x, \end{cases}$$

where $\mathbb{1}\{\cdot\}$ denotes the indicator function. For each class k, the (immediate) cost function $c_k : \mathbb{X} \mapsto \mathbb{R}_+$ is $c_k(x) = h_k x_k$. The performance metric we consider is the expected long-run average cost for each class. In this setting, we can restrict our search for an optimal policy to the class of *stationary policies* [39]. A stationary policy assigns actions based only on the current state of the system, regardless of the point in time at which the decision is being made. More formally, a stationary

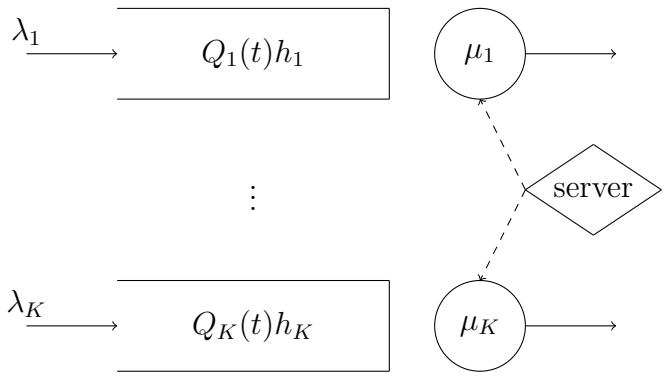


Figure 3.1: A K-class queueing system with a single server. $Q_k(t)$ denotes the number of class k customers at time t.

policy σ is a collection of probability distributions $(\sigma_x)_{x\in\mathbb{X}}$, such that $\sigma_x(A)$ denotes the probability of choosing an action in the set $A \subseteq \mathbb{A}(x)$ in state x under policy σ. We denote the class of stationary policies by Π^S. Every stationary policy σ induces a Markov chain $X^\sigma = \{X^\sigma(t) : t \geq 0\}$ on the state space $\mathbb{X}$. Hence, the expected long-run average costs of each class under σ can be computed (for $k \in [K]$)

$$C_k(\sigma) = \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T c_k(X^\sigma(t))dt\right].$$

To make our analysis easier and our problem well-defined, we impose a the traffic condition

$$\frac{\sum_{k\in[K]} \lambda_k}{\min_{k\in[K]} \mu_k} < 1. \tag{T}$$

This condition, as shown in Proposition 3.2.1, ensures that every stationary policy σ induces a positive-recurrent Markov chain X^σ, allowing for a more tractable cost representation.

Proposition 3.2.1 *Suppose that Assumption* (T) *holds. Then under any stationary policy σ, the Markov chain X^σ is positive-recurrent.*

Proof. We proceed by finding a Lyapunov function and applying Foster's Theorem. Let $e \in \mathbb{R}^K$ denote the vector of all ones. Define the region (for $B > 0$) $\mathbb{X}_B := \{x \in \mathbb{X} : x \leq Be\}$ and the function $f : \mathbb{X} \mapsto \mathbb{R}_+$ returning the total number of customers in the system in state x: $f(x) = e^T x$. Fix a state $x \notin \mathbb{X}_B$ and a stationary policy $\sigma \in \Pi^S$. For ease of notation, let $p_a = \sigma_x(\{a\})$ denote the probability that action $a \in \mathbb{A}(x)$ is chosen in state x under policy σ. For additional simplicity, we consider the equivalent discrete-time Markov chain to X^σ, $\tilde{X}^\sigma$, via *uniformization* (see Lippman [35] and Serfozo [43]) so that the one-step transitions from state x under σ can be represented probabilistically by

$$P(y|x) = \begin{cases} \lambda_k & y = x + e_k \\ p_k \mu_k & y = x - \mathbb{1}\{x_k > 0\} e_k \\ 1 - \sum_{k \in [K]} (\lambda_k + p_k \mu_k) & y = x. \end{cases}$$

Let $\tilde{X}^\sigma(n)$ denote the state of the discrete-time Markov chain $\tilde{X}^\sigma$ at time-step n. Conditioned on $\tilde{X}^\sigma(n) = x$,

$$f(\tilde{X}^\sigma(n+1)) - f(\tilde{X}^\sigma(n)) = \begin{cases} 1 & \sum_{k \in [K]} \lambda_k \\ -1 & \sum_{k \in [K]} p_k \mu_k \\ 0 & \text{otherwise.} \end{cases}$$

Since $x \notin \mathbb{X}_B$, at least one customer class is non-empty. Let $\hat{k}$ denote the non-empty class with the lowest processing rate, $\hat{\mu}$. Since σ is non-idling, $\sum_{k \in [K]} p_k \mu_k \geq \hat{\mu}$. Thus,

$$\mathbb{E}\left[f(\tilde{X}^\sigma(n+1)) - f(\tilde{X}^\sigma(n)) | \tilde{X}^\sigma(n) = x \right] \leq \sum_{k \in [K]} \lambda_k - \hat{\mu} \leq \sum_{k \in [K]} \lambda_k - \min_{k \in [K]} \mu < 0,$$

where the last inequality follows by Assumption (T). ∎

Proposition 3.2.1 implies that every stationary policy σ induces a positive-recurrent

Markov chain X^σ with a unique stationary distribution π^σ. Thus, we can rewrite the expected long-run average costs under σ more compactly (for $k \in [K]$):

$$C_k(\sigma) = \sum_{x \in \mathbb{X}} c_k(x)\pi^\sigma(x). \tag{3.1}$$

A second important consequence of Assumption (T) is that it implies that the expected long-run average costs for each class are bounded over the class of stationary polices. This is summarized in the following proposition.

Proposition 3.2.2 *Suppose that Assumption* (T) *holds. Then*

$$\sup_{\sigma \in \Pi^S} \sum_{k \in [K]} C_k(\sigma) < \infty,$$

and consequently there exists $B > 0$ *such that*

$$\max_{k \in [K]} \sup_{\sigma \in \Pi^S} h_k C_k(\sigma) \leq B.$$

Proof. Let $\hat{\mu} := \min_{k \in [K]}$. Fix any stationary policy $\sigma \in \Pi^S$, and let X^σ denote the uniformized discrete-time Markov chain induced by σ. Let $\hat{X}$ denote a discrete-time Markov chain defined on the same probability space (so that the arrival processes coincide with that seen by X^σ) but with all processing times exponentially distributed with rate $\hat{\mu}$. Note that this can be achieved by taking the processing times of class k customers, $S_1^k, S_2^k, ...$ and adding an exponentially distributed "slack" term. In particular, letting $S_1, S_2, ...$ denote the (random) service times of customers served in the process X^σ, and letting $K_1, K_2, ...$ denote their corresponding classes, define the service times of the customers served in the process $\hat{X}$ to be

$$\tilde{S}_i = \max(S_i, E_i),$$

where $E_1, E_2, ...$ is an independent sequence of random variables with E_i having rate parameter $\mu_{K_i} - \hat{\mu}$, where we take a rate parameter of 0 as meaning that E_i is deterministically 0. By properties of exponential random variables, $\tilde{S}_1, \tilde{S}_2, ...$ are exponentially distributed with rate $\hat{\mu}$. Since the two processes see the same arrivals and since the service times of $\hat{X}$ are longer almost surely, it follows that the expected long-run average total number of customers in system is at least as large as that in X^σ. Noting that the process $\hat{X}$ behaves as an $M/M/1$ queueing system with arrival rate $\sum_{k\in[K]} \lambda_k$ and service rate $\hat{\mu}$, the traffic condition (T) implies that there is a finite average number of customers in the system, $\frac{\rho}{1-\rho}$, where $\rho = \frac{\sum_{k\in[K]} \lambda_k}{\hat{\mu}}$. Since this number does not depend on the (arbitrary) policy σ, the result follows. ∎

The boundedness of costs stated in Proposition 3.2.2 is the necessary and sufficient ingredient in proving the cost continuity result found in Chapter 2. In particular, by following this proof, class k ($k \in [K]$) costs are continuous with respect to pointwise continuity of stationary policies, defined below.

Definition 3.2.3 *Let $(\sigma_n)_{n=0}^\infty$ be a sequence of stationary policies, and let σ be any stationary policy. We say that $\sigma_n \to \sigma$* ***pointwise*** *if, for every $x \in \mathbb{X}$ and every $A \subseteq \mathbb{A}(x)$,*

$$\lim_{n\to\infty} (\sigma_n)_x(A) = \sigma_x(A).$$

The next theorem follows.

Theorem 3.2.4 *Suppose that the sequence of stationary policies $(\sigma_n)_{n=0}^\infty$ converges pointwise to a stationary policy σ. Then*

$$\lim_{n\to\infty} C_k(\sigma_n) = C_k(\sigma),$$

for $k \in [K]$.

An important class of stationary policies that play an instrumental role in our analysis is the *priority policies*. Priority policies are stationary *deterministic* policies: in every state, exactly one action is chosen with probability one. Priority policies are defined by permutations of the K classes. Suppose that the mapping $\varphi : [K] \mapsto [K]$ is a permutation of $[K]$. Then, the priority policy P_φ serves the head-of-the-line customer of the highest priority as determined by the mapping φ, where class $\varphi(1)$ has the highest priority and class $\varphi(K)$ has the lowest priority. We aim to find highly-structured optimal policies for the constrained server allocation problem.

3.3 Main Results

In this section, we consider settings where some classes of customers are high-priority, and hence have quality of service constraints. We first consider the simplest case, with only one high-priority class (and one constraint), and show the optimality of two broad structural classes that are easy to implement in practice. This approach is then generalized for the case with L high-priority classes.

3.3.1 One Constraint

Consider the case where there is a single quality of service constraint. The goal is to minimize the expected long-run average holding costs of the remaining classes while satisfying this constraint. Without loss of generality, we assume that the

first class is the high-priority class. We write our constrained problem, $B_1(V)$

$$C_V = \min_{\sigma \in \Pi^S} \left\{ \sum_{k=2}^{K} h_k C_k(\sigma) : C_1(\sigma) \leq V \right\}. \qquad (B_1(V))$$

Here, $B_1(V)$ denotes the constrained optimization problem, while C_V denotes its optimal objective value. Note that, as in Chapter 2, we have taken $h_1 = 1$ without loss of generality. Also observe that, while the constraint allows us to model the differences in priority between the first class and classes $2, ..., K$, the holding costs $h_2, ..., h_K$ allow us to model preferences among lower-priority customers. Throughout the discussion on the one-constraint problem, assume without loss of generality that

$$h_2\mu_2 \geq h_3\mu_3 \geq ... \geq h_K\mu_K.$$

In our analysis, we need to place some assumptions on the quality of service target, V, to ensure that the constrained problem, $B_1(V)$, is interesting. In particular, we need to make sure that V is restrictive enough so that the unconstrained-optimal policy is infeasible, while not being so restrictive so that no feasible policy exists. To address the former, note that an optimal policy for the unconstrained problem can be obtained by the $c\mu$-rule, and prioritizes classes $2, ..., K$ in order, while prioritizing class 1 last. Let $P_{2,3,...,K,1}$ denote this policy. To ensure that this policy is infeasible for the $B_1(V)$, we assume that

$$V < C_1(P_{2,3,...,K,1}).$$

To address the latter, note that any policy prioritizing class 1 achieves the minimum possible class 1 cost of $\frac{\lambda_1}{\mu_1 - \lambda_1}$, the average number in the system for an $M/M/1$ queue with only class 1 customers. Along with the previous inequality, we have the following assumption on the target service level:

$$V \in \left(\frac{\lambda_1}{\mu_1 - \lambda_1}, C_1(P_{2,3,...,K,1}) \right). \qquad (\text{RHS}_1)$$

For the remainder of the discussion of the one-constraint problem, $B_1(V)$, we refer to any V satisfying Assumption (RHS_1) as *feasible and non-trivial*. In order to find structured optimal policies for $B_1(V)$, we want to relate it to a collection of related unconstrained problems. This is preferable since obtaining structural results for unconstrained problems allows us to appeal to a large literature on the topic. The following theorem from Altman [3] helps in this direction, and is applicable to our setting via conditions that can be verified in Appendix A.

Theorem 3.3.1 (adapted from Thereom 12.7 in Altman)

1. *The optimal value C_V of the problem $B_1(V)$ can be computed,*

$$C_V = \inf_{\sigma} \sup_{\gamma \geq 0} \left\{ \sum_{k=2}^{K} h_k C_k(\sigma) + \gamma(C_1(\sigma) - V) \right\}. \tag{a}$$

2. *A policy σ^* is optimal for $B_1(V)$ if and only if*

$$C_V = \sup_{\gamma \geq 0} \left\{ \sum_{k=2}^{K} h_k C_k(\sigma^*) + \gamma(C_1(\sigma^*) - V) \right\}$$

   That is to say, σ^* attains the infimum in (a).

3. *Suppose Π^D is the set of all stationary, deterministic policies. For any class of policies Π such that $\Pi^D \subseteq \Pi$,*

$$C_V = \sup_{\gamma \geq 0} \min_{\sigma \in \Pi} \left\{ \sum_{k=2}^{K} h_k C_k(\sigma) + \gamma(C_1(\sigma) - V) \right\},$$

   where we can take $\Pi = \Pi^S$, the set of all stationary policies.

The first statement of Theorem 3.3.1 states an equivalence between $B_1(V)$ and its Lagrangian. Statement 2 provides an optimality condition, and Statement 3 relates

this optimality condition to a related problem, the Lagrangian Dual, $LD_1(V)$ below

$$\sup_{\gamma \geq 0} \min_{\sigma \in \Pi^S} \left\{ \sum_{k=2}^{K} h_k C_k(\sigma) + \gamma(C_1(\sigma) - V) \right\}, \qquad (LD_1(V))$$

which can be rewritten

$$\sup_{\gamma \geq 0} \left\{ \min_{\sigma \in \Pi^S} \left\{ \gamma C_1(\sigma) + \sum_{k=2}^{K} h_k C_k(\sigma) \right\} - \gamma V \right\}. \qquad (3.2)$$

Note that the minimization in (3.2) is an unconstrained MDP with class 1 holding cost $h_1 = \gamma$ and class k holding costs h_k for $k = 2, ..., K$, which can be solved via the well-known $c\mu$-rule [16]: order the customer classes by the values of $h_k \mu_k$ from greatest to least, and prioritize service according to that order. We can develop general sufficient optimality conditions for $B_1(V)$ easier to use than the one stated in Statement 2 of Theorem 3.3.1, and then further simplify these conditions by using the $c\mu$-rule. First, define the function

$$g(\gamma) := \min_{\sigma \in \Pi^S} \left\{ \gamma C_1(\sigma) + \sum_{k=2}^{K} h_k C_k(\sigma) \right\} - \gamma V, \qquad (3.3)$$

the function inside the supremum in (3.2), and

$$O_\gamma := \text{argmin}_{\sigma \in \Pi^S} \left\{ \gamma C_1(\sigma) + \sum_{k=2}^{K} h_k C_k(\sigma) \right\}, \qquad (3.4)$$

the set of optimal stationary policies for the K-class server allocation problem with class 1 holding cost $h_1 = \gamma$. This problem is referred to as the Lagrangian relaxation at γ, $LR_1(\gamma)$,

$$\min_{\sigma \in \Pi^S} \left\{ \gamma C_1(\sigma) + \sum_{k=2}^{K} h_k C_k(\sigma) \right\}. \qquad (LR_1(\gamma))$$

The following properties of the function g prove to be useful in developing sufficient optimality conditions. The proof is analogous to that used in Chapter 2, and is hence omitted.

Proposition 3.3.2 *The following hold for all* $\gamma \in \mathbb{R}$

1. *$g(\gamma)$ is concave in γ.*
2. *For any $\sigma_\gamma \in O_\gamma$, $V - C_1(\sigma_\gamma) \in \partial(-g)(\gamma)$, where ∂f is the subdifferential (set of all subgradients) of the function f.*
3. *If $\gamma < \hat{\gamma}$, and $\sigma_\gamma \in O_\gamma$ and $\sigma_{\hat{\gamma}} \in O_{\hat{\gamma}}$, then $C_1(\sigma_\gamma) \geq C_1(\sigma_{\hat{\gamma}})$.*

In the same manner as shown in Chapter 2, Proposition 3.3.2 allows us to develop the sufficient optimality conditions for $B_1(V)$, summarized in Proposition 3.3.3.

Proposition 3.3.3 *(Sufficient optimality conditions) Suppose that $(\sigma^*, \gamma^*) \in \Pi^S \times \mathbb{R}_+$ satisfies*

$$\sigma^* \in O_{\gamma^*} \tag{3.5}$$

$$C_1(\sigma^*) = V. \tag{3.6}$$

The policy σ^* is optimal for $B_1(V)$.

