Dynamic Scheduling of a Multiclass Queue in the Halfin–Whitt Heavy Traffic Regime

J. Michael Harrison*
Stanford University

Assaf Zeevi†
Columbia University

To appear: *Operations Research*
(Last revised: March, 2003.)

Abstract

We consider a Markovian model of a multiclass queueing system in which a single large pool of servers attends to the various customer classes. Customers waiting to be served may abandon the queue, and there is a cost penalty associated with such abandonments. Service rates, abandonment rates and abandonment penalties are generally different for the different classes. The problem studied is that of dynamically scheduling the various classes. We consider the Halfin-Whitt heavy traffic regime, where the total arrival rate and the number of servers both become large in such a way that the system's traffic intensity parameter approaches one. An approximating diffusion control problem is described and justified as a purely formal (i.e., non rigorous) heavy traffic limit. The Hamilton-Jacobi-Bellman equation associated with the limiting diffusion control problem is shown to have a smooth (classical) solution, and optimal controls are shown to have an extremal or "bang-bang" character. Several useful qualitative insights are derived from the mathematical analysis, including a "square root rule" for sizing large systems and a sharp contrast between system behavior in the Halfin-Whitt regime versus that observed in the "conventional" heavy traffic regime. The latter phenomenon is illustrated by means of a numerical example having two customer classes.

Short Title: Dynamic Scheduling of a Multiclass Queue

Keywords: Scheduling, queueing, diffusion approximations, many server limits, Halfin-Whitt regime, stochastic control, numerical methods.

2000 AMS Subject Classification (Primary): 60K25, 93E20, 90B22

*Stanford Graduate School of Business, e-mail: `harrison_michael@gsb.stanford.edu`
†Columbia Graduate School of Business, e-mail: `assaf@gsb.columbia.edu`

1 Introduction

The paper is concerned with dynamic control of a multi-class Markovian service system in which one pool of servers attends to several customer classes. A schematic description is given in Figure 1. There are m customer classes indexed by $i = 1, 2, \dots, m$ and a total of N servers staffing the system, with identical capabilities that will be described below. Customers of each class arrive from outside the system at rates λ_i and require a single service before they depart. The open-ended rectangles in Figure 1 represent (infinite capacity) buffers in which customers of the various classes reside as they await service, and we allow the possibility that customers abandon the system if queueing delays are excessive. Abandonment is represented by the horizontal arrows in Figure 1, with associated abandonment rate γ_i (per customer) for class i, and the circle represents the server pool. Precise details of the arrival processes and abandonments will be provided in Section 2.

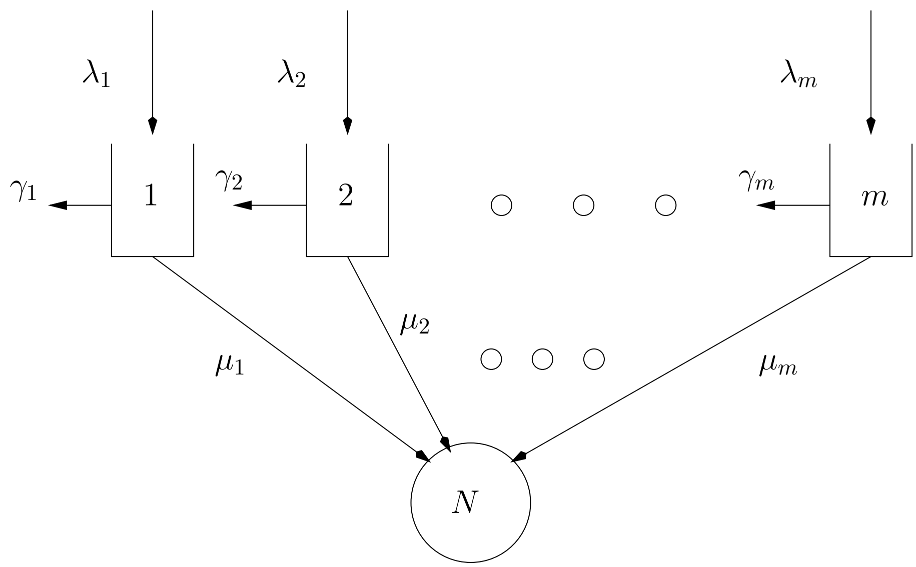

Figure 1: A schematic model of the system

Servers are capable of processing customers from any given class, and the service rate μ_i depends on the class being processed. Because we assume throughout that customers are homogeneous within a given class, there is no loss of generality in assuming that they are served in the order they arrive (FIFO). Thus there are m different processing activities available to the system manager, each of which corresponds to a particular customer class being processed by a server. We assume that customer arrivals are uncontrollable, and with regard to congestion-related costs, we allow both *abandonment penalties* and *linear holding costs*. That is, we take as given parameters $p_i > 0$ and $h_i \geq 0$ for each class i, assuming that a penalty of p_i is incurred each time a class i customer abandons, and that holding costs are continuously incurred at rate h_i for each class

i customer who is waiting in the queue (but not for those being served). Roughly speaking, the system manager wants to allocate servers to waiting customers in such a way that congestion costs are minimized.

The main motivating example of such a queueing system is a telephone call center, where a server pool (typically there are several such) consists of agents or operators that have a particular combination of training and experience, and the customer classes correspond to different caller streams that are distinguishable from one another upon arrival. The sorting of callers into different classes may be accomplished by any of several means, including responses given to an interactive voice response unit. In the context of call centers, and of customer contact centers more broadly, abandonment rates are commonly cited as key measures of system performance, and short waits in queue are obviously preferred to long ones, regardless of whether customers abandon; this serves to motivate the cost structure assumed above.