We aim to simplify these optimality conditions by leveraging the $c\mu$-rule. In particular, we aim to eliminate Condition (3.5), allowing us to simply find a binding policy. Recall that $h_2\mu_2 \geq h_3\mu_3 \geq ... \geq h_K\mu_K$. Let φ_k be the permutation of $1, ..., K$ with $2, ..., K$ in order and 1 in the k^{th} position. Using the $c\mu$-rule, we can partition the space of multipliers $\gamma \in \mathbb{R}_+$ into regions in which each priority policy

P_{φ_k} is optimal. In fact, this partition can be stated explicitly:

$$\begin{cases} P_{\varphi_1} \text{ optimal if} & \gamma \geq \frac{h_2\mu_2}{\mu_1} \\ P_{\varphi_2} \text{ optimal if} & \gamma \in \left[\frac{h_3\mu_3}{\mu_1}, \frac{h_2\mu_2}{\mu_1}\right] \\ \vdots & \\ P_{\varphi_k} \text{ optimal if} & \gamma \in \left[\frac{h_{k+1}\mu_{k+1}}{\mu_1}, \frac{h_k\mu_k}{\mu_1}\right] \\ \vdots & \\ P_{\varphi_K} \text{ optimal if} & \gamma \in \left[0, \frac{h_K\mu_K}{\mu_1}\right]. \end{cases}$$

By Assumption (RHS_1), $V \in (C_1(P_{\varphi_1}), C_1(P_{\varphi_K}))$ and by Statement 3 of Proposition 3.3.2, $C_1(\varphi_k)$ is non-decreasing in $k \in [K]$. Thus, there exists $\ell \in \{2, ..., K\}$ such that

$$C_1(P_{\varphi_{\ell-1}}) < V \leq C_1(P_{\varphi_\ell}), \tag{3.7}$$

For notational simplicity, let R_k denote the interval of $\mathbb{R}_+$ in which P_{φ_k} is optimal, for $k \in [K]$. Furthermore, let a_k and b_k denote the endpoints of R_k, so that $R_k = [a_k, b_k]$, where we take $b_1 = \infty$ and $a_K = 0$. Note that

$$a_{k-1} = b_k = \frac{h_k\mu_k}{\mu_1}, \quad k = 2, ..., K.$$

Hence

$$\begin{aligned} P_{\varphi_{\ell-1}} &\in O_\gamma, \quad \gamma \in [a_{\ell-1}, b_{\ell-1}] \\ P_{\varphi_\ell} &\in O_\gamma, \quad \gamma \in [a_\ell, b_\ell] = [a_\ell, a_{\ell-1}]. \end{aligned}$$

Thus, by the $c\mu$-rule, both $P_{\varphi_{\ell-1}}$ and P_{φ_ℓ} are optimal for $LR_1(\gamma)$ at $\gamma = a_{\ell-1}$. Additionally, since $\ell \leq K - 1$, we know that $a_{\ell-1} = \frac{h_\ell\mu_\ell}{\mu_1}$. Note that these two priority policies only choose different actions when in states x where $x_k = 0$ for $k = 2, ..., \ell - 1$ and both $x_1, x_\ell > 0$. Denote this set of states by $\mathbb{X}_\ell$. By definition

of $\varphi_{\ell-1}, \varphi_\ell$, this means that, for states $x \in \mathbb{X}_\ell$, $P_{\varphi_{\ell-1}}$ serves class 1 and P_{φ_ℓ} instead serves class ℓ. Consider the following class of *one-randomized cμ-rule policies*, $\Pi_1^{c\mu} = \{\sigma_p : p \in [0,1]\}$, where for $p \in [0,1]$, the policy σ_p randomizes between $P_{\varphi_{\ell-1}}$ and P_{φ_ℓ} with probability p. That is, each time a state x is encountered, σ_p selects with probability p the action chosen by $P_{\varphi_{\ell-1}}$ in state x, and otherwise selects the action chosen by P_{φ_ℓ}. Observe that σ_1 coincides with $P_{\varphi_{\ell-1}}$ and σ_0 coincides with P_{φ_ℓ}. Hence

$$C_1(\sigma_1) < V \leq C_1(\sigma_0).$$

Also note that $\sigma_p \to \sigma_0$ pointwise as $p \to 0$. By Theorem 3.2.4, this implies the function $C_1(\sigma_p)$ is continuous in p on the domain $[0,1]$. Putting these two observations together, the intermediate value theorem implies the existence of some $p^* \in (0,1]$ such that $C_1(\sigma_{p^*}) = V$. That is, we can find a one-randomized $c\mu$-rule policy satisfying the optimality condition 3.6. If this policy is optimal for the $LR_1(\gamma)$ at $\gamma = a_{\ell-1} = \frac{h_\ell \mu_\ell}{\mu_1}$, then it is optimal for the constrained problem $B_1(V)$ by way of Proposition 3.3.3. Intuitively, it is reasonable that σ_{p^*} should be optimal for this problem: it is a randomization of two optimal policies. In order to show this claim rigorously, and to also open the door to different structures of optimal policies for $B_1(V)$, we adopt the approach taken in Chapter 2, looking at the average-cost optimality equations (ACOE) for the Lagrangian relaxation at $\gamma = a_{\ell-1}$. Recall that both $P_{\varphi_{\ell-1}}$ and P_{φ_ℓ} are optimal. Since these two policies choose the same action for states $x \notin \mathbb{X}_\ell$, it suffices to only look at the ACOE for states $x \in \mathbb{X}_\ell$. Let J denote the optimal cost of $LR_1(\gamma)$ at $\gamma = a_{\ell-1}$, and let h denote the (unique) relative value function with $h(z) = 0$ for $z = 0 \in \mathbb{R}^K$. Since

$P_{\varphi_{\ell-1}}$ is optimal and serves class 1, we have

$$J + h(x) = \gamma c_1(x) + \sum_{k=2}^{K} c_2(x) + \sum_{k\in[K]} \lambda_k(h(x+e_k) - h(x)) + \mu_1(h(x-e_1) - h(x)) + h(x). \tag{3.8}$$

Similarly, since P_{φ_ℓ} is optimal, and serves class ℓ, we have

$$J + h(x) = \gamma c_1(x) + \sum_{k=2}^{K} c_2(x) + \sum_{k\in[K]} \lambda_k(h(x+e_k) - h(x)) + \mu_\ell(h(x-e_\ell) - h(x)) + h(x). \tag{3.9}$$

Thus, for any $p \in [0,1]$, we can take a convex combination of equations (3.8) and (3.9) to get

$$\begin{aligned}
J + h(x) &= ph(x) + (1-p)h(x) \\
&= \gamma c_1(x) + \sum_{k=2}^{K} c_2(x) + \sum_{k\in[K]} \lambda_k(h(x+e_k) - h(x)) \\
&\quad + p\mu_1(h(x-e_1) - h(x)) + (1-p)\mu_\ell(h(x-e_\ell) - h(x)) + h(x) \\
&= \gamma c_1(x) + \sum_{k=2}^{K} c_2(x) + \sum_{k\in[K]} \lambda_k(h(x+e_k) - h(x)) \\
&\quad + \min_{p\in[0,1]} \left\{(p\mu_1(h(x-e_1) - h(x)) + (1-p)\mu_\ell(h(x-e_\ell) - h(x)) + h(x))\right\},
\end{aligned}$$

where the last equality follows since the choice of $p \in [0,1]$ was arbitrary. This last equality is exactly the ACOE for $LR_1(\gamma)$ with augmented action space

$$\tilde{\mathbb{A}}(x) = \begin{cases} \{(P_{\varphi_\ell})_x\} & x \notin \mathbb{X}_\ell \\ [0,1] & x \in \mathbb{X}_\ell, \end{cases}$$

where (with some abuse of notation) $(P_{\varphi_\ell})_x$ denotes the (deterministic) action chosen by P_{φ_ℓ} in state x, and the action space $[0,1]$ for states $x \in \mathbb{X}_\ell$ represents the probability of serving class 1 in state x, and one minus the probability of serving class ℓ in state x.

Remark 3.3.4 *Since all choices of p satisfy the ACOE for every state, it follows by way of Theorem 7.2.3 and Corollary 7.5.10 of Sennott [42] that any stationary policy coinciding with $P_{\varphi_{\ell-1}}$ and P_{φ_ℓ} in states outside of $\mathbb{X}_\ell$ and picking in any fashion between serving class 1 and class ℓ in states in $\mathbb{X}_\ell$ is optimal for $LR_1(\gamma)$ at $\gamma = a_{\ell-1}$.*

By Remark 3.3.4, σ_{p^*} satisfies optimality Condition (3.5). Since it is also is binding (and hence satisifies (3.6)), it is optimal for $B_1(V)$. The results obtained for the optimality of the one-randomized $c\mu$-rule policies is summarized in Theorem 3.3.5.

Theorem 3.3.5 *For the constrained problem, $B_1(V)$, there exists an optimal one-randomized $c\mu$-rule policy. In particular, let $\ell \in [2, K]$ satisfy* (3.7)

$$C_1(P_{\varphi_{\ell-1}}) < V \leq C_1(P_{\varphi_\ell}).$$

Then there exists $\tilde{w} \in [0, 1]$ such that the policy $\sigma_{\tilde{w}}$ randomizing between $P_{\varphi_{\ell-1}}$ (with probability w) and P_{φ_ℓ} (with probability $1 - w$) is optimal for $B_1(V)$.

Another interesting class of policies to consider for solving $B_1(V)$ is a generalization of the *randomized-threshold policies* found in Chapter 2.

Definition 3.3.6 *A sequence of sets $G = (G_n)_{n=0}^{\infty}$ is called **suitable** if $G_0 = \emptyset$ and $G_n \uparrow \mathbb{R}^2_{++}$, where $\mathbb{R}^2_{++}$ denotes the set of strictly positive two-dimensional vectors. Denote the class of suitable sequences of sets by $\mathcal{G}$.*

Definition 3.3.7 *Let ℓ satisfy* (3.7)*, and let $G = (G_n)_{n=0}^{\infty}$ be a suitable sequence of sets. The class of **randomized-threshold policies with respect to** G is*

$$\Pi^G := \{\sigma^G_{n,p} : n \in \mathbb{Z}_+, p \in [0, 1]\},$$

where, for $n \in \mathbb{Z}_+, p \in [0,1]$*, the non-idling stationary policy* $\sigma^G_{n,p}$ *is defined to choose the same action as the priority policies* $P_{\varphi_{\ell-1}}$ *and* P_{φ_ℓ}*, except in states* $x \in \mathbb{X}_\ell$*, in which*

$$(\sigma^G_{n,p})_x(\{1\}) = 1 - (\sigma^G_{n,p})_x(\{\ell\}) = \begin{cases} 1 & (x_1, x_\ell) \notin G_{n+1} \\ p & (x_1, x_\ell) \in G_{n+1} \setminus G_n \\ 0 & (x_1, x_\ell) \in G_n. \end{cases}$$

Furthermore, the class of ***randomized-threshold policies*** *is defined as*

$$\Pi^{RT} := \bigcup_{G \in \mathcal{G}} \Pi^G.$$

Note that, by Remark 3.3.4, every $\sigma \in \Pi^{RT}$ is optimal for $LR_1(\gamma)$ at $\gamma = a_{\ell-1}$. Thus, if a randomized-threshold policy is binding, then it is optimal for $B_1(V)$. Similar to Chapter 2, we show that for any suitable sequence of sets $G = (G_n)_{n=0}^\infty$, we can find a binding policy $\sigma_{\tilde{n},\tilde{p}} \in \Pi^G$. Fix such a sequence G. Note that, for any $p \in [0,1]$, the policy $\sigma^G_{0,p}$ is equivalent to the priority policy $P_{\varphi_{\ell-1}}$. Note that, for any $p \in [0,1]$, $\sigma^G_{n,p} \to P_{\varphi_\ell}$ pointwise as $n \to \infty$. As in Chapter 2, we can use Theorem 3.2.4 to find $n^* \in \mathbb{Z}_+$ such that

$$C_1(\sigma^G_{n^*,1}) < V \leq C_1(\sigma^G_{n^*+1,1}).$$

By definition, for every $n \in \mathbb{Z}_+$, the policies $\sigma^G_{n,0}$ and $\sigma^G_{n+1,1}$ are equivalent. Hence,

$$C_1(\sigma^G_{n^*,1}) < V \leq C_1(\sigma^G_{n^*,0}).$$

Noting that, for every $n \in \mathbb{Z}_+$, $\sigma^G_{n,p} \to \sigma^G_{n,0}$ as $p \downarrow 0$, applying Theorem 3.2.4 and the intermediate value theorem yields there exists $p^* \in [0,1)$ such that

$$C_1(\sigma^G_{n^*,p^*}) = V,$$

and hence $\sigma^G_{n^*,p^*}$ is optimal for $B_1(V)$. The result is summarized in Theorem 3.3.8.

Theorem 3.3.8 *For any suitable sequence of sets $G = (G_n)_{n=0}^{\infty}$, there exists $n^* \in \mathbb{Z}_+, p^* \in [0,1)$ such that the randomized-threshold policy $\sigma_{n^*,p^*}^G \in \Pi^G$ is optimal for $B_1(V)$.*

The procedure for finding such a randomized-threshold policy is outlined below.

1. Choose a suitable sequence of sets $G = (G_n)_{n=0}^{\infty}$.
2. Initialize $n = 0, p = 1$. Increase n until $C_1(\sigma_{n,1}^G) < V \leq C_1(\sigma_{n+1,1}^G)$, or equivalently

$$C_1(\sigma_{n,1}^G) < V \leq C_1(\sigma_{n,0}).$$

3. Find p so that $C_1(\sigma_{n,p}^G) = V$.

3.3.2 L Constraints

We aim to extend the results of Section 3.3.1 to the case where there are multiple high-priority classes, each with its own quality of service target. In particular, we consider the case where there are L high-priority classes. We assume without loss of generality that the first L classes are of higher priority, and denote the target quality of service levels by $V = (V_1, ..., V_L)^T$. We also assume without loss of generality that the classes $L+1, ..., K$ are ordered so that

$$h_{L+1}\mu_{L+1} \geq ... \geq h_K\mu_K.$$

The constrained optimization problem we consider is $B_L(V)$

$$C_V = \min_{\sigma \in \Pi^S} \left\{ \sum_{k=L+1}^{K} h_k C_k(\sigma) : C(\sigma) \leq V \right\}, \qquad (B_L(V))$$

where $C(\sigma) = (C_1(\sigma), ..., C_L(\sigma))^T$ for a stationary policy σ. We let C_V denote the optimal value of $B_L(V)$. In what follows, let $S_1 = [L]$ (the constrained set) and $S_2 = [K] \setminus S_1$ (the unconstrained set) for notational convenience.

Much of our analysis for this setting depends on the workload process induced by a stationary policy. The results we state regarding this process and its relation to the long-run average number in system can be easily verified as in Shanthikumar and Yao [44]. Fix a stationary policy $\sigma \in \Pi^S$ and let $X^\sigma = \{X^\sigma(t) : t \geq 0\}$ denote its induced Markov chain. Consider the workload process $W^\sigma = \{W^\sigma(t) : t \geq 0\}$, where $W_k^\sigma(t)$ denotes the amount of class k work in the system at time t under policy σ. In particular,

$$W_k^\sigma(t) = \sum_{i=1}^{X_k^\sigma(t)} U^k_{D_k^\sigma(t)+i},$$

where, for $k = 1, ..., K$, $(U_n^k)_{n=1}^\infty$ is an i.i.d. sequence of exponential random variables with rate μ_k and $D_k^\sigma(t)$ denotes the number of class k departures by time t under policy σ. Note that by the memoryless property, the remaining service requirement at time t, $U^k_{D_k^\sigma(t)+2}$, for the class k customer at the head of the line is still exponentially distributed with rate μ_k. Also note that for all $t \geq 0$ and for every class k, the service times $U^k_{D_k^\sigma(t)+2}, ..., U^k_{D_k^\sigma(t)+X_k^\sigma(t)}$ are independent of the number of class k customers in system, $X_k^\sigma(t)$. By Wald's identity,

$$\begin{aligned}
\mathbb{E}[W_k^\sigma(t)] &= \mathbb{E}\left[\sum_{i=1}^{X_k^\sigma(t)} U^k_{D_k^\sigma(t)+i}\right] \\
&= \mathbb{E}\left[U_{D_k^\sigma(t)+1}\right] + \mathbb{E}\left[\sum_{i=2}^{X_k^\sigma(t)} U^k_{D_k^\sigma(t)+i}\right] \\
&= \frac{1}{\mu_k} + \frac{\mathbb{E}[X_k^\sigma(t) - 1]}{\mu_k} \\
&= \frac{\mathbb{E}[X_k^\sigma(t)]}{\mu_k}.
\end{aligned}$$

Furthermore, by the traffic assumption T, the steady-state random vector $X^\sigma(\infty)$ exists almost surely as $t \to \infty$. As a result, the steady-state workload $W^\sigma(\infty)$ also exists and satisfies, for every class k,

$$C_k(\sigma) = \mathbb{E}[X_k^\sigma(\infty)] = \mu_k \, \mathbb{E}[W_k^\sigma(\infty)] =: \bar{W}_k(\sigma).$$

Furthermore, the average workload vector $\bar{W}(\sigma)$ satisfies a conservation law. For a set of classes S, let $\Pi(S)$ denote the set of policies that is non-idling with respect to S: the server always works on a class in S unless there are none in the system. We have, for every $S \subseteq [K]$,

$$\sum_{k \in S} \bar{W}_k(\sigma) = w(S), \quad \sigma \in \Pi(S). \tag{3.10}$$

Here $w(S)$ denotes the (constant) expected total workload induced by any policy in $\Pi(S)$. For notational convenience, in the analysis that follows, we use $\Pi(S)$ to refer to *stationary* policies that are non-idling with respect to S, since we are primarily interested in stationary policies. We can use Statement (3.10) to impose the following restrictions on the service target vector, V.

$$w(U_1) < \sum_{k \in U_1} \frac{V_k}{\mu_k} < w(U_1 \cup U) - w(U), \quad U_1 \subseteq S_1, U \subseteq [K] \setminus U_1. \tag{RHS}$$

Intuitively, the lower bounds of Assumption RHS state that the service targets must be attainable: $w(U_1)$ is the minimum achievable expected workload over classes in U_1 by any policy. On the other hand, the upper bounds ensure that the service targets are restrictive enough so that any subset of unconstrained classes cannot be prioritized over any subset of constrained classes: Among policies that are non-idling in $U_1 \cup U$, $w(U_1 \cup U) - w(U)$ is the expected workload over classes in U_1 achieved by prioritizing classes in U over those in U_1. To see this, take any such policy, σ. Since σ is non-idling in $U_1 \cup U$,

$$\sum_{k \in U_1 \cup U} \frac{C_k(\sigma)}{\mu_k} = w(U_1 \cup U).$$

Since σ also prioritizes classes in U over those in U_1, it must also be non-idling with respect to classes in U. That is,

$$\sum_{k\in U} \frac{C_k(\sigma)}{\mu_k} = w(U).$$

Thus,

$$\sum_{k\in U_1} \frac{C_k(\sigma)}{\mu_k} = \sum_{k\in U_1\cup U} \frac{C_k(\sigma)}{\mu_k} - \sum_{k\in U} \frac{C_k(\sigma)}{\mu_k} = w(U_1 \cup U) - w(U).$$

We again seek structural properties of $B_L(V)$ by applying Theorem 3.3.1 and looking at the Lagrangian dual

$$C_V = \sup_{\gamma\in\mathbb{R}_+^L} \min_{\sigma\in\Pi^S} \left\{ \sum_{k\in S_2}^{K} h_k C_k(\sigma) + \gamma^T (C(\sigma) - V) \right\}, \qquad (LD_L(V))$$

which can be rewritten

$$\sup_{\gamma\in\mathbb{R}_+^L} \left\{ \min_{\sigma\in\Pi^S} \left\{ \sum_{k\in S_1} \gamma_k C_k(\sigma) + \sum_{k\in S_2} h_k C_k(\sigma) \right\} - \gamma^T V \right\}. \qquad (3.11)$$

Note that the minimization in (3.11) is the unconstrained server allocation problem with holding costs γ_k for $k \in S_1$ and h_k for $k \in S_2$.

$$\min_{\sigma\in\Pi^S} \left\{ \sum_{k\in S_1} \gamma_k C_k(\sigma) + \sum_{k\in S_2} h_k C_k(\sigma) \right\}. \qquad (LR_L(\gamma))$$

Hence for any $\gamma \in \mathbb{R}_+^L$, an optimal policy for $LR_L(\gamma)$ can be found via the $c\mu$-rule. Denote the set of all optimal stationary policies for $LR_L(\gamma)$ by O_γ. Similar to Section 3.3.1, define the function $g : \mathbb{R}_+^L \mapsto \mathbb{R}$

$$g(\gamma) := \min_{\sigma\in\Pi^S} \left\{ \sum_{k\in S_1} \gamma_k C_k(\sigma) + \sum_{k\in S_2} h_k C_k(\sigma) \right\} - \gamma^T V. \qquad (3.12)$$

The properties of g as stated in Proposition 3.3.2 extend to the general setting, displayed in Proposition 3.3.9. The proof is analogous, and is omitted for brevity.

Proposition 3.3.9 *The following hold for all* $\gamma \in \mathbb{R}_+^L$

1. *$g(\gamma)$ is concave in γ.*
2. *For any $\sigma_\gamma \in O_\gamma$, $V - C(\sigma_\gamma) \in \partial(-g)(\gamma)$, where ∂f is the subdifferential (set of all subgradients) of the function f.*
3. *Let $k \in [K]$, $\gamma \in \mathbb{R}^L_+$ and $\hat{\gamma}(k) = \gamma + \delta e_k$ for $\delta > 0$. If $\sigma^* \in O_\gamma$ and $\hat{\sigma}^* \in O_{\hat{\gamma}(k)}$, then $C_k(\sigma^*) \geq C_k(\hat{\sigma}^*)$. Here $e_k \in \mathbb{R}^L$ denotes the vector of all zeros with 1 in the k^{th} component.*

Recall $B_L(V)$ and $LD_L(V)$ have the same optimal value. By Assumption RHS, $B_L(V)$ is non-trivial: any optimal priority policy found via the $c\mu$-rule for the problem without constraints is infeasible. Thus, the optimal value C_V must be strictly larger than that of the unconstrained problem. Since this value is equal to $g(0)$, by definition, the supremum of g is not attained at $\gamma = 0$. Suppose now $B_L(V)$ is infeasible, then $C_V = \infty$. By the concavity of g, we have that every subgradient in $\partial(-g)(\gamma)$ is negative for every $\gamma \in \mathbb{R}^L_+$ and the supremum in (3.11) is not attained. Thus, if $B_L(V)$ is feasible, then there exists $\gamma^* \in \mathbb{R}^L_+$ so that $0 \in \partial(-g)(\gamma^*)$. Note that this is equivalent to the sufficient optimality conditions in Proposition 3.3.10 below.

Proposition 3.3.10 *(Sufficient optimality conditions) Suppose that $(\sigma^*, \gamma^*) \in \Pi^S \times \mathbb{R}_+$ satisfies*

$$\sigma^* \in O_{\gamma^*} \tag{3.13}$$

$$C(\sigma^*) = V. \tag{3.14}$$

The policy σ^* is optimal for $B_L(V)$.

Observe that, by the previous discussion, the γ^* in Proposition 3.3.10 is the one attaining the supremum in g. We can use the $c\mu$-rule to find the value of γ^*, given that $B_L(V)$ is feasible. This allows us to simplify the optimality conditions.

Proposition 3.3.11 *Under Assumption RHS, if $B_L(V)$ is feasible, then the supremum in* (3.11) *is attained by $\gamma^* \in \mathbb{R}_+^L$ satisfying*

$$\gamma_1^*\mu_1 = ... = \gamma_L^*\mu_L = h_{L+1}\mu_{L+1}.$$

Proof. Note that, since $B_L(V)$ is feasible, $C_V < \infty$ by Theorem 3.3.1 from Altman, and thus the supremum of g is attained by some γ^*. First we show that $\gamma_k^*\mu_k \geq h_{L+1}\mu_{L+1}$ for all $k \in S_1$. If not, then there exists some $\bar{k} \in S_1$ with $\gamma_{\bar{k}}^*\mu_{\bar{k}} < h_{L+1}\mu_{L+1}$. Let $\bar{S} \subseteq S_1$ denote the set of classes with the lowest $c\mu$ index. For notational convenience, enumerate these classes by $\bar{S} = \{k_1, ..., k_{|\bar{S}|}\}$. By the $c\mu$-rule, there exists an optimal policy σ^* that prioritizes classes in some set $U \ni L+1$ over all those in $\bar{S}$, and prioritizes classes in $\bar{S}$ in the order $k_1, ..., k_{\bar{S}}$. Thus,

$$\begin{aligned}
\frac{C_{k_1}(\sigma^*)}{\mu_{k_1}} &= w(U \cup \{k_1\}) - w(U) > \frac{V_{k_1}}{\mu_{k_1}} \\
&\vdots \\
\frac{C_{k_j}(\sigma^*)}{\mu_{k_j}} &= w(U \cup \{k_1, ..., k_j\}) - w(U \cup \{k_1, ..., k_{j-1}) > \frac{V_{k_j}}{\mu_{k_j}} \\
&\vdots \\
\frac{C_{k_{|\bar{S}|}}(\sigma^*)}{\mu_{k_{|\bar{S}|}}} &= w(U \cup \bar{S}) - w(U \cup \{k_1, ..., k_{|\bar{S}|-1}\}) > \frac{V_{|\bar{S}|}}{\mu_{|\bar{S}|}}.
\end{aligned}$$

Hence $C_k(\sigma^*) - V_k > 0$ for all $k \in \bar{S}$. For sufficiently small $\delta > 0$, we can take

$$\gamma_k' = \begin{cases} \gamma_k^* + \frac{\delta}{\mu_k} & k \in \bar{S} \\ \gamma_k^* & k \notin \bar{S}. \end{cases}$$

Notice that $\sigma^* \in O_{\gamma'}$ since the ordering of the new $c\mu$ indices is unchanged. Thus

$$g(\gamma') - g(\gamma^*) = \delta \sum_{k \in \bar{S}} \frac{C_k(\sigma^*) - V_k}{\mu_k} > 0,$$

contradicting that γ^* attains the supremum in g. We can then proceed forward assuming that $\gamma_k^*\mu_k \geq h_{L+1}\mu_{L+1}$ for all $k \in S_1$. Now suppose that there exists $\hat{k} \in S_1$ so that $\gamma_{\hat{k}}^*\mu_{\hat{k}} > h_{L+1}\mu_{L+1}$. Let $\hat{S} \subseteq S_1$ denote the set of classes in S_1 with the highest $c\mu$ index. By the $c\mu$-rule, there exists an optimal policy σ^* that is non-idling with respect to classes in $\hat{S}$. Thus,

$$\sum_{k\in\hat{S}} \frac{C_k(\sigma^*)}{\mu_k} = w(\hat{S}) < \sum_{k\in\hat{S}} \frac{V_k}{\mu_k}.$$

For sufficiently small $\delta > 0$, we can construct γ' such that

$$\gamma_k' = \begin{cases} \gamma_k^* - \frac{\delta}{\mu_k} & k \in \hat{S} \\ \gamma_k^* & k \notin \hat{S}. \end{cases}$$

Notice that $\sigma^* \in O_{\gamma'}$. Thus

$$g(\gamma') - g(\gamma) = -\delta \sum_{k\in\bar{S}} \frac{C_k(\sigma^*) - V_k}{\mu_k} > 0,$$

again contradicting that γ^* attains the supremum in g. This completes the proof.

■

Note that we can use Proposition 3.3.11 along with the $c\mu$-rule to find optimal priority policies for $LR_L(\gamma)$ at $\gamma = \gamma^*$. Also observe that, if an oracle could provide $\partial(-g)(\gamma^*)$ for this γ^*, then the feasibility of $B_L(V)$ could be determined: if $0 \in \partial(-g)(\gamma^*)$, then $B_L(V)$ is feasible, otherwise it is infeasible. In fact, our main result implies that Assumption RHS results in a feasible problem: by constructing a binding policy that is optimal for $LR_L(\gamma)$ at $\gamma = \gamma^*$, we have shown that $0 \in \partial(-g)(\gamma^*)$.

In order to more easily construct such optimal policies, we find a subset of O_{γ^*} in which finding a binding policy is simple. In particular, O_{γ^*} contains policies that treats the constrained classes, S_1, and the highest weighted unconstrained

class, $L+1$, as a single class with higher holding cost than classes $L+2, ..., K$, and serves according to the $c\mu$-rule. That is, any such policy can serve classes $1, ..., L+1$ arbitrarily, so long as they are prioritized over classes $L+2, ..., K$, which must be prioritized in order. To see this, consider any state x in which there are two non-empty classes in $[L+1]$. Without loss of generality, assume that these are classes 1 and 2. Let $h(\cdot), J$ denote the unique pair of relative value function and objective value for the ACOE of $LR_L(\gamma)$ at $\gamma = \gamma^*$ satisfying $h(z) = 0$ for $z = 0 \in \mathbb{R}^K$. By the $c\mu$-rule, there is an optimal policy that serves class 1 in state x, and one that serves 2. Thus, we can apply the same argument using equations (3.8) and (3.9) as in Section 3.3.1 to get that choosing arbitrarily among the two classes is optimal. In fact, this generalizes for any finite number of classes.

This enables us to extend the results of Section 3.3.1 by constructing optimal policies for $LR_L(\gamma)$ at γ^* that satisfy (3.14). Extending the optimality of the class of randomized-threshold policies as defined in Section 3.3.1 in full generality is difficult. We conjecture that this class of policies does not need to contain an optimal policy for $B_L(V)$. The problem lies in the need to meet multiple constraints: tweaking threshold parameters to meet a particular constraint at equality may change the cost of another class, causing it to no longer be binding. However, we are able to propose a particular class of randomized-threshold policies that we show does contain an optimal policy for $B_L(V)$, if feasible. The intuition is to create a class of policies for which the threshold parameters can be tuned sequentially: each constraint has a set of parameters that must be tweaked to meet the constraint at equality, without affecting previously set constraints.

We define a generalization of the class of threshold policies as follows. Let $n = (n_{L+1}, ..., n_2)^T \in \mathbb{Z}_+^L$ and $p = (p_{L+1}, ..., p_2)^T \in [0,1]^L$. Define the policy $\sigma_{n,p} \in \Pi^S$

that serves according to Algorithm 1. Define the class of *randomized-threshold*

Algorithm 1 chooseAction

procedure CHOOSEACTION(x, n, p)
  served $\leftarrow$ No
  $U \leftarrow \{1, ..., L+1\}$
  for $k = L+1, L, ..., 2$ **do**
    $U \leftarrow U \setminus \{k\}$
    Cond1 $\leftarrow x_k > n_k$
    Cond2 $\leftarrow x_k = n_k$ and $U(0,1) \leq p_k$
    Cond3 $\leftarrow \sum_{i \in U} x_i = 0$
    if Cond1 or Cond2 or Cond3 **then**
      Served $\leftarrow$ Yes
      Serve class k
  if served = No **then**
    Serve classes $L+1, ..., K$ according to $c\mu$-rule.

policies Π^{RT} to be the set of all policies $\sigma_{n,p}$ across all vectors n and p. Note that every $\sigma \in \Pi^{RT}$ is non-idling: $\Pi^{RT} \subseteq \Pi([K])$. Even further, we observe that these policies are all non-idling with respect to the constrained classes and the highest weighted unconstrained class: $\Pi^{RT} \subseteq \Pi([L+1])$. The most important property of this class of policies is summarized in Lemma 3.3.12. This property allows us to find a binding policy sequentially.

Lemma 3.3.12 *Let $\sigma_{n,p}$ denote the policy choosing which class to serve in state x according to Algorithm 1 for $n \in \mathbb{Z}_+^L, p \in [0,1]^L$. For any $k = L+1, L, ..., 2$, the costs $C_{L+1}(\sigma_{n,p}), ..., C_{k+1}(\sigma_{n,p})$ do not depend on the parameters $n_k, p_k, ..., n_2, p_2$.*

Proof. We use a sample path argument. Pick any $k \in \{L+1, ..., 2\}$. Define $\tilde{n}$ and $\tilde{p}$ so that

$$\tilde{n}_{L+1} = n_{L+1}, \tilde{p}_{L+1} = p_{L+1}, ..., \tilde{n}_{k+1} = n_{k+1}, \tilde{p}_{k+1} = p_{k+1}, \tag{3.15}$$

and so that there exists some $i \in \{k, ..., 2\}$ so that either $\tilde{n}_i \neq n_i$ or $\tilde{p}_i \neq p_i$. Define the policies $\sigma_{n,p}$ and $\sigma_{\tilde{n},\tilde{p}}$ accordingly, and for notational convenience refer

to them by σ and $\tilde{\sigma}$, respectively. Define two Markov chains induced by σ and $\tilde{\sigma}$ on the same probability space so that they see the same arrival times, service times, and initial state. Call these Markov chains $X = \{X(t) : t \geq 0\}$ and $\tilde{X} = \{\tilde{X}(t) : t \geq 0\}$, respectively. Suppose that, at some point in time, the two processes are in states that differ in at least one of the components $L+1, ..., k+1$. Then there must be some time τ_0 at which, for the first time, one of $\sigma, \tilde{\sigma}$ serves some class $j \in \{L+1, ..., k+1\}$ while the other does not. Without loss of generality, assume that it is σ that serves j and $\tilde{\sigma}$ does not. By definition of τ_0,

$$X_i(\tau_0) = \tilde{X}_i(\tau_0), \quad i = L+1, ..., k+1. \tag{3.16}$$

Since the two processes are in the same state at time $t = \tau_0$ and have the same threshold conditions for classes $L+1, ...k+1$ by (3.15), class j can only be served in process X if the classes in $U = \{j-1, ..., 1\}$ are all empty. That is,

$$\sum_{i \in U} X_i(\tau_0) = 0. \tag{3.17}$$

Additionally, since j is not served in process $\tilde{X}$, we have

$$\sum_{i \in U} \tilde{X}_i(\tau_0) > 0. \tag{3.18}$$

By definition of τ_0 and the fact that $\sigma, \tilde{\sigma}$ are non-idling with respect to classes in $\{L+1, ..., 1\}$, both processes have spend the same amount of work up to time τ_0 on the classes $j-1, ..., 1$. However, by (3.17) and (3.18),

$$0 = \sum_{i=k}^{L+1} X_i(\tau_0) < \sum_{i=k}^{L+1} \tilde{X}_i(\tau_0), \tag{3.19}$$

which is a contradiction, since it implies that process X has spent more time working on classes $j-1, ..., 1$ than process $\tilde{X}$ has. ∎

With Lemma 3.3.12, we can propose the following procedure to find a binding, and hence optimal, randomized-threshold policy. In order for this procedure to

be well-defined, we introduce the following (artificial) target cost for class $L+1$, C^*_{L+1}, satisfying

$$\frac{C^*_{L+1}}{\mu_{L+1}} = w(\{1, ..., L+1\}) - \sum_{i=1}^{L} \frac{V_i}{\mu_i},$$

representing the class $L+1$ cost achieved by any policy that meets all of the constraints at equality and is also non-idling with respect to classes $1, ..., L{+}1$. Note that, by construction, every randomized-threshold policy fits this description. As a consequence, observe that any randomized-threshold policy meeting the constraints for classes $2, ..., L$ at equality while also meeting the artificial class $L+1$ target at equality also satisfies the class 1 constraint at equality. We use this fact to validate the correctness of Algorithm 2 in Theorem 3.3.15. Finding the parameters

Algorithm 2 findBindingPolicy

procedure FINDBINDINGPOLICY(V)
  $n \leftarrow 0 \in \mathbb{Z}^L_+$
  $p \leftarrow 0 \in [0,1]^L$
  Find n_{L+1}, p_{L+1} so that $C_{L+1}(\sigma_{n,p}) = C^*_{L+1}$
  for $k = L, ..., 2$ **do**
    Find n_k, p_k so that $C_k(\sigma_{n,p}) = V_k$
  return $\sigma_{n,p}$

n_k and p_k as specified in Algorithm 2 can be done similarly as described in Section 3.3.1. Before proving Theorem 3.3.15, we need to define the concept of conditional prioritization, defined below, and its consequences, stated in Lemma 3.3.14.

Definition 3.3.13 *Given $U \subseteq S$ and $k \in S \setminus U$, a policy $\sigma \in \Pi^S$ is said to* ***conditionally prioritize*** *U over k if it prioritizes U over k conditioned on serving a class in $U \cup \{k\}$.*

Lemma 3.3.14 *For $U \subseteq S$ and $k \in S \setminus U$, let $\sigma \in \Pi^S$ be any policy prioritizing*

U over k, and let $\tilde{\sigma}$ conditionally prioritize U over k. Then

$$C_k(\sigma) \leq C_k(\tilde{\sigma}).$$

Proof. It suffices to show that $\bar{W}_k(\sigma) \leq \bar{W}_k(\tilde{\sigma})$. We prove this via a sample-path argument. Consider the two workload processes, $W = \{W(t) : t \geq 0\}$ and $\tilde{W} = \{\tilde{W}(t) : t \geq 0\}$, induced by the respective policies and defined on the same probability space so that arrival times, service requirements, and initial states coincide. Since σ is non-idling with respect to classes in U, we have on every sample path:

$$\sum_{i \in U} W_i(t) \leq \sum_{i \in U} \tilde{W}_i(t), \quad t \geq 0.$$

Thus, if $\tilde{\sigma}$ serves class k at time t, then $\sum_{i \in U} \tilde{W}_i(t) = 0$, and so $\sum_{i \in U} W_i(t) = 0$. Hence, the only way that $\tilde{W}$ sees a class k customer service at time t while W does not see one is if $W_k(t) = 0$. Thus, every class k customer finishes service sooner in process W than in process $\tilde{W}$. This implies

$$W_k(t) \leq \tilde{W}_k(t), \quad t \geq 0,$$

completing the proof. ∎

Lemma 3.3.14 is instrumental in showing that each iteration results in a new constraint being met at equality. In particular, it allows us to show that, at the start of the k^{th} iteration (for $k = L, ..., 2$), $C_k(\sigma_{n,p}) < V_k$, and as we increase $n_k \to \infty$, $C_k(\sigma_{n,p})$ approaches a value greater than V_k.