In the call center context, the problem of dynamically allocating customers to server pools is called *skills-based routing*, because the classes that agents in a given pool can process, and the average rates at which they can process them, are reflective of those agents' skills. The model that we study here is suitable in situations where there is only one such agent pool. Although this is rarely the case in most call centers, it is certainly an important first step in studying the skills-based routing problem. Moreover, as we hope to demonstrate here, it is quite an interesting problem in its own right.

Our analysis is mainly inspired by the work of Halfin and Whitt [13], and its extension in a recent paper by Garnett, Mandelbaum and Reiman [12], henceforth referred to as GMR. GMR consider a model that is similar to the one depicted in Figure 1, with just one customer class and one server pool. This is considered as a simple operating model of a call center with one class of customers and a single agent pool. In their model, as in [13], there are no scheduling decisions to be made, but they allow customer abandonment. Assuming a Markovian model (that is, service times, inter-arrival times and abandonment times are all exponentially distributed), they consider the asymptotic regime where the number of servers N becomes large. Assuming that the customer arrival rate is approximately balanced with the total processing capacity of the server pool, GMR [12, Theorem 2] show that a normalized version of the queue-length process in their single-class model is well approximated, as N becomes large, by a certain one-dimensional diffusion process. This extends the pioneering work of Halfin and Whitt [13] on heavy traffic analysis of many-server queueing systems (see also [29]), the extension being to incorporate customer abandonment. Moreover, GMR show that this asymptotic analysis is useful in deriving certain rules of thumb for system staffing that strike a good balance between abandonments and delays. In particular, the analysis via diffusion approximations supports the well known "square-root rule" for safety staffing. The stationary distribution of the GMR Markovian model is known "explicitly" (the associated queue length process is a birth-and-death process), but their diffusion approximation is still useful because it yields such ready insights about the relationship between system staffing and system performance.

The Halfin-Whitt regime has been studied recently in various contexts. Fleming et. al. [10] use the Halfin-Whitt approximation in analyzing a wireless access network, while Fleming and Simon [9] discuss load balancing and CDMA communications applications. In a recent paper, Das and Srikant [8] develop approximations for performance of a congested communication link in a packet-switched data network. Some extensions of the basic Halfin-Whitt model taking into account phase type service distributions and static priorities have been studied recently by Puhalskii and Reiman [24]. Recent work by Atar et. al. [1] and [2] also considers dynamic control in the Halfin-Whitt regime; [2] which is the more recent of the two starts from an "exact" problem formulation which is similar to the one in our paper. They also focus on a formal derivation of a diffusion control problem, however they consider more general cost structures and derive asymptotic optimality results for the controls derived from the (formal) limiting control problem.

Unlike the "conventional" heavy traffic parameter regime considered in [15], [16], [30] and other recent papers, the Halfin-Whitt regime is one where high utilization goes hand-in-hand with "small" waiting time. That is, high-quality service can be achieved together with high resource utilization in the Halfin-Whitt regime. Whitt [29] has argued that this regime is the "right" one to consider, for purposes of system design, in many large scale service operations, and the GMR [12] extension to allow customer abandonment is crucial for modeling call centers.

What we develop in this paper is a multi-class version of the GMR diffusion model, in the context of which one can undertake an approximate analysis of the multi-class scheduling problem. The main contributions are the following.

1. With regard to system modeling, we describe a multi-class generalization of the GMR model, one that incorporates a dynamic control capability, and justify it by means of a purely formal heavy traffic limit.

2. With regard to mathematical analysis, we assume an infinite-horizon discounted cost criterion for our diffusion control problem and prove that the associated Hamilton-Jacobi-Bellman (HJB) equation admits a smooth (classical) solution, which is the value function (Theorem 1). Moreover, it is shown that there exists an optimal control policy having an extremal or "bang-bang" character: in each state the controller's optimal action is to hold customers of at most one class in the queue, but the class that is so distinguished is generally different in different states (Theorem 2).

3. With regard to qualitative insights, there are two findings worthy of note. First, in the Halfin-Whitt heavy traffic regime where N and the total arrival rate become large simultaneously, the amount of "excess capacity" required to achieve any given level of performance is of order $\sqrt{N}$, as Whitt [29] and GMR [12] have observed for single-class models without dynamic control; this less-than-proportional growth in required capacity is consistent with the general finding that stochastic service systems are characterized by economies of scale. Second, the character of optimal controls and associated system behavior is very different in the Halfin-Whitt heavy traffic regime from what one sees in the "conventional" heavy traffic regime. In particular, one does not see any degree of "asymptotic state space collapse" in the Halfin-Whitt regime, whereas such model simplification is the hallmark of analogous control problems obtained as conventional heavy traffic limits by Harrison and Lopez [16] and Kelly and Laws [18] among others. (For more on dynamic scheduling in the conventional heavy traffic regime see the recent paper by Williams [30].) This matter is discussed in Section 6 and illustrated by means of a numerical example in Section 7, where one sees that solutions of control problems in the Halfin-Whitt limiting regime are not simple priority rules of the kind occurring in "conventional" limiting control problems, such as the classical $c\mu$ rule or the generalized $c\mu$ rule investigated by Van Mieghem [27].

The paper is structured as follows. Section 2 describes our "conventional" stochastic scheduling problem with abandonment, as opposed to the approximating diffusion model emphasized later. Section 3 discusses the natural parameter regime for diffusion approximations proposed by Halfin and Whitt [13]. The proposed approximate diffusion control problem is formulated in Section 4, which contains the main modeling contributions of the paper. We do not attempt to justify this approximation by a rigorous limit theorem, but the limit theory developed by Halfin and Whitt [13] and by GMR [12] in their more restrictive settings provide indirect support for our more complex, multi-dimensional diffusion model. Section 5 contains the statements of the main analytical results: the study of the HJB equation and the characterization of the optimal control policy. In section 6 we discuss the qualitative implications of that analysis and some insights that it provides. We do not attempt to develop general numerical methods for solving our diffusion control problem, but in Section 7 we undertake a numerical analysis of a simple example with two classes that illustrates the form of the optimal policy. Finally, there are two appendices, one containing proofs of the main results described above, and another containing proofs of auxiliary results.