Theorem 3.3.15 *Algorithm 2 returns an optimal policy for $B_L(V)$.*

Proof. Recall that for any value of n, p, $\sigma_{n,p}$ is non-idling with respect to classes $1, ..., L+1$. It suffices to show that, for each iteration $k = L+1, ..., 2$ of Algorithm

2, we can find n_k, p_k so that $C_k(\sigma_{n,p})$ meets its target. Note that if we are able to do so, then Lemma 3.3.12 implies that a binding (and hence optimal) policy is returned. Also note that by Theorem 3.2.4 and the intermediate value theorem, it suffices to show that, for each iteration $k = L+1, L, ..., 2$, the class k cost at the start of the iteration is below its target, and converges to a value above its target as $n_k \to \infty$. Take $k = L+1$ as our base case. Note that, at the start of this iteration, $n = p = 0 \in \mathbb{R}^L$: all parameters are set to zero. Thus, $\sigma_{n,p}$ is a policy that prioritizes class $L+1$ over all other classes, and thus

$$\frac{C_{L+1}(\sigma_{n,p})}{\mu_{L+1}} = w(\{L+1\}). \tag{3.20}$$

Furthermore, note that, as $n_{L+1} \to \infty$ as all other parameters stay constant, $\sigma_{n,p}$ converges pointwise to a policy that prioritizes classes $1, ..., L$ over class $L+1$. Since this limiting policy is also non-idling with respect to classes $1, ..., L+1$, and since costs are continuous via Theorem 3.2.4,

$$\lim_{n_{L+1}\to\infty} \frac{C_{L+1}(\sigma_{n,p})}{\mu_{L+1}} = w(\{1, ..., L+1\}) - w(\{1, ..., L\}), \tag{3.21}$$

where we take the limit on the LHS to mean that $n_{L+1} \to \infty$ while all other components of n and all the components of p remain constant. Note that Assumption (RHS) implies

$$w(\{1, ..., L\}) < \sum_{i=1}^{L} \frac{V_i}{\mu_i} < w(\{1, ..., L+1\}) - w(\{L+1\}).$$

Combining this with the definition of C^*_{L+1} and (3.20), we have

$$\begin{aligned}
\frac{C^*_{L+1}}{\mu_{L+1}} &= w(\{1, ..., L+1\}) - \sum_{i=1}^{L} \frac{V_i}{\mu_i} \\
&> w(\{1, ..., L+1\}) - (w(\{1, ..., L+1\}) - w(\{L+1\})) \\
&= w(\{L+1\}) \\
&= \frac{C_{L+1}(\sigma_{n,p})}{\mu_{L+1}}.
\end{aligned}$$

Similarly, leveraging (3.21) gives us

$$\begin{aligned}\frac{C^*_{L+1}}{\mu_{L+1}} &= w(\{1,...,L+1\}) - \sum_{i=1}^{L} \frac{V_i}{\mu_i} \\ &< w(\{1,...,L+1\}) - w(\{1,...,L\} \\ &= \lim_{n_{L+1}\to\infty} \frac{C_{L+1}(\sigma_{n,p})}{\mu_{L+1}},\end{aligned}$$

completing our base case. Now suppose that we are at the beginning of iteration $k \in \{L,...,2\}$, and have successfully completed iterations $L+1,...,k+1$, resulting in n,p such that

$$\begin{aligned}&C_{L+1}(\sigma_{n,p}) = C^*_{L+1} \\ &C_i(\sigma_{n,p}) = V_i, \quad i = L,...,k+1 \\ &n_i = p_i = 0, \quad i = k,k-1,...,2.\end{aligned}$$

Using our inductive hypothesis, along with the fact that $\sigma_{n,p}$ is non-idling with respect to classes $1,...,L+1$, we have

$$\begin{aligned}\sum_{i=1}^{k} \frac{C_i(\sigma_{n,p})}{\mu_i} &= w(\{1,...,L+1\}) - \sum_{i=k+1}^{L+1} \frac{C_i(\sigma_{n,p})}{\mu_i} \\ &= w(\{1,...,L+1\}) - \sum_{i=k+1}^{L} \frac{V_i}{\mu_i} - \frac{C^*_{L+1}}{\mu_{L+1}} \\ &= \sum_{i=1}^{k} \frac{V_i}{\mu_i}.\end{aligned}$$

Hence, either $C_i(\sigma_{n,p}) = V_i$ for all $i = 1,...,k$, or there exists $j \in \{1,...,k\}$ such that $C_j(\sigma_{n,p}) < V_j$. In the former case, we have found a binding policy and are done. In the latter case, we proceed to show that $j = k$. To see this, note that each class $i = 1,...,k-1$ is conditionally prioritized after classes $\{k,...,i+1\}$. Hence, by Assumption (RHS),

$$\frac{C_i(\sigma_{n,p})}{\mu_i} \geq w(\{k,k-1,...,i\}) - w(\{k,k-1,...,i+1\}) > \frac{V_i}{\mu_i}.$$

Thus, we must have that $C_k(\sigma_{n,p}) < V_k$ by elimination. By noting that, as $n_k \to \infty$, $\sigma_{n,p}$ converges pointwise to a policy that conditionally prioritizes classes $k-1, ..., 1$ over class k, a similar argument yields that

$$\lim_{n_k \to \infty} \frac{C_k(\sigma_{n,p})}{\mu_k} \geq w(\{k, k-1, ..., 1\}) - w(\{k-1, ..., 1\}) > \frac{V_k}{\mu_k},$$

completing the proof. ∎

3.4 Conclusion

We examined a K-class, parallel queueing system staffed by a single server and considered a dynamic server allocation problem in the presence of higher-priority customer classes with target service level requirements. The problem was formulated as a constrained Markov decision process (CMDP), and optimal policies exhibiting structural properties lending themselves to easy application were sought. In the case with only one high-priority class, two broad classes of such policies were found: one-randomized $c\mu$-rule policies and randomized-threshold policies. In addition, for the one-constraint case, we showed how the ordering of policies in the optimal one-randomized $c\mu$-rule policy changed depending on the target quality of service level. In the general case with L high-priority classes, this discussion was skipped in favor of a tighter assumption on the quality of service targets to simplify our analysis. The L-constraint counterparts to the randomized-threshold class of policies was shown to be optimal. The randomized-threshold class is easier to implement algorithmically, and is hence an appealing choice when applied to settings in which the practitioner does not have a good sense of how to weigh high-priority customers.

CHAPTER 4

CMDP APPROACHES FOR PERSONALIZED MEDICINE

4.1 Introduction

We consider the problem of dynamically prescribing treatments to a patient over time. The patient's state of health is described by a state space, $\mathbb{X}$. In practice, $\mathbb{X}$ may consist of a set of vectors whose components correspond to various measures of health (e.g. blood pressure, cholesterol level, etc.). For our problem to be tractable, we need to assume an ordering of these states (vectors). In doing so, we lose no generality in also assuming that states can be represented by natural (scalar) numbers. For simplicity, we model $\mathbb{X} := \{1, 2, ..., n\}$ for some positive integer $n \geq 2$. Here we take state 1 to be the best state of health and state n to be the worst. The practitioner (e.g. a physician) has a finite number, k, of possible treatments to prescribe to the patient during each appointment. Let $\mathbb{A} := \{1, ..., k\}$ denote the set of treatments. Associated with each treatment $a \in \mathbb{A}$ is a treatment cost $c(a)$, and two rates $\lambda(a)$ and $\mu(a)$. The treatment cost function $c : \mathbb{A} \mapsto \mathbb{R}_+$ is assumed to be independent of the patient's state of health, and represents the total cost of a particular treatment incurred from the time of prescription until the next appointment. The rate functions $\lambda : \mathbb{A} \mapsto \mathbb{R}_+$ and $\mu : \mathbb{A} \mapsto \mathbb{R}_+$ are also assumed to be independent of the patient's health and represent treatment response. More specifically, $\lambda(a)$ can be thought of as the rate at which the patient's health deteriorates to the next highest state when undergoing treatment a. Similarly, $\mu(a)$ is interpreted as the rate at which the patient's health improves to the next lowest state while using treatment a. It should be noted that we are implicitly assuming that the patient's health can only jump to neighboring states

between appointments. This assumption can be rationalized if the disease being treated develops slowly or if the patient's state of health is modelled using a sufficiently coarse state space. In more rapidly progressing diseases, this assumption can still be reasonable if the frequency of treatment is high enough so that the state of health is unlikely to change between treatment sessions. We assume that the treatments are ordered by (increasing) cost and effectiveness: treatments are ordered from least expensive to most expensive, and more expensive treatments should also be more effective. That is, we assume

$$c(1) \leq c(2) \leq ... \leq c(k)$$

$$\lambda(1) \geq \lambda(2) \geq ... \geq \lambda(k)$$

$$\mu(1) \leq \mu(2) \leq ... \leq \mu(k).$$

Given the dynamics of this model, there is an immediate trade-off between the patient's overall state of health and cost of treatment: treatment policies which are more expensive tend to put the patient in better overall health than a less expensive policy. To put this into more concrete terms, we consider an infinite time-horizon proxy for this problem where we consider both the average cost of a treatment policy as well as the resulting long-run fraction of time the patient spends in each state of health. For a given state ℓ and a fraction of time V, we are interested in prescribing the cheapest treatment policy possible while ensuring that the patient spends at most a fraction V of the time in a state of health ℓ or worse. Both metrics here are defined with respect to the infinite-horizon, expected long-run average cost criterion. In this setting, it suffices to consider treatment policies that are **stationary**: the treatment that is prescribed at any point in time depends only on the patient's current state of health [39]. We let Π^S denote the set of stationary policies. We can now formally define our problem. Given a

stationary policy $\sigma \in \Pi^S$, let

$$C_1(\sigma) := \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T \mathbb{1}\{X^\sigma(t) \geq \ell\}dt\right]$$
$$C_2(\sigma) := \limsup_{T\to\infty} \mathbb{E}\left[\frac{1}{T}\int_0^T c\left(\sigma\left(X^\sigma(t)\right)\right)dt\right],$$

where $\{X^\sigma(t) : t \geq 0\}$ is the Markov process induced by the policy σ and $\sigma(x) \in \mathbb{A}$ is the action chosen by σ when in state $x \in \mathbb{X}$. We can write our problem, denoted by B(V,ℓ), as

$$\inf_{\sigma\in\Pi^S}\{C_2(\sigma) : C_1(\sigma) \leq V\}. \tag{B(V,ℓ)}$$

4.2 Assumptions and Preliminaries

In this section, we introduce an assumption and review some preliminary results. For notational convenience, let P_a denote the policy that always chooses action $a \in \mathbb{A}$. We operate under the following assumption:

$$C_1(P_k) < V < C_1(P_1), \tag{4.1}$$

Observe that every stationary policy induces an irreducible, finite-state (and hence positive-recurrent) Markov chain. As a result, every stationary policy $\sigma \in \Pi^S$ yields a unique occupation measure $\phi^\sigma : \mathbb{X} \times \mathbb{A} \mapsto [0, 1]$ representing the long-run average fraction of time the induced process spends in each state-action pair [3]. This allows us, given $\sigma \in \Pi^S$, to represent $C_1(\sigma)$ and $C_2(\sigma)$ in an alternative form:

$$C_1(\sigma) = \sum_{i=\ell}^{n}\sum_{a=1}^{k}\phi^\sigma(i, a)$$
$$C_2(\sigma) = \sum_{i=1}^{n}\sum_{a=1}^{k}c(a)\phi^\sigma(i, a).$$

In addition, this allows us to write B(V,ℓ) as the following linear program

$$\begin{aligned}
&\min \sum_{i,a} c(a)\phi(i,a) \\
&\text{s.t.} \quad \sum_{a} \lambda(a)\phi(i,a) - \sum_{a} \mu(a)\phi(i+1,a) = 0, \qquad i = 1, ..., n-1 \\
&\qquad \sum_{i,a} \phi(i,a) = 1 \\
&\qquad \sum_{i \geq \ell, a} \phi(i,a) \leq V \\
&\qquad \phi(i,a) \geq 0 \qquad i = 1, ..., \ell - 1, \quad a = 1, ..., k.
\end{aligned}$$

We now state a result from constrained Markov decision process (CMDP) theory that we later use to show that, under Assumption (4.1), it suffices to seek a policy that satisfies the constraint at equality. This fact helps us decompose the problem in later steps. The result is adapted from Altman [3].

Theorem 4.2.1 (adapted from Thereom 12.7 in Altman)

1. *The optimal value, $C_{V,\ell}$, of the problem B(V,ℓ) can be computed,*

$$C_{V,\ell} = \inf_{\sigma} \sup_{\gamma \geq 0} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\}.$$

2. *A policy σ^* is optimal for B(V,ℓ) if and only if*

$$C_{V,\ell} = \sup_{\gamma \geq 0} \{C_2(\sigma^*) + \gamma(C_1(\sigma^*) - V)\}$$

3. *Suppose Π^D is the set of all stationary, deterministic policies. For any class of policies Π such that $\Pi^D \subseteq \Pi$,*

$$C_{V,\ell} = \sup_{\gamma \geq 0} \min_{\sigma \in \Pi} \{C_2(\sigma) + \gamma(C_1(\sigma) - V)\},$$

where we can take $\Pi = \Pi^S$, the set of all stationary policies.

Statement 3 of Theorem 4.2.1 deserves some further comment. Recall that $C_{V,\ell}$ is the value of the problem B(V,ℓ). A policy that achieves the minimum at the supremum over γ may be an optimal policy. On the other hand, it is possible that the policy achieving the value $C_{V,\ell}$ is not feasible; Statement 3 allows for a method to compute $C_{V,\ell}$.

4.3 Decomposition Approach Using Lagrangian

In this section we show through a series of steps that solving B(V,ℓ) can be decomposed into finding optimal policies for two unconstrained MDPs and then stitching these solutions together. The first step involves showing that we can restrict our search for an optimal stationary policy for B(V,ℓ) to the set of **binding** stationary policies: $\Pi^B(V,\ell) := \{\sigma \in \Pi^S : C_1(\sigma) = V\}$. We introduce the"traditional" Lagrangian dual problem, LD(V,ℓ):

$$\sup_{\gamma \geq 0}\{\min_{\sigma \in \Pi^S}\{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V\}. \qquad \text{(LD(V,}\ell\text{))}$$

Note that LD(V,ℓ) is a simplified version of the expression appearing in Statement 3 of Theorem 4.2.1. The simplification involves noting that $-\gamma V$ does not depend on $\sigma \in \Pi^S$. Let $g(\gamma) := \min_{\sigma \in \Pi^S}\{\gamma C_1(\sigma) + C_2(\sigma)\} - \gamma V$ be the function within the supremum in LD(V,ℓ). Additionally, for $\gamma \in \mathbb{R}$, let

$$O_\gamma := \arg\min_{\sigma \in \Pi^S}\{\gamma C_1(\sigma) + C_2(\sigma)\}.$$

The following lemma summarizes some useful facts about $g(\gamma)$.

Lemma 4.3.1 *The function $g(\gamma)$ satisfies*

1. *$g(\gamma)$ is concave in γ.*

2. *For any* $\sigma_\gamma \in O_\gamma$, $V - C_1(\sigma_\gamma) \in \partial(-g)(\gamma)$, *where* ∂f *is the subdifferential (set of all subgradients) of the function* f.

Proof. Note that since sums of concave functions are concave, and the minimum of concave functions is concave the first result holds. To show the second result we need to show that for any $\gamma \in \mathbb{R}, \sigma_\gamma \in O_\gamma$,

$$-g(\gamma_0) \geq -g(\gamma) + (V - C_1(\sigma_\gamma))(\gamma_0 - \gamma) \qquad \forall \gamma_0 \in \mathbb{R}.$$

Fix $\gamma \in \mathbb{R}$. We have, for any $\gamma_0 \in \mathbb{R}$,

$$\begin{aligned} g(\gamma_0) - g(\gamma) &= \min_{\sigma \in \Pi^S} \{\gamma_0 C_1(\sigma) + C_2(\sigma)\} - \min_{\sigma \in \Pi^S} \{\gamma C_1(\sigma) + C_2(\sigma)\} - V(\gamma_0 - \gamma) \\ &= \min_{\sigma \in \Pi^S} \{\gamma_0 C_1(\sigma) + C_2(\sigma)\} - \gamma C_1(\sigma_\gamma) - C_2(\sigma_\gamma) - V(\gamma_0 - \gamma) \\ &\leq \gamma_0 C_1(\sigma_\gamma) + C_2(\sigma_\gamma) - \gamma C_1(\sigma_\gamma) - C_2(\sigma_\gamma) - V(\gamma_0 - \gamma) \\ &= -(V - C_1(\sigma_\gamma))(\gamma_0 - \gamma). \end{aligned}$$

Hence

$$-g(\gamma_0) \geq -g(\gamma) + (V - C_1(\sigma_\gamma))(\gamma_0 - \gamma),$$

as desired. ∎

The following proposition states that there always exists a binding constrained-optimal policy for B(V,ℓ).

Proposition 4.3.2 *For any* $V \in (C_1(P_k), C_1(P_1)), \ell \in \mathbb{X}$,

$$\min_{\sigma \in \Pi^B(V,\ell)} C_2(\sigma) = C_{V,\ell}.$$

Proof. Consider the maximization in LD(V,ℓ). Since $g(\gamma)$ is concave, $0 \in \partial(-g)(\gamma^*)$, unless γ^* attaining the supremum is on the boundary (i.e. $\gamma^* = 0$). By Statement 2 of Lemma 4.3.1, this means that there exists $\sigma^* \in O_{\gamma^*}$ satisfying $C_1(\sigma^*) - V = 0$, hence σ^* is binding. Thus,

$$\begin{aligned} C_{V,\ell} &= \sup_{\gamma \geq 0} g(\gamma) = g(\gamma^*) \\ &= \gamma^* C_1(\sigma^*) + C_2(\sigma^*) - \gamma^* V = C_2(\sigma^*) \\ &= \sup_{\gamma \geq 0} \{C_2(\sigma^*) + \gamma(C_1(\sigma^*) - V)\}, \end{aligned}$$

where the the last two equalities follow since $C_1(\sigma^*) - V = 0$. By Statement 2 of Theorem 4.2.1, σ^* is optimal for B(V,ℓ). Thus, it suffices to show that we cannot have $\gamma^* = 0$. We first evaluate $g(0)$. Noting that $O_\gamma = \{P_1\}$ for $\gamma = 0$ (in the unconstrained problem, always choose the cheapest action), we have that

$$g(0) = \min_{\sigma \in \Pi^S} \{C_2(\sigma)\} = \min_{a \in \mathbb{A}} c(a) = c(1).$$

So we must find $\bar{\gamma} > 0$ with $g(\bar{\gamma}) > c(1)$. Let Π^D denote the set of stationary deterministic policies. It is well-known that this is a dominating class of policies for unconstrained MDPs [39]: it suffices to search for an optimal stationary deterministic policy. Since there are only a finite number of such policies, there exists a minimum gap between the expected long-run average cost of P_1 and that of other stationary deterministic policies :

$$\delta := \min_{\sigma \in \Pi^D \setminus \{P_1\}} \{C_2(\sigma) - C_2(P_1)\} > 0.$$

Fix arbitrary $\bar{\gamma} \in \left(0, \frac{\delta}{V - C_1(P_k)}\right)$. Note that this interval is well-defined by Assumption 4.1. Let $\sigma_{\bar{\gamma}} \in \Pi^D \cap O_{\bar{\gamma}}$. Then

$$g(\bar{\gamma}) = \bar{\gamma}(C_1(\sigma_{\bar{\gamma}}) - V) + C_2(\sigma_{\bar{\gamma}}),$$

and so

$$\begin{aligned} g(\bar{\gamma}) - g(0) &= \bar{\gamma}(C_1(\sigma_{\bar{\gamma}}) - V) + (C_2(\sigma_{\bar{\gamma}}) - C_2(P_1)) \\ &\geq \bar{\gamma}(C_1(P_k) - V) + \delta \\ &> \frac{\delta}{C_1(P_k) - V}(C_1(P_k) - V) + \delta \\ &= 0. \end{aligned}$$

If $\sigma_{\bar{\gamma}} = P_1$, then

$$g(\bar{\gamma}) - g(0) = \bar{\gamma}(C_1(P_1) - V) + (C_2(P_1) - C_2(P_1)) > 0.$$

Hence the supremum of $g(\gamma)$ cannot be attained at 0, completing the proof. ∎

Using Proposition 4.3.2, we can now simplify the LP formulation of B(V,ℓ):

$$\min \sum_{i=1}^{\ell-1}\sum_{a=1}^{k} c(a)\phi(i,a) + \sum_{i=\ell}^{n}\sum_{a=1}^{k} c(a)\phi(i,a) \tag{4.2}$$

$$\text{s.t.} \quad \sum_{a=1}^{k} \lambda(a)\phi(i,a) - \sum_{a=1}^{k} \mu(a)\phi(i+1,a) = 0, \qquad i = 1,...,\ell-2 \tag{4.3}$$

$$\sum_{i=1}^{\ell-1}\sum_{a=1}^{k} \phi(i,a) = 1 - V \tag{4.4}$$

$$\phi(i,a) \geq 0 \qquad i = 1,...,\ell-1, \quad a = 1,...,k \tag{4.5}$$

$$\sum_{a=1}^{k} \lambda(a)\phi(i,a) - \sum_{a=1}^{k} \mu(a)\phi(i+1,a) = 0, \qquad i = \ell,...,n-1 \tag{4.6}$$

$$\sum_{i=\ell}^{n}\sum_{a=1}^{k} \phi(i,a) = V \tag{4.7}$$

$$\phi(i,a) \geq 0 \qquad i = \ell,...,n, \quad a = 1,...,k \tag{4.8}$$

$$\sum_{a=1}^{k} \lambda(a)\phi(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi(\ell,a) = 0. \tag{4.9}$$

Note that the problem is nearly separable: only constraint (4.9) prevents us from optimizing over $\{\phi(i,a) : i = 1,...,\ell-1, \quad a = 1,...,k\}$ and $\{\phi(i,a) : i =$

$\ell, ..., n, \quad a = 1, ..., k\}$ separately. Additionally, the constraints involving each set of variables ((4.3)-(4.5) and (4.6)-(4.8), respectively) resemble scaled versions of those found in the LP formulations for unconstrained MDPs. Note that, given a solution ϕ to the LP, we extract a stationary policy by choosing action a in state i with probability

$$\frac{\phi(i,a)}{\sum_{d=1}^{k}\phi(i,d).}$$

We formally define these subproblems

$$\min_{x\in Q_1}\left\{\sum_{i=1}^{\ell-1}\sum_{a=1}^{k}c(a)x(i,a)+\gamma\sum_{a=1}^{k}\lambda(a)x(\ell-1,a)\right\}, \qquad (SP_1(\gamma))$$

$$\min_{y\in Q_2}\left\{\sum_{a=1}^{k}(-\gamma\mu(a))y(1,a)+\gamma\sum_{i=1}^{n-\ell+1}\sum_{a=1}^{k}c(a)y(i,a)\right\}. \qquad (SP_2(\gamma))$$

Since the scaling involved in $SP_1(\gamma)$ and $SP_2(\gamma)$ do not affect the optimal policy, solving the scaled and unscaled versions are equivalent. This observation motivates us to take the following "alternate" Lagrangian, L(V,ℓ):

$$\sup_{\gamma\in\mathbb{R}}\{g_1(\gamma)+g_2(\gamma)\}, \qquad (\mathrm{L(V,}\ell\mathrm{)})$$

where, letting x and y take the place of $\{\phi(i,a) : i = 1, ..., \ell-1, \quad a = 1, ..., k\}$ and $\{\phi(i,a) : i = \ell, ..., n, \quad a = 1, ..., k\}$, respectively, we define

$$g_1(\gamma) := \min_{x\in Q_1}\left\{\sum_{i=1}^{\ell-1}\sum_{a=1}^{k}c(a)x(i,a)+\sum_{a=1}^{k}\gamma\lambda(a)x(\ell-1,a)\right\}$$
$$g_2(\gamma) := \min_{y\in Q_2}\left\{\sum_{a=1}^{k}(-\gamma\mu(a))y(1,a)+\sum_{i=1}^{n-\ell+1}\sum_{a=1}^{k}c(a)y(i,a)\right\}$$

and the sets Q_1 and Q_2 are defined by constraints (4.3)-(4.5) and (4.6)-(4.8), respectively. Note that for every fixed $\gamma \in \mathbb{R}$, the minimizations in $g_1(\gamma)$ and $g_2(\gamma)$ are unconstrained MDPs with modified cost functions that depend on γ. We refer to these problems as subproblems $SP_1(\gamma)$ and $SP_2(\gamma)$, respectively. The following

properties of these subproblems and their corresponding functions help establish the equivalence of L(V,ℓ) and B(V,ℓ).

Lemma 4.3.3 *The functions $g_1(\gamma)$ and $g_2(\gamma)$ satisfy the following properties.*

1. *$g_1(\gamma)$ and $g_2(\gamma)$ are concave in γ on $\mathbb{R}$.*
2. *Let $\Omega_1(\gamma), \Omega_2(\gamma)$ denote the set of occupation measures corresponding to the optimal stationary policies for $SP_1(\gamma)$ and $SP_2(\gamma)$, respectively. Then, for any $\gamma \in \mathbb{R}$,*

$$\left\{ \sum_{a=1}^{k} \lambda(a) x_1^\gamma(\ell - 1, a) : x_1^\gamma \in \Omega_1(\gamma) \right\} = \partial(-g_1)(\gamma),$$
$$\left\{ -\sum_{a=1}^{k} \mu(a) x_2^\gamma(1, a) : x_2^\gamma \in \Omega_2(\gamma) \right\} = \partial(-g_2)(\gamma),$$

and hence, letting $g(\gamma) = (-g_1)(\gamma) + (-g_2)(\gamma)$,

$$\left\{ \sum_{a=1}^{k} \lambda(a) x_1^\gamma(\ell - 1, a) - \sum_{a=1}^{k} \mu(a) x_2^\gamma(1, a) : x_1^\gamma \in \Omega_1(\gamma), x_2^\gamma \in \Omega_2(\gamma) \right\} = \partial(g)(\gamma).$$

Proof. The first statement and the fact that

$$\left\{ \sum_{a=1}^{k} \lambda(a) x_1^\gamma(\ell - 1, a) : x_1^\gamma \in \Omega_1(\gamma) \right\} \subseteq \partial(-g_1)(\gamma),$$
$$\left\{ -\sum_{a=1}^{k} \mu(a) x_2^\gamma(1, a) : x_2^\gamma \in \Omega_2(\gamma) \right\} \subseteq \partial(-g_2)(\gamma).$$

follow using a similar argument as found in the proof of 4.3.1. To complete the second statement of the proof, we need to show that

$$\left\{ \sum_{a=1}^{k} \lambda(a) x_1^\gamma(\ell - 1, a) : x_1^\gamma \in \Omega_1(\gamma) \right\} \supseteq \partial(-g_1)(\gamma),$$
$$\left\{ -\sum_{a=1}^{k} \mu(a) x_2^\gamma(1, a) : x_2^\gamma \in \Omega_2(\gamma) \right\} \supseteq \partial(-g_2)(\gamma).$$

We prove the claim for $g_1(\gamma)$, as the proof for $g_2(\gamma)$ follows a similar argument. Fix $\tilde{\gamma} \in \mathbb{R}$. We make use of the fact that there exists $\hat{\delta} > 0, \hat{\phi} \in \Pi^D$ such that

$$\hat{\phi} \in \Omega_1(\gamma), \quad \gamma \in [\tilde{\gamma}, \tilde{\gamma} + \hat{\delta}].$$

To see this, pick any positive sequence $(\delta_n)_{n=1}^{\infty}, \delta_n \downarrow 0$ and pick a sequence of stationary deterministic policies $(\sigma_n)_{n=1}^{\infty}$ with corresponding occupation measures $(\phi_n)_{n=1}^{\infty}$ such that

$$\phi_n \in \Omega_1(\tilde{\gamma} + \delta_n), \quad n \in \mathbb{N}.$$

Since the state space is finite, there are a finite number of stationary deterministic policies, and thus there must exist a policy $\hat{\sigma}$ that appears infinitely often in the sequence $(\sigma_n)_{n\in\mathbb{N}}$. That is, letting $\hat{\phi}$ denote the occupation measure corresponding to $\hat{\sigma}$, there exists a subsequence $(n_d)_{d\in\mathbb{N}}$ so that $\hat{\phi} \in \Omega_1(\tilde{\gamma} + \delta_{n_d})$ for every $d \in \mathbb{N}$. Thus

$$g_1(\tilde{\gamma}) = \sum_{i=1}^{\ell-1} \sum_{a=1}^{k} c(a)\hat{\phi}(i,a) + (\tilde{\gamma} + \delta_{n_d}) \sum_{a=1}^{k} \lambda(a)\hat{\phi}(\ell-1,a), \quad d \in \mathbb{N}.$$

By concavity of $g_1(\gamma)$, this means that, letting $\hat{\delta} = \delta_{n_1}$,

$$g_1(\gamma) = \sum_{i=1}^{\ell-1} \sum_{a=1}^{k} c(a)\hat{\phi}(i,a) + \gamma \sum_{a=1}^{k} \lambda(a)\hat{\phi}(\ell-1,a), \quad \gamma \in [\tilde{\gamma}, \tilde{\gamma} + \hat{\delta}].$$

Hence $g_1(\gamma)$ is linear on the interval $[\tilde{\gamma}, \tilde{\gamma} + \hat{\delta}]$, and thus any subgradient $\Delta \in \partial(-g_1)(\tilde{\gamma})$ must, by definition, satisfy $\Delta \geq \sum_{a=1}^{k} \lambda(a)\hat{\phi}(\ell-1,a)$. A similar argument yields the existence of $\bar{\delta} > 0, \bar{\phi}$ so that

$$g_1(\gamma) = \sum_{i=1}^{\ell-1} \sum_{a=1}^{k} c(a)\bar{\phi}(i,a) + \gamma \sum_{a=1}^{k} \lambda(a)\bar{\phi}(\ell-1,a), \quad \gamma \in [\tilde{\gamma} - \bar{\delta}, \tilde{\gamma}].$$

Thus, $g_1(\gamma)$ is linear on the interval $[\tilde{\gamma} - \bar{\delta}, \tilde{\gamma}]$, and hence any subgradient $\Delta \in \partial(-g_1)(\tilde{\gamma})$ must satisfy $\Delta \leq \sum_{a=1}^{k} \lambda(a)\bar{\phi}(\ell-1,a)$. Thus, for any $\Delta \in \partial(-g_1)(\tilde{\gamma})$,

there exists $\alpha \in [0,1]$ such that

$$\begin{aligned}\Delta &= \alpha \sum_{a=1}^{k} \lambda(a)\hat{\phi}(\ell-1,a) + (1-\alpha)\sum_{a=1}^{k}\lambda(a)\bar{\phi}(\ell-1,a) \\ &= \sum_{a=1}^{k}\lambda(a)(\alpha\hat{\phi}(\ell-1,a) + (1-\alpha)\bar{\phi}(\ell-1,a)).\end{aligned}$$

Note that since both $\hat{\sigma}, \bar{\sigma} \in O_{\tilde{\gamma}}$, we must have $\phi^{\alpha} = \alpha\hat{\phi} + (1-\alpha)\bar{\phi} \in \Omega_1(\tilde{\gamma})$. Thus $\Delta \in \left\{\sum_{a=1}^{k}\lambda(a)x_1^{\gamma}(\ell-1,a) : x_1^{\gamma} \in \Omega_1(\gamma)\right\}$, as desired. ∎

We aim to show the following relationship between L(V,ℓ) and B(V,ℓ), which tells us how to use optimal solutions to the unconstrained subproblems to construct an optimal policy for the constrained problem, B(V,ℓ).

Proposition 4.3.4 *The following hold.*

1. $C_{V,\ell} = \sup_{\gamma\in\mathbb{R}}\{g_1(\gamma) + g_2(\gamma)\}$
2. *If* $\gamma^* \in \mathbb{R}, x_1^* \in \Omega_1(\gamma^*), x_2^* \in \Omega_2(\gamma^*)$ *satisfies*

$$\sum_{a=1}^{k}\lambda(a)x_1^*(\ell-1,a) = \sum_{a=1}^{k}\mu(a)x_2^*(1,a), \tag{4.10}$$

then the concatenated occupation measure

$$\phi^*(i,a) = \begin{cases} x_1^*(i,a) & i = 1,...,\ell-1 \\ x_2^*(i-\ell+1,a) & i = \ell,...,n \end{cases}$$

corresponds to an optimal policy for B(V,ℓ).

The statement of Proposition 4.3.4 tells us that we can construct a constrained-optimal policy for B(V,ℓ) by stitching together two particular optimal solutions to each of the subproblems SP1(γ^*) and SP2(γ^*), where we do not know γ^* explicitly. To prove Statement 1 of Proposition 4.3.4, we leverage the structure of

the stationary distribution of a controlled birth-death chain with $\tilde{m}$ states under a stationary deterministic policy. For any $\sigma \in \Pi^D$, let σ_i denote the (single) action chosen by σ when in state i. Define

$$b_i^\sigma = \prod_{j=1}^{i-1} \frac{\lambda(\sigma_j)}{\mu(\sigma_{j+1})}, \quad i = 1, 2, ..., \tilde{m},$$

where we adopt the convention that $\prod_{j=1}^{0} \frac{\lambda(\sigma_j)}{\mu(\sigma_{j+1})} = 1$. Then π^σ, the stationary distribution induced by σ, is

$$\pi_i^\sigma = \frac{b_i^\sigma}{\sum_{j=1}^{\tilde{m}} b_j^\sigma}, \quad i = 1, 2, ..., \tilde{m}.$$

We use the form of this stationary distribution for $\tilde{m} = \ell - 1$ (for $SP_1(\gamma)$) and $\tilde{m} = n - \ell + 1$ (for $SP_2(\gamma)$) to prove the results in Lemmas 4.3.6 and 4.3.7, which are instrumental in proving the first statement of Proposition 4.3.4. To show these results, we first determine how switching between stationary deterministic policies changes the induced stationary distribution.

Proposition 4.3.5 *Consider a controlled birth-death process on state space* $\{1, ..., \tilde{m}\}$. *Fix a stationary deterministic policy* $\sigma \in \Pi^D$ *such that in some state,* s, $\sigma_s \in \{1, ..., k-1\}$. *Let* $\sigma' \in \Pi^D$ *be identical to* σ *except that* $\sigma'_s > \sigma_s$. *Then, the induced stationary distributions* π *and* π' *satisfy*

$$\pi'_1 > \pi_1, \pi_i < \pi'_i \text{ for } i = 2, ..., \tilde{m}, \qquad s = 1$$

$$\pi'_i > \pi_i \text{ for } i \leq x - 1, \pi'_i < \pi_i \text{ for } i \geq x, \qquad s = 2, ..., k-1$$

$$\pi'_i > \pi_i \text{ for } i = 1, ..., \tilde{m} - 1, \pi'_{\tilde{m}} < \pi_{\tilde{m}} \qquad s = \tilde{m}.$$

Proof. For simplicity, let $b_i = b_i^\sigma, b'_i = b_i^{\sigma'}$ for $i = 1, ..., \tilde{m}$. Note that

$$b'_i = b_i \quad i \leq \max(s-1, 1)$$

$$b'_i = pb_i \quad i \geq \max(s, 2),$$

where

$$p = \begin{cases} \frac{\lambda(\sigma'_s)}{\lambda(\sigma_s)} & s = 1 \\ \frac{\lambda(\sigma'_s)}{\lambda(\sigma_s)} \frac{\mu(\sigma_s)}{\mu(\sigma'_s)} & s = 2, ..., \tilde{m} \\ \frac{\mu(\sigma_s)}{\mu(\sigma'_s)} & s = \tilde{m}. \end{cases}$$

Since $\sigma'_s > \sigma_s$, observe that $p < 1$. For $s = 1$, we have

$$\pi'_i = \frac{b'_1}{\sum_{j=1}^{\tilde{m}} b'_j} = \begin{cases} \frac{b_1}{b_1 + p\sum_{j=2}^{\tilde{m}} b_j} > \frac{1}{\sum_{j=1}^{\tilde{m}} b_j} = \pi_i & i = 1 \\ \frac{pb_i}{b_1 + p\sum_{j=2}^{\tilde{m}} b_j} < \frac{pb_i}{p\sum_{j=1}^{\tilde{m}} b_j} = \pi_i & i = 2, ..., \tilde{m}. \end{cases}$$

Similarly, for $s \geq 2$,

$$\pi'_i = \frac{b'_i}{\sum_{j=1}^{\tilde{m}} b'_j} = \begin{cases} \frac{b_i}{\sum_{j=1}^{s-1} b_j + p\sum_{j=2}^{\tilde{m}} b_j} > \frac{b_i}{\sum_{j=1}^{\tilde{m}} b_j} = \pi_i & i = 1, ..., s-1 \\ \frac{pb_i}{b_1 + p\sum_{j=2}^{\tilde{m}} b_j} < \frac{pb_i}{p\sum_{j=1}^{\tilde{m}} b_j} = \pi_i & i = s, ..., \tilde{m}. \end{cases}$$

■

Lemma 4.3.6 *For every $\gamma \leq 0$, it is optimal to pick action 1 in every state for both $SP_1(\gamma)$ and $SP_2(\gamma)$.*

Proof. For $SP_1(\gamma)$, note that picking $\sigma_i = 1$ for all $i = 1, ..., \ell - 1$ minimizes $\sum_{i=1}^{\ell-1} c(\sigma_i)\pi_i^\sigma$. By Proposition 4.3.5, this choice of σ also maximizes $\pi_{\ell-1}$. Furthermore, $\lambda(\sigma_{\ell-1})$ is maximized by $\sigma^*_{\ell-1} = 1$. Hence, it is optimal to always pick action 1 in every state for $SP_1(\gamma)$ for $\gamma \leq 0$.

Similarly, for $SP_2(\gamma)$, picking $\sigma_i = 1$ for all $i = 1, ..., n - \ell + 1$ minimizes $\pi^\sigma_{\ell-1}$ (by Proposition 4.3.5) and $\sum_{i=1}^{n-\ell+1} c(\sigma_i)\pi_i^\sigma$, and picking $\sigma_1 = 1$ minimizes $\mu(\sigma_1)$. Thus, it is optimal to always pick action 1 in every state for $SP_2(\gamma)$ when $\gamma \leq 0$. ■

Lemma 4.3.7 *There exists $\tilde{\gamma} < \infty$ such that, for all $\gamma \geq \tilde{\gamma}$, for both subproblems $SP_1(\gamma)$ and $SP_2(\gamma)$, it is optimal to choose action k in every state.*

Proof. We show how to find such a $\tilde{\gamma}_1$ for $SP_1(\gamma)$. The same argument can be used for $SP_2(\gamma)$ to find another $\tilde{\gamma}_2$, yielding $\tilde{\gamma} = \max(\tilde{\gamma}_1, \tilde{\gamma}_2)$ as in the statement of the lemma. The objective value for $SP_1(\gamma)$ of the policy that chooses action k at every state can be written

$$c(k) + \gamma\lambda(k)\pi^*_{\ell-1},$$

where π^* is the induced stationary distribution. Since stationary deterministic policies are dominant for unconstrained MDPs, we need to show that for large enough γ,

$$c(k) + \gamma\lambda(k)\pi^*_{\ell-1} < \sum_{i=1}^{\ell-1} c(\sigma_i)\pi^\sigma_i + \gamma\lambda(\sigma_{\ell-1})\pi^\sigma_{\ell-1}$$

for every $\sigma \in \Pi^D$. By Proposition 4.3.5,

$$\pi^*_{\ell-1} = \min_{\sigma\in\Pi^D} \pi^\sigma_{\ell-1}.$$

Since Π^D is finite, there exists a minimum gap

$$\delta = \pi^*_{\ell-1} - \min\{\pi^\sigma_{\ell-1} : \pi^\sigma_{\ell-1} > \pi^*_{\ell-1}, \sigma \in \Pi^D\} < 0.$$

Thus, for any $\sigma \in \Pi^D$,

$$\left(c(k) + \gamma\lambda(k)\pi^*_{\ell-1}\right) - \left(\sum_{i=1}^{\ell-1} c(\sigma_i)\pi^\sigma_i + \gamma\lambda(\sigma_{\ell-1})\pi^\sigma_{\ell-1}\right) \leq (c(k) - c(1)) + \gamma\delta,$$

which is negative for all $\gamma \geq \frac{c(1)-c(k)}{\delta}$, completing the proof. ∎

Proof of Proposition 4.3.4.