2 Conventional Model Formulation

Customers of classes $1, \ldots, m$ arrive according to independent Poisson processes, the arrival rate for class i being $\lambda_i > 0$. Also, each class i customer is characterized by an exponentially distributed service time random variable with mean $1/\mu_i$ and an exponentially distributed "patience" random variable with mean $1/\gamma_i$: a customer departs when either the total time it has spent "in service" accumulates to equal the former random variable (the departure is then called a *service completion*), or else the total time spent waiting accumulates to the latter random variable (the departure is then called an *abandonment*). Thus we call μ_i and γ_i the *service rate* and *abandonment rate*, respectively, for class i. It is assumed that $\mu_i > 0$ and $\gamma_i \geq 0$ for all $i = 1, \ldots, m$. (Thus, we allow the case with no abandonment for some or even all classes.) The service time random variable and "patience" random variable for a customer are independent of one another, independent of those for other customers, and independent of the arrival process.

The assumption that a customer's decision to abandon is independent of the state of the system is driven by the main motivating example we have in mind, namely, a telephone call center. In these service operations queues are "invisible", that is, customers waiting in the queue have no access to state information. The assumption that customer "patience" is exponentially distributed has only partial support in real call center data, but it greatly simplifies the analysis. (The recent work by Puhalskii and Reiman [24] illustrates some implications of relaxing this assumption.)

In the interest of tractability, we assume that the service of any customer can be interrupted at any time and resumed later (perhaps by another server) without penalty or loss of efficiency. This assumption, of course, is unreasonable in most settings, however, one feels intuitively that optimal policies would have similar structure with and without preemptions. Moreover, it is plausible that the difference between the two problems with and without preemption should be negligible in the heavy-traffic limit, and in fact such a result is proved in the recent work of Atar et. al. [2]. Given this assumption and the memoryless character of our arrival processes, service processes and abandonment processes, the state of system at time t is adequately summarized by the vector $Q(t) = (Q_1(t), \ldots, Q_m(t))$, where $Q_i(t)$ is the number of class i customers then present in the system, either waiting or being served. We denote by $S := \mathbb{Z}_+^m$ the state space of the process $Q = (Q(t) : t \geq 0)$, and by q a generic point in S. For each state $q \in S$ we define an associated *action set* $\mathbb{A}(q)$ as follows:

$$\mathbb{A}(q) = \left\{ a \in \mathbb{Z}_+^m : \ a \leq q \quad \text{and} \quad e \cdot a = (e \cdot q) \wedge N \right\} \quad , \tag{1}$$

where e is the m-vector of ones, $x \wedge y := \min\{x, y\}$, and $x \cdot y$ denotes the usual scalar product for vectors $x, y \in \mathbb{R}^m$. The ith component of the vector a in (1) is interpreted as the number of servers assigned to process class i customers when the system is in state q; the two inequalities appearing in (1) express the following constraints. First, one cannot assign to class i more servers than there are customers of this class present in the system. Second, the total numbers of servers assigned cannot exceed N, nor do we allow "idling" of servers when work is present in the system.

If the system is in state q at some point in time and action $a \in \mathbb{A}(q)$ is selected, then we have a_i

customers of class i "in service," and $q_i - a_i$ customers of that class waiting. Thus, the probability of a class i service completion in the next t time units is $a_i\mu_i t + o(t)$ for small t, where $f(t) = o(t)$ if $f(t)/t \to 0$ as $t \to 0$. The corresponding probability of class i abandonment is $(q_i - a_i)\gamma_i t + o(t)$. Of course, the probability of a class i arrival is $\lambda_i t + o(t)$ for small t. These transition intensities completely specify the probabilistic structure of our dynamic control problem.

With respect to the problem's economic structure, we shall assume an infinite-horizon discounted cost criterion, the interest rate for discounting being $\alpha > 0$. Given that objective, and the stationary character of our problem data, attention will be restricted to stationary Markov policies. That is, a *policy* is defined as a function $\pi : S \mapsto \mathbb{Z}_+^m$ such that $\pi(q) \in \mathbb{A}(q)$ for all $q \in S$; in the obvious way, one interprets $\pi(q)$ as the action to be taken when the system is in state q. We denote by Π the set of such stationary, Markov policies. Given the transition intensities described above, we have the following relationships for any initial state $Q(0) = q \in S$ and any policy $\pi \in \Pi$: if $\pi(q) = a$, then for small $t > 0$,

$$\mathbb{E}\left[Q_i(t) - q_i\right] = \left[\lambda_i - \gamma_i(q_i - a_i) - \mu_i a_i\right] t + o(t) \quad , \tag{2}$$

and

$$\mathbb{E}\left[Q_i(t) - q_i\right]^2 = \left[\lambda_i + \gamma_i(q_i - a_i) + \mu_i a_i\right] t + o(t) \quad , \tag{3}$$

whereas

$$\mathbb{E}\left\{\left[Q_i(t) - q_i\right]\left[Q_j(t) - q_j\right]\right\} = o(t) \quad \text{for } i \neq j \quad . \tag{4}$$

Recall from Section 1 that h_i and p_i are the holding cost rate and abandonment penalty, respectively, for class i. To avoid trivialities, it is assumed throughout that

$$c_i := h_i + \gamma_i p_i > 0 \quad \text{for all } \; i = 1, \ldots, m.$$

(If c_i were zero, the system manager would have no motivation to ever serve class i, and so that class would be dropped from the model.) Abusing notation somewhat, we shall use the letter c to denote both the vector $(c_1, \ldots, c_m)$ and the *instantaneous expected cost rate function*

$$c(q, a) = c \cdot (q - a) \quad . \tag{5}$$

From the discussion of transition intensities above, one sees that if state q is observed and the action $a \in \mathbb{A}(q)$ is selected, then the expected cost incurred during the next t time units is $c(q, a)t + o(t)$ for small $t > 0$.