Let P^1_a and P^2_a denote the priority policies that always pick action a $(a = 1, ..., k)$

for the problems $SP_1(\gamma)$ and $SP_2(\gamma)$, respectively. By Lemma 4.3.6, we know that P_1^1 and P_1^2 are both optimal for their respective subproblems when $\gamma \leq 0$. By Lemma 4.3.7, there exists $\tilde{\gamma} < \infty$ so that P_k^1 and P_k^2 are optimal for these subproblems. By Lemma 4.3.3,

$$\sum_{a=1}^{k} \lambda(a)\phi^{P_1^1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_1^2}(1,a) \in \partial(-g_1-g_2)(0)$$
$$\sum_{a=1}^{k} \lambda(a)\phi^{P_k^1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_k^2}(1,a) \in \partial(-g_1-g_2)(\tilde{\gamma}).$$

Note that $\sum_{a=1}^{k} \lambda(a)\phi^{P_1^1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_1^2}(1,a) > 0$. To see why, note that the policy P_1 has occupation measure ϕ^{P_1} satisfying

$$\left(\frac{1-V}{1-C_1(P_1)}\phi^{P_1}(i,a)\right)_{i=1}^{\ell-1} \in Q_1$$
$$\left(\frac{V}{C_1(P_1)}\phi^{P_1}(i,a)\right)_{i=\ell}^{n} \in Q_2$$
$$\sum_{a=1}^{k} \lambda(a)\phi^{P_1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_1}(\ell,a) = 0,$$

where, by Assumption 4.1 $C_1(P_1) > V$. Note that the scaled and truncated occupation measures $(\frac{1-V}{1-C_1(P_1)}\phi^{P_1}(i,a))_{i=1}^{\ell-1}$ and $(\frac{V}{C_1(P_1)}\phi^{P_1}(i,a))_{i=\ell}^{n}$ correspond to the policies P_1^1 and P_1^2, respectively, since scaling an occupation measure does not affect the associated policy. Thus,

$$\begin{aligned}
&\sum_{a=1}^{k} \lambda(a)\phi^{P_1^1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_1^2}(1,a) \\
&= \sum_{a=1}^{k} \lambda(a)\frac{1-V}{1-C_1(P_1)}\phi^{P_1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\frac{V}{C_1(P_1)}\phi^{P_1}(\ell,a) \\
&> \sum_{a=1}^{k} \lambda(a)\phi^{P_1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_1}(\ell,a) = 0,
\end{aligned}$$

where the last inequality follows since $\frac{V}{C_1(P_1)} < 1 < \frac{1-V}{1-C_1(P_1)}$. Applying a similar argument with P_k shows that

$$\sum_{a=1}^{k} \lambda(a)\phi^{P_k^1}(\ell-1,a) - \sum_{a=1}^{k} \mu(a)\phi^{P_k^2}(1,a) < 0.$$

By Statement 2 of Lemma 4.3.3, this implies that $g_1(\gamma) + g_2(\gamma)$ monotonically decreases towards $-\infty$ as γ moves away from the interval $[0, \tilde{\gamma}]$. Hence the supremum is finite and is attained by some $\gamma^* \in [0, \tilde{\gamma}]$. Since $g_1(\gamma) + g_2(\gamma)$ is concave and maximized at γ^*, $0 \in \partial(-g_1 - g_2)(\gamma^*)$. By Statement 2 of Lemma 4.3.3, there e xists $x_1^* \in \Omega_1(\gamma^*), x_2^* \in \Omega_2(\gamma^*)$ such that

$$\sum_{a=1}^{k} \lambda(a) x_1^*(\ell - 1, a) - \sum_{a=1}^{k} \mu(a) x_2^*(1, a) = 0.$$

Hence the concatenated measure

$$\phi^*(i,a) = \begin{cases} x_1^*(i,a) & i = 1, ..., \ell - 1 \\ x_2^*(i - \ell + 1, a) & i = \ell, ..., n \end{cases}$$

is feasible for the LP representation of B(V,ℓ). Since the problem is separable when Constraint (4.9) is removed, and noting that x_1^* and x_2^* are optimal for each of these separate optimization problems, it follows that ϕ^* is optimal for B(V,ℓ), as desired. ∎

Proposition 4.3.4 implies that to prove structural results for B(V,ℓ), we can instead prove structural results for SP1(γ^*) and SP2(γ^*). However, we need to be careful. First, we do not know what γ^* is, so we need to prove structural results for SP1(γ^*) and SP2(γ^*) *regardless* of the value of γ^*. Second, we do not know *which* optimal policies for each of the subproblems jointly satisfy (4.10), although we do know that they belong to Π^S. Hence, rather than proving the *existence* of optimal policies exhibiting a particular structure for each subproblem, we need to prove that *all* optimal stationary policies conform to this structure. In section 4.4, we consider the case where $c(\cdot), \lambda(\cdot)$, and $\mu(\cdot)$ are linear and prove that all optimal stationary policies for each subproblem exhibit an almost-monotone structure. Combining these results yields the existence of an almost-unimodal constrained optimal policy for B(V,ℓ). In Section 4.5 we extend these results to the convex case.

4.4 Linear Case

We first consider the case where the functions $c(\cdot), \lambda(\cdot)$, and $\mu(\cdot)$ are of the following forms:

$$c(a) = ca$$
$$\lambda(a) = \Lambda - \lambda a$$
$$\mu(a) = M + \mu a,$$

where $c > 0, \lambda, \mu > 0$, $\Lambda > \lambda k$, and $M > -\mu$ (to ensure that all three functions are positive). In addition, for our analysis we assume that $\lambda < \mu$.

4.4.1 Subproblem 1

The first subproblem, $SP_1(\gamma)$, is an unconstrained MDP on the augmented state space $\mathbb{X}_L := \{1, 2, ..., \ell - 1\}$ with modified cost function $c'(i, a) := c(a) + \gamma\lambda(a)\mathbb{1}_{\{\ell-1\}}(i)$. We aim to find conditions under which every optimal stationary policy is monotone. Note that for $\gamma \leq 0$, $c'(i, \cdot)$ is still increasing, and so the (only) optimal policy is to prescribe treatment 1 only. To see this, note that taking the cheapest action in every state yields a Markov chain which spends the maximum amount of time in state 1, which in turn has lower costs of treatment than any other state. So, moving forward we assume $\gamma > 0$. Since the transition rates $\lambda(a)$ and $\mu(a)$ are bounded, we apply the traditional uniformization techniques to consider each problem in discrete-time. In the analysis that follows, assume without loss of generality that the uniformization constant is 1, so that we can interpret $\lambda(a)$ and $\mu(a)$ as transition probabilities. For $m \in \mathbb{N}$, let $v_m(i)$ denote the m-horizon *value function*: the optimal (modified) cost accrued over m appointments, given

that the patient starts in state i. The finite-horizon optimality equations (FHOE) are then, for $i = 2, ..., \ell - 2$

$$v_{m+1}(i) = \min_a \{c(a) + \lambda(a)v_m(i+1) + \mu(a)v_m(i-1) + (1 - \lambda(a) - \mu(a))v_m(i)\}. \tag{FHOE}$$

For $i = 1$ and $i = \ell - 1$, we have

$$v_{m+1}(1) = \min_a \{c(a) + \lambda(a)v_m(2) + (1 - \lambda(a))v_m(1)\}$$
$$v_{m+1}(\ell - 1) = \min_a \{c(a) + \gamma\lambda(a) + \mu(a)v_m(\ell - 2) + (1 - \mu(a))v_m(\ell - 1)\}.$$

Define for $i = 1, ..., \ell - 2$

$$\Delta_m(i) := v_m(i+1) - v_m(i),$$

the marginal m-horizon benefit of starting from state $i + 1$ rather than i. Since every stationary policy yields an irreducible, positive-recurrent Markov chain, we know the following hold for the average-cost problem [39]

1. There exists a constant, J, representing the optimal expected long-run average cost, satisfying

   $$\lim_{m\to\infty} (v_{m+1}(i) - v_m(i)) = J$$

   for every $i \in \mathbb{X}$.

2. Let $z \in \mathbb{X}$ be a distinguished state. Then there exists a vector $h \in \mathbb{R}^{\ell-1}$ called the *relative value function* which satisfies

   $$h(i) = \lim_{m\to\infty} (v_m(i) - v_m(z))$$

   for every $i \in \mathbb{X}$. Additionally, there exists $\Delta \in \mathbb{R}^{\ell-2}$ satisfying

   $$\Delta(i) = \lim_{m\to\infty} \Delta_m(i)$$

   for every $i = 1, ..., \ell - 2$.

3. Every average-optimal stationary policy can be found by picking an action attaining the minimum (for each state) in the average cost optimality equations (ACOE). For $i = 2, ..., \ell - 2$, these equations are

$$h(i) + J = \min_a \{c(a) + \lambda(a)h(i+1) + \mu(a)h(i-1) + (1 - \lambda(a) - \mu(a))h(i)\}. \tag{ACOE}$$

For $i = 1$ and $i = \ell - 1$, they are

$$h(1) + J = \min_a \{c(a) + \lambda(a)h(2) + (1 - \lambda(a))h(1)\}$$
$$h(\ell - 1) + J = \min_a \{c(a) + \gamma\lambda(a) + \mu(a)h(\ell - 2) + (1 - \mu(a))h(\ell - 1)\}.$$

Substituting $\Delta(i), i = 1, ..., \ell - 2$ for $h(i), i = 1, ..., \ell - 1$, and using the linearity of $c(\cdot), \lambda(\cdot)$, and $\mu(\cdot)$, we can rewrite the ACOE

$$h(1) + J = \min_a \{(c - \lambda\Delta(1))a\} + \Lambda\Delta(1) + h(1)$$
$$h(i) + J = \min_a \{(c - \lambda\Delta(i) - \mu\Delta(i-1))a\} - M\Delta(i-1) + h(i) \quad i = 2, ..., \ell - 2$$
$$h(\ell - 1) + J = \min_a \{(c - \lambda\gamma - \mu\Delta(\ell - 2))a\} - M\Delta(\ell - 2) + h(\ell - 1),$$

where we recall that M and Λ are the intercepts for $\mu(\cdot)$ and $\lambda(\cdot)$, respectively. We can now express the argmins of the minimizations in terms of $\Delta(\cdot)$:

$$A_1^*(\gamma) = \begin{cases} \{1\} & \Delta(1) < \frac{c}{\lambda} \\ A & \Delta(1) = \frac{c}{\lambda} \\ \{k\} & \Delta(1) > \frac{c}{\lambda} \end{cases},$$

$$A_i^*(\gamma) = \begin{cases} \{1\} & \lambda\Delta(i) + \mu\Delta(i-1) < c \\ A & \lambda\Delta(i) + \mu\Delta(i-1) = c \\ \{k\} & \lambda\Delta(i) + \mu\Delta(i-1) > c \end{cases},$$

$$A_{\ell-1}^*(\gamma) = \begin{cases} \{1\} & \Delta(\ell-2) < \frac{c-\lambda\gamma}{\mu} \\ A & \Delta(\ell-2) = \frac{c-\lambda\gamma}{\mu} \\ \{k\} & \Delta(\ell-2) > \frac{c-\lambda\gamma}{\mu} \end{cases}.$$

Note that even though the dependence of $A_i^*(\gamma)$ on γ is explicit when $i = \ell - 1$, $A_1^*(\gamma), ..., A_{\ell-2}^*(\gamma)$ also depend on γ since it impacts the value functions. At first glance, it also appears that if $\Delta(\cdot)$ is strictly increasing, then any optimal stationary policy should be non-decreasing in the following sense:

$$\max_{a \in A_i^*(\gamma)} a \leq \min_{a \in A_{i+1}^*(\gamma)} a$$

for every $i = 1, ..., \ell - 2$, $\gamma \in \mathbb{R}$. For simplicity, we write $\mathbb{A}_i^*(\gamma) \leq \mathbb{A}_j^*(\gamma)$ if

$$\max_{a \in A_i^*(\gamma)} a \leq \min_{a \in A_j^*(\gamma)} a.$$

The following results more rigorously specify this concept.

Proposition 4.4.1 *For $i = 1, ..., \ell - 2$, $\Delta(i) \geq 0$.*

The proof of Proposition 4.4.1 easily follows by induction on v_m, and is omitted for brevity. The next result allows us to find structure in $SP_1(\gamma)$.

Lemma 4.4.2 *$\Delta(i)$ is strictly increasing in i.*

Proof. By induction. Our base case involves showing that $\Delta(2) > \Delta(1)$. First note that we must have $\Delta(1) > 0$, for if $\Delta(1) = 0$, then the ACOE yield

$$J = \min_{a \in \mathbb{A}} c(a) = c(1),$$

which means that P_1 is optimal for B(V,ℓ), a contradiction by Assumption 4.1. Letting $a_1^* \in \mathbb{A}_1^*(\gamma), a_2^* \in \mathbb{A}_2^*(\gamma)$, we have from the ACOE

$$\begin{aligned} c(a_1^*) + \lambda(a_1^*)\Delta(1) &= J \\ &= c(a_2^*) + \lambda(a_2^*)\Delta(2) - \mu(a_2^*)\Delta(1) \\ &\leq c(a_1^*) + \lambda(a_1^*)\Delta(2) - \mu(a_1^*)\Delta(1). \end{aligned}$$

Thus

$$0 \geq (\lambda(a_1^*) + \mu(a_1^*))\Delta(1) - \lambda(a_1^*)\Delta(2),$$

so

$$\Delta(2) \geq \frac{(\lambda(a_1^*) + \mu(a_1^*))}{\lambda(a_1^*)}\Delta(1) > \Delta(1),$$

completing our base case. Now supose that $\Delta(i+1) > \Delta(i)$ for $i = 1, ..., \ell - 4$. By the ACOE, we have

$$\begin{aligned} & c(a_{i+1}^*) + \lambda(a_{i+1}^*)\Delta(i+1) - \mu(a_{i+1}^*)\Delta(i) \\ & = J \\ & \leq c(a_{i+1}^*) + \lambda(a_{i+1}^*)\Delta(i+2) - \mu(a_{i+1}^*)\Delta(i+1), \end{aligned}$$

so rearranging terms yields

$$\begin{aligned} 0 &\geq (\lambda(a_{i+1}^*) + \mu(a_{i+1}^*))\Delta(i+1) - \mu(a_{i+1}^*)\Delta(i) - \lambda(a_{i+1}^*)\Delta(i+2) \\ &> \lambda(a_{i+1}^*)\Delta(i+1) - \lambda(a_{i+1}^*)\Delta(i+2), \end{aligned}$$

where the last inequality follows by the inductive hypothesis. This completes the proof. ∎

Lemma 4.4.2 yields the following proposition.

Proposition 4.4.3 *In the linear case, for any* $\gamma \in \mathbb{R}$ *and any* $a_1^* \in A_1^*(\gamma), ..., a_{\ell-1}^* \in A_{\ell-1}^*(\gamma)$,

$$a_1^* \leq a_2^* \leq ... \leq a_{\ell-2}^*.$$

Additionally, if $\gamma \geq \frac{c}{\lambda+\mu}$, *we have that* $a_{\ell-2}^* \leq a_{\ell-1}^*$.

Proof. To show that $a_1^* \leq a_2^*$, we consider the form of the argmins $A_1^*(\gamma)$ and $A_2^*(\gamma)$. If $A_1^*(\gamma) = \{1\}$, then we do not care what $A_2^*(\gamma)$ is since 1 is the lowest action. If $A_1^*(\gamma) = A$ or $\{k\}$, then $\Delta(1) \geq \frac{c}{\lambda}$. This implies $A_2^*(\gamma) = \{k\}$, since

$$\lambda\Delta(2) + \mu\Delta(1) > (\lambda+\mu)\Delta(1) > c + \frac{c\mu}{\lambda} > c.$$

To show that $a_i^* \leq a_{i+1}^*$ for $i = 2, ..., \ell-2$, we see that (via Lemma 4.4.2)

$$\lambda\Delta(i+1) + \mu\Delta(i) > \lambda\Delta(i) + \mu\Delta(i-1).$$

To show that $a_{\ell-2}^* \leq a_{\ell-1}^*$, we need to specify additional conditions on γ. We can ignore the case where $A_{\ell-2}^*(\gamma) = \{1\}$, since the claim follows trivially. So suppose that $A_{\ell-2}^*(\gamma) = A$ or $\{k\}$ and so

$$\lambda\Delta(\ell-2) + \mu\Delta(\ell-3) \geq c.$$

We would like to find conditions on γ under which this implies that $A_{\ell-1}^*(\gamma) = \{k\}$. This involves showing that $\Delta(\ell-2) > \frac{c-\gamma\lambda}{\mu}$. Suppose not. So $\Delta(\ell-2) \leq \frac{c-\gamma\lambda}{\mu}$. Note that, since $\Delta(\ell-2) > \Delta(\ell-3)$, we have

$$\lambda\Delta(\ell-2) + \mu\Delta(\ell-3) < (\lambda+\mu)\frac{c-\gamma\lambda}{\mu} = RHS(\gamma).$$

We arrive at a contradiction if γ satisfies $RHS(\gamma) \leq c$. Solving for γ yields $\gamma \geq \frac{c}{\lambda+\mu}$. ∎

4.4.2 Subproblem 2

The second subproblem in $L(V, \ell)$ is an unconstrained MDP defined on the states $\{\ell, \ell+1, ..., n\}$ with cost function $c(i, a) = c(a) + \gamma \mathbb{1}_{\{\ell\}}(i)$. Similar to $SP_1(\gamma)$, the ACOE are

$$\begin{aligned} h(\ell) + J &= \min_{a \in \mathbb{A}} \{c(a) - \gamma\mu(a) + \lambda(a)\Delta(\ell)\} + h(\ell) \qquad \text{(ACOE2)} \\ h(i) + J &= \min_{a \in \mathbb{A}} \{c(a) + \lambda(a)\Delta(i) - \mu(a)\Delta(i-1)\} \\ h(n) + J &= \min_{a \in \mathbb{A}} \{c(a) - \mu(a)\Delta(n-1)\}. \end{aligned}$$

Similar analysis to that done in Section 4.4.1 yields the following series of results.

Proposition 4.4.4 *$\Delta(i) \geq 0$ for $i = \ell, \ell+1, ..., n$.*

Lemma 4.4.5 *$\Delta(\ell) > \Delta(\ell+1) > ... > \Delta(n-1)$.*

Proposition 4.4.6 *For any $\gamma \in \mathbb{R}$ and any $a_\ell^* \in A_\ell^*(\gamma), ..., a_n^* \in A_n^*(\gamma)$,*

$$a_{\ell+1}^* \geq ... \geq a_n^*.$$

Additionally, if $\gamma \geq \frac{c}{\lambda+\mu}$, we have that $a_\ell^* \leq a_{\ell+1}^*$.

Applying Proposition 4.3.4 yields the following result.

Theorem 4.4.7 *Consider the control problem B(V,ℓ) with linear $c(\cdot), \lambda(\cdot)$, and $\mu(\cdot)$, and V satisfying Assumption 4.1. For a stationary policy σ, let $A_i(\sigma)$ be the set of actions that are chosen by σ in state i with strictly positive probability. Then there exists an optimal stationary policy σ^* satisfying*

$$A_1(\sigma^*) \leq ... \leq A_{\ell-2}(\sigma^*)$$
$$A_{\ell+1}(\sigma^*) \geq ... \geq A_n(\sigma^*).$$

4.5 Convex Case

We now extend most of the results from the previous section to the case where the cost function c as well as the rate functions $\lambda(\cdot)$ and $\mu(\cdot)$ are convex. To make our analysis easier, we also assume that these functions are differentiable, when extended to the reals. We also make the assumption that $\lambda(\cdot)$ and $\mu(\cdot)$ are strictly monotone. We make use of the following observation.

Remark 4.5.1 *We did not use linearity in the proofs for Propositions 4.4.1, 4.4.4 and for Lemmas 4.4.2 and 4.4.5, so they hold in the general case.*

Theorem 4.5.2 *In the convex case, for any $\gamma \in \mathbb{R}$ and any $a_1^* \in A_1^*(\gamma), ..., a_{\ell-1}^* \in A_{\ell-1}^*(\gamma)$,*

$$a_1^* \leq a_2^* \leq ... \leq a_{\ell-2}^*$$

for the first subproblem $SP_1(\gamma)$. Additionally, for any $\gamma \in \mathbb{R}$ and any $a_\ell^* \in A_\ell^*(\gamma), ..., a_n^* \in A_n^*(\gamma)$,

$$a_{\ell+1}^* \geq ... \geq a_n^*.$$

To prove that this is the case, we first consider the relaxed problem of minimizing $c(a) + \alpha\lambda(a) + \beta(-\mu)(a)$ over $a \in [1, k]$, where $[1, k]$ is the closed interval of reals. After showing that the argmin is non-decreasing as α and β jointly increase, we show that the same holds true when restricted to $a \in \{1, ..., k\}$. These two steps are summarized in the following lemmas.

Lemma 4.5.3 *Let $\alpha, \beta \geq 0$ and $\hat{\alpha} \geq \alpha, \hat{\beta} \geq \beta$. Define*

$$A := \arg\min_{a \in [1,k]} \{c(a) + \alpha\lambda(a) + \beta(-\mu)(a)\}$$

$$\hat{A} := \arg\min_{a \in [1,k]} \{c(a) + \hat{\alpha}\lambda(a) + \hat{\beta}(-\mu)(a)\}.$$

Then

$$\max\{a : a \in A\} \leq \min\{a : a \in \hat{A}\}.$$

For notational convenience, we write $A \leq \hat{A}$.

Proof. Differentiating yields

$$\frac{d}{da}(c(a) + \alpha\lambda(a) - \beta\mu(a)) = c'(a) + \alpha\lambda'(a) - \beta\mu'(a) =: f(a, \alpha, \beta).$$

Noting that $\lambda'(a), \mu'(a) > 0$, it follows that the f is decreasing as α and β jointly increase, for any $a \in [1, k]$. Let a^* be optimal for α, β. It follows from convexity that $f(a^*, \alpha, \beta) = 0$. Let $\alpha' \geq \alpha, \beta' \geq \beta$. Then

$$0 = f(a^*, \alpha, \beta) > f(a^*, \alpha', \beta').$$

Since $c(a) + \alpha'\lambda(a) - \beta'\mu(a)$ is convex in a, it follows that $f(\cdot, \alpha', \beta')$ is non-decreasing, and so $f(a^{**}, \alpha', \beta') = 0$ for some $a^{**} \geq a^*$, completing the proof. ∎

Lemma 4.5.4 *The result from Lemma 4.5.3 holds with $[1, k]$ replaced by $\{1, ..., k\}$.*

Proof. Again pick $\alpha, \beta \geq 0$ and $\alpha' \geq \alpha, \beta' \geq \beta$. Let $a^* \in [1, k]$ be the optimal action for α, β, and let $a^{**} \in [a^*, k]$ denote the optimal action for α', β'. For simplicity, assume that the minimum for α, β restricted to $\{1, ..., k\}$ is attained at $\lfloor a^* \rfloor$ or $\lceil a^* \rceil$, since this can be generalized by assuming they are attained at $\lfloor \bar{a} \rfloor$ or $\lceil \hat{a} \rceil$, where $\bar{a}$ is the smallest minimizer of the function at α, β and $\hat{a}$ is the largest. If $a^{**} \geq \lceil a^* \rceil$, then the result is trivial. So assume that $a^{**} \in [a^*, \lceil a^* \rceil)$. Additionally, if the minimum for α, β is attained only at $\lfloor a^* \rfloor$ the result holds trivially, so assume that the minimum is attained at $\lceil a^* \rceil$. We want to show that the minimum for α', β' must be attained only at $\lceil a^* \rceil$. This holds if

$$c(\lfloor a^* \rfloor) + \alpha' \lambda(\lfloor a^* \rfloor) - \beta' \mu(\lfloor a^* \rfloor) > c(\lceil a^* \rceil) + \alpha' \lambda(\lceil a^* \rceil) - \beta' \mu(\lceil a^* \rceil).$$

For notational convenience, let $F(a, \alpha, \beta) = c(a) + \alpha\lambda(a) - \beta\mu(a)$. By assumption, we know that $F(\lfloor a^* \rfloor, \alpha, \beta) \geq F(\lceil a^* \rceil, \alpha, \beta)$. Realizing that

$$\begin{aligned} F(\lfloor a^* \rfloor, \alpha', \beta') &= F(\lfloor a^* \rfloor, \alpha, \beta) + (\alpha' - \alpha)\lambda(\lfloor a^* \rfloor) - (\beta' - \beta)\mu(\lfloor a^* \rfloor) \\ F(\lceil a^* \rceil, \alpha', \beta') &= F(\lceil a^* \rceil, \alpha, \beta) + (\alpha' - \alpha)\lambda(\lceil a^* \rceil) - (\beta' - \beta)\mu(\lceil a^* \rceil), \end{aligned}$$

we obtain that

$$\begin{aligned} F(\lfloor a^* \rfloor, \alpha, \beta) - F(\lceil a^* \rceil, \alpha', \beta') &= (F(\lfloor a^* \rfloor, \alpha, \beta) - F(\lceil a^* \rceil, \alpha, \beta)) \\ &\quad + (\alpha' - \alpha)(\lambda(\lfloor a^* \rfloor) - \lambda(\lceil a^* \rceil)) \\ &\quad + (\beta' - \beta)(\mu(\lceil a^* \rceil) - \mu(\lfloor a^* \rfloor)) \\ &> 0, \end{aligned}$$

where the inequality follows since $F(\lfloor a^* \rfloor, \alpha, \beta) \geq F(\lceil a^* \rceil, \alpha, \beta)$, $\lambda(\cdot)$ is decreasing, and $\mu(\cdot)$ is increasing. This completes our proof. ∎

We now combine Lemma 4.4.5 with Lemma 4.5.4 to prove Theorem 4.5.2.

Proof of Theorem 4.5.2.

It immediately follows from 4.4.5 with Lemma 4.5.4 that for $SP_1(\gamma)$ we have

$$a_2^* \leq a_3^* \leq ... \leq a_{\ell-2}^*$$

and for $SP_2(\gamma)$ we have

$$a_{\ell+1}^* \geq a_{\ell+2}^* \geq ... \geq a_{n-1}^*.$$

To get that $a_1^* \leq a_2^*$, note that a_1^* is a minimizer of

$$c(a) + \Delta(1)\lambda(a) - 0\mu(a)$$

and that a_2^* is a minimizer of

$$c(a) + \Delta(2)\lambda(a) - \Delta(1)\mu(a).$$

Since $\Delta(2) > \Delta(1)$ and $\Delta(1) \geq 0$, this is covered by the general case, and so $a_1^* \leq a_2^*$. A similar argument yields that $a_{n-1}^* \geq a_n^*$ for $SP_2(\gamma)$, completing the proof. ∎

By combining Theorem 4.5.2 and Proposition 4.3.4, we get the following result for B(V,ℓ) in the convex case.

Corollary 4.5.5 *Consider the control problem B(V,ℓ) with convex, strictly monotone, and differentiable $c(\cdot), \lambda(\cdot)$, and $\mu(\cdot)$, and V satisfying Assumption 4.1. For a stationary policy σ, let $A_i(\sigma)$ be the set of actions that are chosen by σ in state i with strictly positive probability. Then there exists an optimal stationary policy σ^* satisfying*

$$A_1(\sigma^*) \leq ... \leq A_{\ell-2}(\sigma^*)$$
$$A_{\ell+1}(\sigma^*) \geq ... \geq A_n(\sigma^*).$$

4.6 General Models with Non-decreasing Optimal Policies

In this section we consider a personalized medicine model with more general transition dynamics. By considering different transition dynamics, we make the results of this chapter applicable to a wider variety of illnesses, particularly those in which the patient's condition can deteriorate drastically between appointments. We aim to find conditions under which there exists a non-decreasing constrained-optimal policy.

4.6.1 Formulation

We retain the definitions introduced in section 4.1 with the exception of the transition probabilities. We now refer to these probabilities more generally as $p(j|i,a)$, which represents the probability that a patient in health state i undergoing treatment a will be in health state j at the beginning of the next treatment period. We impose the following assumptions of the transition probabilities:

(A2) **Stochastic dominance in state of health**: For every $a \in A$ and every $z \in \mathbb{X}$, $\sum_{j=z}^{n} p(j|i,a)$ is non-decreasing in i.

(A3) **Stochastic dominance in treatment**: For every $i \in \mathbb{X}$ and every $z \in \mathbb{X}$, $\sum_{j=z}^{n} p(j|i,a)$ is non-increasing in a.

(A4) **Marginal effectiveness of treatments**: Define, for a function $f : \mathbb{X} \mapsto \mathbb{R}_+$, state $x \in \mathbb{X}$, and action $a \in A$,

$$q_f(x,a) := \sum_{j=1}^{n} p(j|x,a) f(j).$$

We assume that, for every $i \in [n-1], a \in [k-1]$, and for any non-decreasing f,

$$q_f(x,a) - q_f(x,a+1) \leq q_f(x+1,a) - q_f(x+1,a+1).$$

We can interpret these assumptions as follows. Assumption (A2) says that the patient's state of health is more likely to worsen the "less healthy" the patient currently is. Assumption (A3) states that more expensive (and hence more effective) treatments are less likely to leave a patient in a poor state of health. Finally, the last assumption (A4) implies that the marginal benefit of picking a more effective treatment increases as the patient's state of health worsens. In other words, the healthier the patient, the less valuable it is to switch to a more effective treatment. Under these assumptions, we can show that there exists a constrained-optimal treatment policy that is non-decreasing: more effective treatments are used in worse states of health. Furthermore, we show that there exists such a policy that is also one-randomized: there is at most one state of health in which we randomize between treatments. The result is summarized in the following theorem.

Theorem 4.6.1 *For any cutoff level $\ell \in \mathbb{X}$ and any feasible, non-trivial $V \in (0,1)$, there exists an optimal stationary policy for the constrained problem B(V,ℓ) which is non-decreasing and one-randomized.*

In order to prove Theorem 4.6.1, we use the Lagrangian dual defined in section 4.3 problem, LD(V,ℓ). Note that this Lagrangian is different from the decomposition-based Lagrangian used in the previous sections. This is due to the fact that the more general transition probabilities prevent us from decomposing the problem into two simpler sub-problems. The inner minimization of the Lagrangian dual problem is an unconstrained MDP with cost function $c(a) + \gamma \mathbb{1}\{i \geq \ell\}$, and can

be written as the linear program

$$\begin{aligned}
&\min \sum_{i=1}^{n}\sum_{a=1}^{k} \left(c(a) + \gamma \mathbb{1}\{i \geq \ell\}\right) \phi(i,a) \\
&\text{s.t.} \sum_{j=1}^{n}\sum_{a=1}^{k} \left(\mathbb{1}\{j=i\} - p(i|j,a)\right) \phi(j,a) = 0, \qquad i = 1, ..., n \\
&\quad \sum_{i=1}^{n}\sum_{a=1}^{k} \phi(i,a) = 1 \\
&\quad \phi(i,a) \geq 0, \qquad i = 1, ..., n, a = 1, ..., k.
\end{aligned}$$

This formulation of the inner minimization allows us to characterize the structure of constrained-optimal policies. First we show, for fixed $\gamma \geq 0$, structural properties of the unconstrained MDP. The following proposition states that for any γ, the relative value function is non-decreasing.

Proposition 4.6.2 *For any $\gamma \geq 0$, the relative value function of the Lagrangian relaxation problem, $h_\gamma(\cdot)$, is non-decreasing.*

Proof. It suffices to show the claim is true for the finite-horizon value functions, $v_{n,\gamma}$. For simplicity, we will omit the dependence on γ. The claim is trivial for $n = 0$ as $v_0 = 0$ identically. Assume that $v_n(\cdot)$ is non-decreasing and look at $n+1$. Fix $i \in [n-1]$, and let a_x denote an optimal action in state $x \in \mathbb{X}$. Then, since a_{i+1} is potentially sub-optimal in state i,

$$\begin{aligned}
v_{m+1}(i+1) - v_{m+1}(i) &\geq \left(c(a_{i+1}) + \gamma \mathbb{1}\{i+1 \geq \ell\} + \sum_{j=1}^{n} p(j|i+1, a_{i+1}) v_m(j)\right) \\
&\quad - \left(c(a_{i+1}) + \gamma \mathbb{1}\{i \geq \ell\} + \sum_{j=1}^{n} p(j|i, a_{i+1}) v_m(j)\right) \\
&= \gamma \left(\mathbb{1}\{i+1 \geq \ell\} - \mathbb{1}\{i \geq \ell\}\right) \\
&\quad + \sum_{j=1}^{n} p(j|i+1, a_{i+1}) v_m(j) - \sum_{j=1}^{n} p(j|i, a_{i+1}) v_m(j).
\end{aligned}$$

Since $i \geq \ell$ implies $i+1 \geq \ell$ and $\gamma \geq 0$, the first term is non-negative. To see that $\sum_{j=1}^{n} p(j|i+1, a_{i+1}) v_m(j) - \sum_{j=1}^{n} p(j|i, a_{i+1}) v_m(j) \geq 0$, note that by assumption the distribution $p(\cdot|i+1, a_{i+1})$ is stochastically greater than $p(\cdot|i, a_{i+1})$ and that $v_m(\cdot)$ is non-decreasing by the inductive hypothesis. Hence $v_{m+1}(i+1) - v_{m+1}(i) \geq 0$, completing the proof by induction. ∎

Proposition 4.6.3 *For any $\gamma \geq 0$, every optimal stationary deterministic policy for the unconstrained Lagrangian relaxation MDP with multiplier γ is non-decreasing.*

Proof. Suppose not. Then there exists states x and y with $x < y$ and optimal actions $a_x > a_y$. By potential sub-optimality, we have

$$c(a_x) + \gamma \mathbb{1}\{x \geq \ell\} + q_h(x, a_x) \leq c(a_y) + \gamma \mathbb{1}\{x \geq \ell\} + q_h(x, a_y)$$
$$c(a_y) + \gamma \mathbb{1}\{y \geq \ell\} + q_h(y, a_y) \leq c(a_x) + \gamma \mathbb{1}\{y \geq \ell\} + q_h(y, a_x).$$

Rearranging terms, we get that

$$q_h(x, a_y) - q_h(x, a_x) \geq c(a_x) - c(a_y) \geq q_h(y, a_y) - q_h(y, a_x),$$

contradicting assumption (A3). ∎

4.7 Conclusion

We considered a controlled Markov chain model motivated by a personalized medicine problem in which treatments are prescribed to a patient at discrete

points in time. Patient health was modeled as a Markov chain with state space $\mathbb{X} = \{1, ..., n\}$, where larger states correspond to worse patient health. The goal is to minimize the cost of treatment while keeping the amount of time spent in "undesirable" states of health is below a given level. We first considered a simple model in which patient health is modeled by a controlled truncated birth-death process (or random walk), which is reasonable assuming that a patient's state of health cannot deteriorate too quickly between appointments. For this model, we proved that if the birth rates, death rates, and costs of treatments are convex that a unimodal optimal policy exists. We then considered a more general model where patient health can deteriorate more quickly, and found conditions under which a monotone increasing optimal policy is optimal. Finding more specific conditions under which such an optimal policy exists is a promising direction for future work.

APPENDIX A

APPENDIX FOR CHAPTER 2

A.1 Verification of Assumptions

We verify assumptions (A1)-(A4), that are needed to apply **Theorem 12.7** from Altman [3] and hence establish the equivalence between the constrained problem and its Lagrangian dual in both applications. The assumptions are stated as follows. Let $\mathcal{K} := \{(x, a) : x \in \mathbb{X}, a \in A(x)\}$ and let $\mathbb{K}$ denote its Borel σ-algebra generated by the rectangles $\{(x, A) : x \in \mathbb{X}, A \subseteq A(x)\}$. We need to show:

(A1) For any state $x \in \mathbb{X}$, if $\{a_n\} \subseteq A(x)$ is a sequence of actions with $a_n \to a \in A(x)$ as $n \to \infty$, then for every $y \in \mathbb{X}$, $\lim_{n\to\infty} P(y|x, a_n) = P(y|x, a)$.

(A2) Under any stationary policy, the induced MC contains a single ergodic class, and absorption into the positive recurrent class takes place in a finite expected time. This corresponds to assumption (B1) in Chapter 11 of Altman [3].

(A3) The immediate cost functions c_1 and c_2 are bounded below.

(A4) There exists an increasing sequence of compact sets $(K_n)_{n=1}^{\infty} \subseteq \mathcal{K}$ with $\bigcup_{n\in\mathbb{N}} K_n = \mathcal{K}$ and $\liminf_{n\to\infty}\{r(k) : k \notin K_n\} = \infty$ for $r = c_1, c_2$.

Assumption (A1) trivially holds for finite action sets (for each $x \in \mathbb{X}$). It can be easily checked that every stationary policy yields an irreducible MC. Under the traffic conditions stated in the main body of the chapter for both the parallel and tandem queue settings, we showed that every stationary policy yields a positive-recurrent MC, and so assumption (A2) holds. (A3) holds trivially since $c_1(\cdot)$ and

$c_2(\cdot)$ are non-negative. To see that (A4) holds, take $K_n = \{((i,j),0) : i,j \leq n\}$ so that for every n,

$$\inf\{r(k) : k \notin K_n\} = n+1$$

for $r(\cdot) = c_1(\cdot), c_2(\cdot)$, which diverges as $n \to \infty$.

A.2 Notation and Definitions

Recall a stationary policy $\sigma \in \Pi^S$ is defined as a sequence of measures $(\sigma_x)_{x\in\mathbb{X}}$. For $A \subseteq A(x)$, $\sigma_x(A) \in [0,1]$ is the probability that an action $a \in A$ is taken in state x under policy σ. For a sequence of stationary policies $(\sigma_n)_{n=0}^{\infty}$, we say that σ_n **converges pointwise** to σ (for some $\sigma \in \Pi^S$) if $(\sigma_n)_x(A) \to \sigma_x(A)$ as $n \to \infty$ for all $x \in \mathbb{X}, A \subseteq \mathbb{A}(x)$. If, for all bounded and continuous functions $g : A(x) \mapsto \mathbb{R}$, we have

$$\lim_{n\to\infty} \int_{A(x)} g(z) d(\sigma_n)_x(z) = \int_{A(x)} g(z) d\sigma_x(z)$$

for every $x \in \mathbb{X}$, we say that σ_n **converges weakly** to σ. Under the assumption (A2), each stationary policy ν induces an **occupation measure** f^ν satisfying

$$f^\nu(x,A) = \nu_x(A)\pi^\nu(x), \;\; x \in \mathbb{X}, A \subseteq A(x),$$

where π^ν is the stationary distribution of the process induced by policy ν. Intuitively, the occupation measure of a particular policy σ describes the long-run average fraction of time the induced process spends in each state-action pair. An important property of a set of occupation measures $\{f^\nu : \nu \in I\}$ over some class of policies I is tightness. In our context, $\{f^\nu : \nu \in \Pi^S\}$ being tight means that for any desired probability level, p, we can find a set of state-action pairs $K \subseteq \mathcal{K}$ such that the long-run fraction of time spent in K is at least p under any stationary policy. A more formal definition is provided below.