With the instantaneous expected cost rate defined above, the expected present value of the total future costs under policy π, given an initial state q, is

$$J(q, \pi) := \mathbb{E}_q^\pi \left\{ \int_0^\infty e^{-\alpha t} c(Q(t), \pi(Q(t))) dt \right\} \quad , \tag{6}$$

where $\mathbb{E}_q^\pi\{\cdot\}$ denotes the expectation with respect to the probability distribution on the path space of Q that corresponds to initial state q and control policy π. Hereafter the function $J(\cdot,\pi) : S \mapsto \mathbb{R}$ will be called the *cost function* under policy π.

Now set

$$V(q) = \inf_{\pi \in \Pi} J(q,\pi) \quad \text{for} \quad q \in S \quad . \tag{7}$$

We call $V(\cdot)$ the system manager's optimal value function, or simply *value function*; a policy π is said to be *optimal* if it achieves the infimum in (7) for each $q \in S$.

It will be convenient for future purposes to define a small amount of additional notation. First, let

$$z_i = \frac{(\lambda_i/\mu_i)}{\sum_{j=1}^m (\lambda_j/\mu_j)} \qquad \text{for } i = 1,\ldots,m. \tag{8}$$

The sum appearing as the denominator on the right side of (8) represents the total workload input rate to our server pool (that is, average time units of server work arriving per time unit), so obviously z_i is the fraction of that input attributed to class i; hereafter $z = (z_1,\ldots,z_m)$ will be called the vector of *relative workload contributions*. Also, we define as usual the system's *traffic intensity parameter*

$$\rho = N^{-1} \sum_{j=1}^m (\lambda_j/\mu_j) \quad . \tag{9}$$

The problem described in this section is a continuous-time Markov decision process (MDP), and as such is amenable to the method of dynamic programming; Bertsekas [5] provides a general account of the mathematical theory associated with such problems, and Sennott [26] provides a more specialized account focused on queueing models. At least in theory, our dynamic control problem can be solved using methods described there, but the size of the problem and its fine-grained structure (e.g., discreteness of the state space) make that task quite complicated; those factors cause computational difficulty on the one hand, and overly-detailed characterization of the optimal policies (assuming that such a characterization is even possible) on the other. Our analysis will not pursue a "direct attack" on the above formulation. Rather, we propose an approximate analysis of the scheduling problem to obtain more insight.

3 The Halfin-Whitt Asymptotic Regime

Let us consider now a sequence of models, each having the structure described in section 2, indexed by $N = 1, 2, \ldots$, attaching a superscript N to the notation established previously in order to indicate the dependence of a parameter or process on N. (Accordingly, the absence of such a superscript shows that the quantity in question is independent of N.) For maximum simplicity, we shall vary only one other model parameter as the number of servers N increases, that being the total arrival rate; following the example set by Halfin and Whitt [13], we want to do this in such a way that $\sqrt{N}(1-\rho^N) \to \theta$ as $N \to \infty$, where θ is a real valued constant representing the server pool's "excess capacity" in a suitable asymptotic sense. (If θ is negative, then its absolute value represents a capacity *shortfall*, in a suitable asymptotic sense.) To minimize formulational complexity, we shall construct our sequence in such a way that $\sqrt{N}(1-\rho^N)$ is simply *equal* to a fixed constant θ for all N, or equivalently,

$$\rho^N = 1 - \theta/\sqrt{N} \quad \text{for all } N = 1, 2, \ldots. \tag{10}$$

Given that both the average service rates $\mu_1, \ldots, \mu_m$ and the relative workload contributions $z_1, \ldots, z_m$ are to be held fixed, one can substitute in (10) the definition (9) of ρ so as to obtain the following explicit formula for the arrival rates of each of the customer classes:

$$\lambda_i^N = z_i \mu_i \left(N - \theta\sqrt{N}\right) \quad \text{for } i = 1, \ldots, m \text{ and } N = 1, 2, \ldots \,. \tag{11}$$

It may appear at first glance that a more general analysis would be obtained by allowing a different parameter θ_i for each class i in (11), but as we shall explain in Section 6, the added generality is in fact illusory, and characterizing the asymptotic regime by means of a single excess capacity parameter θ facilitates interpretation and application of the analysis.

Our "heavy-traffic" assumption (10) says that as N gets large, the total workload input equals the total capacity N of the server pool plus a perturbation of order $\sqrt{N}$. To make connection with the analysis of Halfin and Whitt [13] and GMR [12] it is useful to imagine that each class $i = 1, \ldots, m$ is statically allocated $N_i = z_i N$ servers for its exclusive use (that is, servers are dedicated to the various classes in numbers proportional to those classes' relative workload inputs). With this understanding, let a sequence of normalized vector processes $X^N = (X_1^N, \ldots, X_m^N)$ be defined as follows:

$$X_i^N(t) := \frac{Q_i^N(t) - z_i N}{\sqrt{N}} \tag{12}$$

for all $t \geq 0$, $i = 1, \ldots, m$ and $N = 1, 2, \ldots$. The following result is due to GMR [12, Theorem 2], extending the earlier analysis by Halfin and Whitt [13, Theorem 2] for systems without abandonments; see also [10]. Here and later, '$\Rightarrow$' denotes weak convergence in the space $D[0,\infty)$, or the associated product space $D^m[0,\infty)$, endowed with the usual Skorohod topology. Also, readers should recall the definition (8) of z.

Proposition 1 *If $X^N(0) \Rightarrow \xi \in \mathbb{R}^m$, then $X^N \Rightarrow X$, where the limit $X = (X_1, \ldots, X_m)$ is an m-dimensional diffusion process with independent components. Specifically, X is the unique strong solution of the following stochastic differential equation:*

$$\begin{aligned} dX(t) &= b(X(t))dt + \Sigma dW(t) \quad , \\ X(0) &= \xi \quad , \end{aligned}$$

where $W = (W(t) : t \geq 0)$ *is standard Brownian motion in* $\mathbb{R}^m$*, the infinitesimal drift function* $b_i(\cdot)$ *for the* i*th component is*

$$b_i(x) = \begin{cases} -\mu_i x_i - \mu_i \theta & x_i \leq 0 \\ -\gamma_i x_i - \mu_i \theta & x_i > 0 \end{cases} ,$$

$\Sigma := \mathrm{diag}(\sigma_1, \ldots, \sigma_m)$*, and* $\sigma_i^2 = 2\mu_i z_i$*.*

Each component of the limiting diffusion X in Proposition 1 is essentially obtained by "pasting together" two Ornstein-Uhlenbeck processes; for further details on diffusions with piecewise linear drift coefficients the reader is referred to Browne and Whitt [7]. The fact that the properly centered and scaled occupancy process X_i^N has a weak limit as stated in Proposition 1 has many important consequences, and quite a bit of insight can be gleaned from it (for more details, we refer the reader to [29] and to the more recent contributions of Fleming et. al. [10], and GMR [12]). One important observation is that, in the many-server heavy-traffic regime identified by Halfin and Whitt, the number of idle servers is either zero or else a positive quantity of order $\sqrt{N}$, and the same statement applies to the total number of customers waiting for service. It follows (cf. [12]) that both average waiting times and abandonment rates are of order $1/\sqrt{N}$ in the Halfin-Whitt regime, so large service systems can be designed to deliver both high quality service and high server utilization. This is one of the main powerful messages delivered in Whitt's paper [29], and further pursued by GMR [12].

4 The Diffusion Control Model

The Halfin-Whitt scaling in (12), and the resulting diffusion limit appearing in Proposition 1, serve to motivate the diffusion approximation proposed in this section for our "original" control problem (see Section 2). In Section 6 we shall state as conjectures two heavy-traffic limit theorems which, if true, would provide a rigorous justification for the proposed diffusion approximations, but the argument provided here is purely formal.

To form a diffusion approximation for the original control problem, we first define the *normalized* state and control vectors

$$X^N(t) := \frac{Q^N(t) - zN}{\sqrt{N}}$$

and

$$U^N(t) := \frac{A^N(t) - zN}{\sqrt{N}} ,$$

where z is given in (8), and $A^N(t)$ denotes the *allocation* of servers under whatever policy π^N is selected for use in the Nth system; that is, $A^N(t) = \pi^N(Q^N(t))$ for all $t \geq 0$. Note that elements of the vector control U^N express dynamic server allocation as (scaled) deviations from the *nominal server allocations* $z_1 N, \ldots, z_m N$ that were fixed in Proposition 1.

To motivate the proposed diffusion approximation, let us consider a fixed N and a fixed but arbitrary policy π^N for the Nth system, writing $A^N(t) = \pi^N(Q^N(t))$ as above. Let us also fix an initial state $q := Q^N(0)$ and set $a := \pi^N(q) = A^N(0)$, then define $x_i := X_i^N(0) = (q_i - N z_i)/\sqrt{N}$ and $u_i := U_i^N(0) = (a_i - N z_i)/\sqrt{N}$ for $i = 1, \ldots, m$. Using the identity (11), readers can then verify that (2) - (4) are equivalent to the following: for small $t > 0$ one has

$$\mathbb{E}\left[X_i^N(t) - x_i\right] = \left[-\theta\mu_i \ - \gamma_i(x_i - u_i) - \mu_i u_i\right] t + o(t) \quad , \tag{13}$$

and

$$\mathbb{E}\left[X_i^N(t) - x_i\right]^2 = N^{-1/2}\left[-\theta\mu_i + \gamma_i(x_i - u_i) + \mu_i u_i\right] t + 2\mu_i z_i t + o(t) \quad , \tag{14}$$

for $i = 1, \ldots, m$, whereas

$$\mathbb{E}\left\{\left[X_i^N(t) - x_i\right]\left[X_j^N(t) - x_j\right]\right\} = o(t) \quad \text{for } i \neq j \ . \tag{15}$$

Taking a formal limit as $N \to \infty$, we are then led to approximate our original MDP by a classical diffusion control problem in which the infinitesimal drift function depends on the observed state x and chosen control u exactly as in (13), and the infinitesimal covariance matrix is the limit as $N \to \infty$ of that computed in (14) - (15). That is, the infinitesimal drift function $b(x, u)$ is given by

$$b_i(x, u) = -\theta\mu_i - \gamma_i(x_i - u_i) - \mu_i u_i \tag{16}$$

for all $i = 1, \ldots, m$, while the infinitesimal covariance matrix is $\Sigma^2 = \mathrm{diag}(2\mu_1 z_1, \ldots, 2\mu_m z_m)$, independent of x and u.