Definition A.2.1 *Let $\mathcal{K} := \mathbb{X} \times A$ and let $\mathbb{K}$ denote its Borel σ-algebra. We say the set of occupation measures $\{f^\nu : \nu \in \Pi^S\}$ over the set of stationary policies is* ***tight*** *if, for every $\epsilon > 0$, there exists $K_\epsilon \in \mathbb{K}$ such that*

$$f^\nu(K_\epsilon) > 1 - \epsilon$$

for every $\nu \in \Pi^S$.

A.3 Proof of Lemma 2.5.4

Given these preliminaries, we can now state the following theorem from Altman [3].

Theorem A.3.1 (adapted from Theorem 11.2 (i) in Altman [3])
Let $\Pi \subseteq \Pi^S$. If $\{f^\nu : \nu \in \Pi\}$ is tight, then f is weakly continuous over Π. That is, for any sequence of polices $(\nu_n)_{n=0}^\infty \subseteq \Pi$ converging weakly to $\nu \in \Pi$, we have that f^{ν_n} converges to f^ν weakly.

Verifying the assumptions of the theorem involves two steps: showing that pointwise convergence of policies implies weak convergence, and showing that the set of occupation measures over the class of stationary policies is tight. We verify both of these with the following lemmas.

Lemma A.3.2 *Let σ be a stationary policy, and let $(\sigma_n)_{n=0}^\infty$ be a sequence of stationary policies. Suppose that $\sigma_n \to \sigma$ pointwise. Then $\sigma_n \to \sigma$ weakly.*

Proof. Since $\sigma_n \to \sigma$ pointwise, and the action sets $\mathbb{A}(x)$ are finite for every $x \in \mathbb{X}$, we have

$$\lim_{n\to\infty} (\sigma_n)_x(\{0\}) = \sigma_x(\{0\})$$
$$\lim_{n\to\infty} (\sigma_n)_x(\{1\}) = \sigma_x(\{1\})$$

for every $x \in \hat{\mathbb{X}}$. Hence for any bounded, continuous function $g : \mathbb{A} \mapsto \mathbb{R}$ and every $x \in \mathbb{X}$

$$\begin{aligned}\lim_{n\to\infty} \int_{A(x)} g(a) d(\sigma_n)_x(a) &= \lim_{n\to\infty} \sum_{a=0}^{1} g(a)(\sigma_n)_x(\{a\}) \\ &= \sum_{a=0}^{1} g(a)\sigma_x(\{a\}) \\ &= \int_{A(x)} g(a) d(\sigma_n)_x(a),\end{aligned}$$

completing the proof. ∎

Lemma A.3.3 *The set of occupation measures $\{f^\nu : \nu \in \Pi^S\}$ is tight.*

Proof. Note that under our assumptions, every stationary policy yields a stable Markov process with finite expected long-run average number of customers in system. Hence for $k = 1, 2$, we have that $\sup_{\nu\in\Pi^S} C_k(\nu) =: \hat{C}_k < \infty$. Define for $N \in \mathbb{N}$ the compact set $S_N := \{(i, j, A(i, j)) : i, j \leq N\}$. If $\{f^\nu : \nu \in \Pi^S\}$ is not tight, then by definition, there exists $\tilde{\epsilon} > 0$ such that for every $N \in \mathbb{N}$, there is a policy $\nu_N \in \Pi^S$ with

$$1 - \tilde{\epsilon} \geq f^{\nu_N}(S_N) = \sum_{(i,j):i,j\leq N} \pi^{\nu_N}(i, j),$$

and thus

$$\sum_{(i,j):i,j>N} \pi^{\nu_N}(i, j) \geq \tilde{\epsilon}.$$

Hence, picking $\hat{N}$ such that $\tilde{\epsilon}\hat{N} > \hat{C}_1$, we have that

$$C_1(\nu_{\hat{N}}) \geq \sum_{(i,j):i,j>\hat{N}} i\pi^{\nu_{\hat{N}}}(i,j) > \hat{N} \sum_{(i,j):i,j>\hat{N}} \pi^{\nu_{\hat{N}}}(i,j) \geq \tilde{\epsilon}\hat{N} > \hat{C}_1,$$

a contradiction. Thus $\{f^\nu : \nu \in \Pi^S\}$ is tight, completing the proof. ∎

Proof of Lemma 2.5.4:

Proof. We know for any bounded and continuous $h : \mathcal{K} \mapsto \mathbb{R}$, **Theorem** A.3.1 yields

$$\lim_{n\to\infty} \int_{\mathcal{K}} h(z) df^{\sigma_n}(z) = \int_{\mathcal{K}} h(z) df^{\sigma}(z).$$

Define, for $y \in \mathbb{X}, a \in \mathbb{A}(y)$, the indicator

$$\mathbb{1}_x(y,a) = \begin{cases} 1 & y = x \\ 0 & y \neq x. \end{cases}$$

Since $\mathbb{1}_x(y,a)$ is bounded and continuous on $\mathcal{K}$, applying the theorem yields that, for any $x \in \mathbb{X}$,

$$\lim_{n\to\infty} \pi^{\sigma_n}(x) = \lim_{n\to\infty} \int_{\mathcal{K}} \mathbb{1}_x(z) df^{\sigma_n}(z) = \int_{\mathcal{K}} \mathbb{1}_x(z) df^{\sigma}(z) = \pi^{\sigma}(x),$$

as desired. ∎

A.4 Proof of Theorem 2.5.3

We prove the result for class 1 costs ($k = 1$). The proof for $k = 2$ is analogous. For each policy σ_n in the sequence, define the process $\{X^{\sigma_n}(t) : t \geq 0\}$ on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that, for each policy, this is a two-dimensional

Markov chain: $X^{\sigma_n}(t) = (I^{\sigma_n}(t), J^{\sigma_n}(t))$, where the k^{th} component represents the number of class k customers in the system at time t. For a stationary policy $\tilde{\sigma}$, let $X^{\tilde{\sigma}}(\infty) = (I^{\tilde{\sigma}}(\infty)), J^{\tilde{\sigma}}(\infty))$ denote the limiting number of customers in the system under policy $\tilde{\sigma}$, and note that $X^{\tilde{\sigma}}(\infty)$ is distributed according to the stationary distribution $\pi^{\tilde{\sigma}}$. By Proposition 2.5.2, there exists a random variable $\hat{X}$ defined on $(\Omega, \mathcal{F}, \mathbb{P})$ such that,

$$|I^{\sigma_n}(\infty)| = I^{\sigma_n}(\infty) \leq \hat{X} \quad a.s.$$
$$\mathbb{E}[|\hat{X}|] = \mathbb{E}[\hat{X}] < \infty.$$

If $I^{\sigma_n}(\infty)$ converges in distribution to $I^{\sigma}(\infty)$ as $n \to \infty$, then we can apply the dominated convergence theorem to yield

$$\lim_{n\to\infty} C_1(\sigma_n) = \lim_{n\to\infty} \mathbb{E}[I^{\sigma_n}(\infty)] = \mathbb{E}[I^{\sigma}(\infty)] = C_1(\sigma),$$

proving the result. Thus it suffices to show this distributional convergence. Formally, we need to show that for each $i = 0, 1, 2, ...,$

$$\lim_{n\to\infty} \mathbb{P}(I^{\sigma_n}(\infty) \leq i) = \mathbb{P}(I^{\sigma}(\infty) \leq i).$$

Using total probability, the fact that $X^{\tilde{\sigma}} \sim \pi^{\tilde{\sigma}}$, and the stationary distribution convergence from Lemma 2.5.4, this is equivalent to showing

$$\lim_{n\to\infty} \sum_{m=0}^{i} \sum_{j=0}^{\infty} \pi^{\sigma_n}(m,j) = \sum_{m=0}^{i} \sum_{j=0}^{\infty} \lim_{n\to\infty} \pi^{\sigma_n}(m,j) = \sum_{m=0}^{i} \sum_{j=0}^{\infty} \pi^{\sigma}(m,j).$$

Note that if we can interchange the limit (taking $n \to \infty$) and the infinite sum (over j), then we have proved the result. To justify this interchange we show for every $m = 0, 1, 2, ...,$ the sequence $(\sum_{j=0}^{N} \pi^{\sigma_n}(m,j))_{N=0}^{\infty}$ is uniformly convergent. That is, for every $\epsilon > 0$, there exists $N(\epsilon)$ such that for all $N \geq N(\epsilon)$,

$$\sum_{j=N+1}^{\infty} \pi^{\sigma_n}(m,j) < \epsilon, \quad n = 0, 1, 2, ...$$

This indeed holds by the tightness of $\{f^{\tilde{\sigma}} : \tilde{\sigma} \in \Pi^S\}$: by picking the appropriate $N(\epsilon)$ as in Lemma A.3.3, for all $N \geq N(\epsilon)$, we have, for every $\sigma_n \in \Pi^S$,

$$\begin{aligned}\sum_{j=N+1}^{\infty} \pi^{\sigma_n}(m,j) &< \sum_{m=0}^{\infty}\sum_{j=N+1}^{\infty} \pi^{\sigma_n}(m,j) + \sum_{m=N+1}^{\infty}\sum_{j=0}^{N} \pi^{\sigma_n}(m,j) \\ &= 1 - \sum_{m=0}^{N}\sum_{j=0}^{N} \pi^{\sigma_n}(i,j) \\ &< 1 - (1-\epsilon) = \epsilon,\end{aligned}$$

completing the proof. ∎

BIBLIOGRAPHY

[1] Hyun-soo Ahn, Izak Duenyas, and Mark E. Lewis. Optimal control of a two-stage tandem queuing system with flexible servers. *Probability in the Engineering and Informational Sciences*, 16(4):453469, 2002.

[2] Hyun-Soo Ahn and Mark E Lewis. Flexible Server Allocation and Customer Routing Policies for Two Parallel Queues When Service Rates Are Not Additive. *Operations Research*, 61(2):344–358, April 2013.

[3] Eitan Altman. *Constrained Markov Decision Processes*. Chapman and Hall/CRC, 1999.

[4] Sigrún Andradóttir and Hayriye Ayhan. Throughput maximization for tandem lines with two stations and flexible servers. *Operations Research*, 53(3):516–531, 2005.

[5] Sigrún Andradóttir, Hayriye Ayhan, and Douglas G Down. Server assignment policies for maximizing the steady-state throughput of finite queueing systems. *Management Science*, 47(10):1421–1439, 2001.

[6] A. Arapostathis, A. Biswas, and G. Pang. Ergodic control of multi-class m/m/n+m queues in the halfin-whitt regime. *Ann. Appl. Probab.*, 25:3511–3570, 2015.

[7] N. T. Argon, S. Ziya, and R. Righter. Scheduling impatient jobs in a clearing system with insights on patient triage in mass casualty incidents. *Probability in The Engineering and Informational Sciences*, 22:301–332, 2008.

[8] M. Armony and C. Maglaras. Contact centers with a call-back option and real-time delay information. *Operations Research*, 52:527–545, 2004.

[9] M. Armony and C. Maglaras. On customer contact centers with a call-back option: Customer decisions, routing rules and system design. *Operations Research*, 52:271–292, 2004.

[10] M. Armony, N. Shimkin, and W. Whitt. The impact of delay announcements in many-server queues with abandonment. *Operations Research*, 57:66–81, 2009.

[11] Rami Atar, Giat Chanit, and Nahum Shimkin. The $c\mu/\theta$ rule for many-server queues with abandonment. *Operations Research*, Sept-Oct, 2010.

[12] U. Ayesta, P. Jacko, and V. Novak. A nearly-optimal index rule for scheduling of users with abandonment. *Proceedings of IEEE INFOCOM*, 2011.

[13] S.L. Bell and R.J. Williams. Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: Asymptotic optimality of a threshold policy. *Annals of Applied Probability*, 11(3):608–649, 2001.

[14] O. Berman, J. Wang, and K. P. Sapna. Optimal management of cross-trained workers in services with negligible switching costs. *Eur. J. Oper. Res.*, 167:349–369, 2005.

[15] Sandjai Bhulai and Ger Koole. A queueing model for call blending in call centers. *IEEE Transactions on Automatic Control*, 48:1434–1438, 2000.

[16] C. Buyukkoc, P. Varaiya, and J. Walrand. The $c\mu$-rule revisited. *Advances in Applied Probability*, 17(1):237–238, 1985.

[17] Center for Disease Control. National hospital ambulatory medical care survey: 2012 emergency department summary tables. `https://www.cdc.gov/nchs/data/ahcd/nhamcs_emergency/2012_ed_web_tables.pdf`, 2012. Table 2.

[18] J. Dai and S. He. Many-server queues iwth customer abandonment: A survey of diffusion and fluid approximations. *Journal of Systems Science and Systems Engineering*, 21:1–36, 2012.

[19] Hyun-soo Ahn David L. Kaufman and Mark E. Lewis. On the introduction of an agile, temporary workforce into a tandem queueing system. *Queueing Systems: Theory and Applications*, 51(1-2):135–171, 2005.

[20] Douglas G. Down, Ger Koole, and Mark E. Lewis. Dynamic control of a single server system with abandonments. *Queueing Systems: Theory and Applications*, 67, January 2011.

[21] I Duenyas, D. Gupta, and T. L. Olsen. Control of a single-server tandem queueing system with setups. *Operations Research*, 46(2):218–230, March-April 1998.

[22] Izak Duenyas, Diwakar Gupta, and Tava Lennon Olsen. Control of a single-server tandem queueing system with setups. *Operations Research*, 46(2):218–230, 1998.

[23] Eugene Feinberg. Constrained semi-markov decision processes with average rewards. *Mathematical Methods of Operations Research*, 39:257–288, 1994.

[24] Noah Gans, Ger Koole, and Avishai Mandelbaum. Telephone call centers: Tutorial, review, and research prospects. *Manufacturing & Service Operations Management*, 5(2):79–141, 2003.

[25] Noah Gans and Yong-Pin Zhou. A call-routing problem with service-level constraints. *Operations Research*, 51:255–271, 2003.

[26] S. Ghamami and A. Ward. Dynamic scheduling of an n-system with reneging. *Preprint*, 2010.

[27] S. Ghamami and A. Ward. Dynamic scheduling of a two-server parallel server sysetm with complete resource pooling and renegin in heavy traffic: asymptotic optimality of a two-threshold policy. *Mathematics of Operations Research*, 38:761–824, 2013.

[28] J. Harrison and A. Zeevi. Dynamic scheduling of a multiclass queue in the halfin and whitt heavy traffic regime. *Operations Research*, 52:243–257, 2004.

[29] Junfei Huang, Boaz Carmeli, and Avishai Mandelbaum. Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. *Operations Research*, 63(4):892–908, 2015.

[30] Seyed MR Iravani, Morton J. M. Posner, and John A Buzacott. A two-stage tandem queue attended by a moving server with holding and switching costs. *Queueing Systems*, 26(3-4):203–228, 1997.

[31] SMR Iravani, MJM Posner, and JA Buzacott. An n-stage tandem queueing system attended by a moving server with holding and switching costs. Technical report, Tech. Rep. 96-09, Department of Industrial Engineering, University of Toronto, 1996.

[32] Tara Javidi, Nah-Oak Song, and Demosthenis Teneketzis. Expected makespan minimization on identical machines in two interconnected queues. *Probability in the Engineering and Informational Sciences*, 15(4):409–443, 2001.

[33] Pravin K Johri and Michael N Kateiiakis. Scheduling service in tandem queues attended by a single server. *Stochastic Analysis and Applications*, 6(3):279–288, 1988.

[34] Vidyadhar G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Texts in Statistical Science. Chapman & Hall, Boca Raton, second edition, 2010.

[35] S. Lippman. Applying a new device in the optimization of exponential queueing systems. *Operations Research*, 23:687–710, 1975.

[36] P. Nain. Interchange arguments for classical scheduling problems in queues. *Systems and Control Letters*, 12:177–184, 1989.

[37] Rosser T Nelson. Labor assignment as a dynamic control problem. *Operations Research*, 14(3):369–376, 1966.

[38] MARK P Van OYEN, Esma GS Gel, and Wallace J Hopp. Performance opportunity for workforce agility in collaborative and noncollaborative work systems. *Iie Transactions*, 33(9):761–777, 2001.

[39] Martin Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. John Wiley & Sons, Inc., 1994.

[40] Soroush Saghafian, Wallace J Hopp, Mark P Van Oyen, Jeffrey S Desmond, and Steven L Kronick. Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research*, 60(5):1080–1097, 2012.

[41] Alexandre Salch, J-P Gayon, and Pierre Lemaire. Optimal static priority rules for stochastic scheduling with impatience. *Operations Research Letters*, 41(1):81–85, 2013.

[42] Linn I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley Series In Probablllty And Statistics. John Wiley & Sons, Inc, New York, NY, 1999.

[43] R. Serfozo. An equivalence between continuous and discrete time markov decision processes. *Operations Research*, 27:616–620, 1978.

[44] J George Shanthikumar and David D Yao. Multiclass queueing systems: Polymatroidal structure and optimal scheduling control. *Operations Research*, 40(3-supplement-2):S293–S299, 1992.

[45] O. A. Soremekun, R. Capp, P. D. Biddinger, B. A. White, Y. Chang, S. B. Carignan, and D. F. Brown. Impact of physician screening in the emergency

department on patient flow. *The Journal of Emergency Medicine*, 43(3):509 515, September 2012.

[46] F Subash, F Dunn, B McNicholl, and J Marlow. Team triage improves emergency department efficiency. *Emergency Medicine Journal*, 21(5):542–544, 2004.

[47] T. Tezcan and J. Dai. Dynamic control of n-systems with many servers, asymptotic optimality of a static priority policy in heavy traffic. *Operations Research*, 58:94–110, 2010.

[48] A. Ward. Asymptotic analysis of queueing systems with reneging: A survey of results for fifo, single class models. *Oper. Res. Management Sci.*, 16:1–14, 2011.

[49] A. R. Ward and P. W. Glynn. A diffusion approximation for a markovian queue with reneging. *Queueing Systems*, 43:103–128, 2003.

[50] A. R. Ward and P. W. Glynn. A diffusion approximation for a $gi/gi/1$ queue with balking or reneging. *Queueing Systems*, 50:371–400, 2005.

[51] Cheng-Hung Wu, Douglas G. Down, and Mark E. Lewis. Heuristics for allocation of reconfigurable resources in a serial line with reliability considerations. *IIE Transactions*, 40(6):595–611, June 2008.

[52] R. Yang, S. Bhulai, and R. Mei. Optimal resource allocation for multiqueue systems with a shared server pool. *Queueing Systems*, 68:133–163, 2011.

[53] Gabriel Zayas-Cabán, Jingui Xie, Linda V. Green, and Mark E. Lewis. Dynamic control of a tandem system with abandonments. *Queueing Systems: Theory and Applications*, 84(3):279–293, December 2016.

[54] Gabriel Zayas-Cabán, Jingui Xie, Linda V Green, and Mark E Lewis. Policies for physcian allocation to triage and treatment in emergencey departments. *Working paper*, 2017.