The state space for our diffusion control problem is all of $\mathbb{R}^m$, and as above, we denote by x and u a generic state and control, respectively. To determine for each state $x \in \mathbb{R}^m$ the associated set of feasible controls $\mathbb{U}(x)$, one starts with the specification (1), then applies the same centering and scaling that define $U^N(t)$ in terms of $A^N(t)$, and finally takes a limit as $N \to \infty$ to arrive at the following:

$$\mathbb{U}(x) = \{u \in \mathbb{R}^m : \ u \leq x \quad \text{and} \quad e \cdot u = (e \cdot x) \wedge 0\} \quad \text{for all } x \in \mathbb{R}^m \ . \tag{17}$$

As in Section 2, we shall restrict attention to stationary, Markov control policies, and the letter π will be re-used in this setting to denote a generic policy. That is, in the context of our proposed diffusion approximation, a policy is defined as a measurable function $\pi : \mathbb{R}^m \mapsto \mathbb{R}^m$ such that $\pi(x) \in \mathbb{U}(x)$ for all $x \in \mathbb{R}^m$. Formalizing the conclusion reached immediately above about the appropriate "infinitesimal parameters" for the diffusion control problem, the state dynamics under an arbitrary policy π are specified in differential form as follows:

$$dX(t) = b(X(t), \pi(X(t)))dt + \Sigma dW(t) \quad , \tag{18}$$

where $W = (W(t) : t \geq 0)$ is standard Brownian motion in $\mathbb{R}^m$. Again re-using notation employed in Section 2, we denote by Π the set of all (stationary, Markov, measurable) control policies in our diffusion context.

As above, let $W = (W(t) : t \geq 0)$ be standard Brownian motion in $\mathbb{R}^m$, with $t \in [0, \infty)$ and fix a reference complete filtered probability system $(\Omega, \mathcal{F}, \{\mathcal{F}_t\}, \mathbb{P})$, such that the Brownian motion is adapted to $\mathcal{F}_t$, and the filtration satisfies the usual conditions (cf. [17]). The state process X solving (18) will be considered with respect to this probability system. Due to the structure of the constraints in the feasible control set $\mathbb{U}(x)$, Girsanov's theorem can be applied (as in [17, Proposition 5.3.6]) even though the control set is not uniformly bounded; this yields the existence of a *weak* solution. Moreover, the law of this solution is unique; see the proof of Theorem 2. For the future purpose of proving our main contributions we would like to establish a stronger result, namely, that the solution to (18) exists in a pathwise sense. (The proof of this result is technical and is therefore given in Appendix B.)

Proposition 2 *For any initial state $X(0) = x$ and any policy $\pi \in \Pi$, the stochastic differential equation (18) admits a strong solution.*

Hereafter $\mathbb{E}_x^\pi\{\cdot\}$ denotes expectation with respect to the probability distribution on the path space of X that corresponds to the solution of (18) with initial state x and policy π. Using the same centering and scaling of the state and control processes, we can now write the dynamic optimization problem as a formal limit of the original one (6).

$$\mathcal{J}(x, \pi) = \mathbb{E}_x^\pi \left\{ \int_0^\infty e^{-\alpha t} c(X(t), \pi(X(t)))dt \right\} \quad , \tag{19}$$

with $c : \mathbb{R}^m \times \mathbb{R}^m \mapsto \mathbb{R}_+$ defined exactly as in (5). Note that $\mathcal{J}(\cdot, \pi)$ is always well defined as an extended positive real number. (As one might expect, any "reasonable" policy $\pi \in \Pi$ will render a finite cost).

The main goal is thus to solve the approximating diffusion control problem by seeking an admissible policy that minimizes (19), with the associated *value function* given by

$$V(x) = \inf_{\pi \in \Pi} \mathcal{J}(x, \pi) \quad . \tag{20}$$

5 Analytical Characterization of Optimal Controls

Following standard practice, we proceed with the analysis of the control problem by formally writing down the associated Hamilton-Jacobi-Bellman (HJB) equation The latter is a non-linear partial differential equation that should have as its solution the value function V. If the solution is sufficiently smooth then this should lead to a sharp characterization of the optimal control policy. Since the economic framework we have introduced uses an infinite horizon discounted cost criterion (19), the associated HJB equation is given by (cf. [11, §IV.5])

$$\frac{1}{2}\sum_{i=1}^{m}\sigma_i^2\frac{\partial^2 V(x_i)}{\partial x_i^2}+\mathcal{H}(x,\nabla V(x))-\alpha V(x)=0 \quad . \tag{21}$$

Here, $\mathcal{H}:\mathbb{R}^m\times\mathbb{R}^m\mapsto\mathbb{R}$ is the system *Hamiltonian*

$$\mathcal{H}(x,\delta):=\inf\left\{b(x,u)\cdot\delta+c(x,u):\ u\in\mathbb{U}(x)\right\} \quad , \tag{22}$$

where $b(x,u)$ is given by (16).

Unlike many stochastic control problems, our formulation poses several potential difficulties in that: (i) the controls have pathwise state constraints; (ii) the feasible control set is not compact or even bounded; (iii) the HJB equation is solved over all of $\mathbb{R}^m$ with no boundary conditions; (iv) the cost function is not bounded. Consequently, more generic results, surveyed for example in [11], [4], and [21] do not cover this case.

Before stating the main result we need the following definitions. Let $C^2(\mathbb{R}^m)$ denote the class of functions which are twice continuously differentiable over $\mathbb{R}^m$, and set

$$\mathcal{C}^2:=\left\{f\in C^2(\mathbb{R}^m)\ :\ f\geq 0\ \text{ and }\ \sup_{x\in\mathbb{R}^m}f(x)/(1+\|x\|)<\infty\right\} \quad .$$

Note that if $V\in\mathcal{C}^2$, it is non-negative and can exhibit at most linear growth. The following are our main results.

Theorem 1 *The Hamilton-Jacobi-Bellman equation (21) has a unique solution in $\mathcal{C}^2$, and that solution is the value function defined by (20).*

Theorem 2 *Let $V\in\mathcal{C}^2$ be as in Theorem 1, and define $\delta(x)=\nabla V(x)$, $x\in\mathbb{R}^m$. For each $x\in\mathbb{R}^m$ let $i^*(x)$ be the largest element $i\in\{1,\ldots,m\}$ that minimizes $c_i+(\mu_i-\gamma_i)\delta_i(x)$. The following policy π^* is optimal:*

$$\pi_i^*(x)=\begin{cases} x_i-[(e\cdot x)\vee 0] & \text{if} \quad i=i^*(x)\\ x_i & \text{otherwise} \quad .\end{cases} \tag{23}$$

6 Discussion and Qualitative Insights

Construction and interpretation of the optimal control. To see that the policy π^* defined mathematically by (23) corresponds to the verbal description provided earlier in Section 1, recall that the (scaled) queue length for class i in state x is $x_i - \pi_i^*(x)$. Thus the optimal policy π^* maintains a positive queue for at most one class i, that being the "least costly" class $i^*(x)$ identified in Theorem 2.

Let us consider now the question of how to "interpret" the policy π^* in the context of a single system with N large and ρ near 1, taking as given the model data and derived quantities (like c, z, and ρ) identified in Section 2. The first step is to calculate the data for our "approximating diffusion control problem." Inverting (10), the drift parameter θ is computed as $\theta = \sqrt{N}(1-\rho)$, and then the infinitesimal drift rate function $b(x,u)$, covariance matrix Σ, and cost rate function $c(x,u)$ are as specified in Sections 2 and 3. Given this data plus the interest rate α, one can in principle solve the approximating diffusion control problem to determine the gradient function $\delta(x) = \nabla V(x)$ that appears in the specification of our optimal policy π^* (Theorem 2).

Given an observed state $q \in S$, one first determines the corresponding normalized state vector $x = (q - Nz)/\sqrt{N}$, then determines for each class i the index $\kappa_i := c_i + (\mu_i - \gamma_i)\delta_i(x)$. Numbering the classes so that $\kappa_1 \geq \ldots \geq \kappa_m$, the obvious way to implement π^* is by giving class 1 highest priority, $\ldots$, and class m lowest priority in the allocation of servers. That is, one sets $a_1 = q_1 \wedge N$, $a_2 = q_2 \wedge (N - a_1), \ldots$, and $a_m = q_m \wedge [N - (a_1 + \ldots + a_{m-1})]$.

Asymptotic optimality. Given a family of models indexed by $N = 1, 2, \ldots$ that satisfy the hypotheses of Section 3 (that is, a family of models approaching heavy traffic in the Halfin-Whitt sense), let us denote by $V^N(\cdot)$ the value function for the Nth model. Because of the simple way in which we have specified our parametric family of models, with (11) determining the average arrival rates in terms of a single parameter θ and all else fixed, our approximating diffusion control problem is the same for every N; let us denote by π^N the implementation of its optimal policy π^* for purposes of the Nth system, as described in the two paragraphs preceding this one. Also, let $J^N(q^N, \pi^N)$ denote the cost function associated with the policy π^N in the Nth system, and let $q^N = Q^N(0)$ be the sequence of initial states such that

$$x^N := \frac{1}{\sqrt{N}}\left(q^N - Nz\right) \to x \in \mathbb{R}^m \quad \text{as} \quad N \to \infty \quad .$$

In the original version of this paper we conjectured the asymptotic optimality of our proposed policy, as expressed by the following two statements:

$$\frac{1}{\sqrt{N}} V^N(x^N) \to V(x) \quad \text{as} \quad N \to \infty \quad , \tag{24}$$

where $V(\cdot)$ is the value function for the diffusion control model of Section 5, and

$$\frac{1}{\sqrt{N}}\left[V^N(x^N) - J^N(q^N,\pi^N)\right] \to 0 \quad \text{as} \quad N \to \infty \quad . \tag{25}$$

Together, (24) and (25) imply that the *percentage* deviation from optimal performance under our proposed policy vanishes as $N \to \infty$, because $V(x) > 0$. Subject to additional technical restrictions, the above conjecture has been verified in subsequent work of Atar et. al. [2].

Scaling relations and capacity decisions. Assuming that they are true, (24) and (25) extend to our dynamic control setting the important qualitative insights obtained in [13] and [12] about "scaling relationships" in the Halfin-Whitt heavy traffic regime, as follows. First, the vector Q of population sizes for the various classes varies about its "nominal value" of zN by amounts of order $\sqrt{N}$, and because instantaneous expected cost rates are proportional to the queue lengths in our formulation, the best achievable performance (in terms of expected discounted cost) is of order $\sqrt{N}$ as well. The conjectured asymptotic relationship (25) says that performance under our proposed policy differs from the best achievable performance by an amount that is $o(\sqrt{N})$.

Of course, our whole theory is predicated on the crucial "heavy traffic" assumption (10), which deserves further discussion. Taking the time unit to be hours for the sake of concreteness, the quantity $R = \lambda_1/\mu_1 + \ldots + \lambda_m/\mu_m$ represents the average amount of work, expressed in server-hours, that arrives per hour. In telecommunications it is said that R represents the *offered load* (in *Erlangs*). Alternatively, one can describe R as the *nominal capacity requirement* (in servers) to handle the incoming workload. We originally defined the traffic intensity parameter ρ as $\rho = R/N$, and assumption (10) says that $\rho = 1 - \theta/\sqrt{N}$ for each of the systems in our parametric family, where θ is a fixed constant. Combining those two relationships, readers can confirm that $N = R + \theta\sqrt{R} + o(R)$ in our family of models, which can be interpreted as follows: we are considering systems in which capacity is set using roughly the *square root rule* advocated by Whitt [29]; that is, the *excess capacity* $N - R$ is taken to be approximately proportional to $\sqrt{R}$, but nothing has been said thus far about how the constant of proportionality θ might be chosen.

In our limiting diffusion control problem (see Section 4), for any initial state x it is intuitively clear that $V(x) \downarrow 0$ as $\theta \uparrow \infty$, because for large θ the controlled diffusion process X can be very nearly confined to the non-positive orthant (where congestion costs are zero). This suggests that the value function in the "original system" (with large but finite N) can be made arbitrarily small by the capacity choice $N = R + \theta\sqrt{R}$ with θ sufficiently large. In practice, the value of θ should be chosen to optimize the tradeoff between congestion costs on the one hand and capacity costs on the other. Assuming that the cost of capacity is proportional to the number of servers employed, the analysis presented here suggests the following: if both static capacity decisions and dynamic control decisions are made optimally, then both congestion costs and the cost of "excess capacity" are asymptotically proportional to $\sqrt{R}$ as $R \to \infty$.

Dynamic control in the Halfin-Whitt regime versus "conventional" heavy-traffic. The diffusion control problem described in Section 5 is one of *drift rate control*, or "bounded control", as opposed to the "singular" diffusion control problems obtained as heavy traffic limits in papers like [16]. In any given state x, the set of available controls $\mathbb{U}(x)$ is compact, so the rate at which the system manager can drive state changes is locally bounded. In contrast, it is instructive to consider a model exactly like the one laid out in Section 2 , except that there is a single server who processes class i customers at rate $N\mu_i$ $(i = 1, \ldots, m)$. Allowing the arrival rates λ_i to increase with N according to (11) as before, and dividing all population sizes by $\sqrt{N}$ (there is no need for centering in this setting), one approaches a "conventional" heavy traffic limit without any scaling of time; the acceleration of both arrival rates and service rates by a factor of N is equivalent to the scaling of time that one usually sees in the conventional heavy traffic theory.

Even without considering in detail the limiting diffusion control problem obtained in this alternate setup, one can readily see the distinction between the conventional heavy traffic regime and the Halfin-Whitt regime. In each case there exists a total service capacity of order N, and in the conventional regime it is concentrated in a "point source," so the full service capacity can be applied to a single non-empty buffer if the system manager so desires. The effect of such a deployment is to drain the targeted buffer at a rate of order N, while other buffer contents *increase* at a similar rate, because their input processes are no longer balanced against comparable capacity allocations. Thus, the system manager in the conventional model can effectively exchange customers of one class for customers of another class (the exchange rate depends on the ratio of the two classes' average service rates) in a time span of order $1/N$; such exchanges can be effected instantaneously in the limiting control problem, and so it is characterized by a dramatic "state space collapse" [14]. In the Halfin-Whitt regime considered here, the amount of service capacity that can be directed to any one customer class is limited by the number of such customers present, so instantaneous state changes cannot be effected, even in the limit as $N \to \infty$, and the phenomenon of state space collapse is absent.

Why is there a single drift parameter θ? Let us return now to the question of what happens if we allow each class i to have its own "drift parameter" θ_i in the relationship (11) that specifies average input rates for our parametric family of models. To avoid excessive complexity, we shall again take the average service rates μ_i to be independent of N, but readers will readily see how that restriction can be relaxed in the argument that follows. Also, we take as given a vector $z = (z_1, \ldots, z_m)$ with $z_i > 0$ and $e \cdot z = 1$, interpreting its components as *limiting* relative workload contributions for the various customer classes. Now for each $N = 1, 2, \ldots$ let λ_i^N be defined by (11) with θ_i in place of θ, where $\theta_1, \ldots, \theta_m$ are arbitrary scalars. Defining the "traffic intensity parameter" ρ^N via the obvious extension of (10), it follows that $\sqrt{N}(1 - \rho^N) \to \bar{\theta}$, as $N \to \infty$, where

$$\bar{\theta} := z_1\theta_1 + \ldots + z_m\theta_m \quad .$$

This parameter $\bar{\theta}$ represents, in a suitable asymptotic sense, the *aggregate* excess capacity of our server pool, so it is precisely analogous to the single parameter θ used to construct a parametric family of models in Section 3.

Now one constructs relative workload contributions z_i^N for the Nth system by means of (8), but using λ_i^N and λ_j^N in place of λ_i and λ_j, respectively. Defining vectors z^N in the obvious way, it is easy to check that $z^N \to z$ as $N \to \infty$. In defining the normalized processes X^N and U^N (see Section 4), it is now natural to use Nz^N rather than Nz as the centering term. That is, for each system N we express population sizes and server allocations as (scaled) deviations from nominal values that reflect relative workload contributions for that particular system. After tedious, but straightforward calculations, one then arrives at exactly the same limiting diffusion control problem described in Section 4 except that, in the specification of the infinitesimal drift function (16) for an arbitrary class i, θ is replaced by $\bar{\theta}$. That is, we arrive again at a limiting diffusion control problem having the apparently special structure exhibited in Section 4; the individual parameters $\theta_1, \ldots, \theta_m$, having been accounted for in the normalization, appear in the limiting problem only through their weighted average $\bar{\theta}$.

To see this conclusion from a different perspective, suppose that we use Nz rather then Nz^N as the centering term in defining X^N and U^N (that is, we center about nominal server allocations based on *limiting* relative workload contributions). The limiting diffusion control problem eventually obtained has $-\mu_i\theta_i$ as the first term on the right side of (8), but it can be converted to the limiting problem described immediately above by means of a simple transformation of state and control variables (of course the translation involves $\theta_1, \ldots, \theta_m$).

A special symmetric case and a "greedy" heuristic. An illuminating special case of our model is that where $\mu_1 = \ldots = \mu_m = \mu$ and $\gamma_1 = \ldots = \gamma_m = \gamma$. Given our restriction to full-allocation policies (that is, policies which never have idle servers and positive queues simultaneously), the dynamics of the total population size $e \cdot Q(t)$ are then independent of the policy chosen. In the limiting diffusion control problem of Section 4 this translates as follows: defining the driftless Brownian motion $\xi(t) = e \cdot W(t)$ and the normalized total "population size" $Z(t) = e \cdot X(t)$, one has

$$dZ(t) = [-m\mu\theta - \mu Z(t) - \gamma Z(t) \vee 0]\, dt + d\xi(t) \quad ,$$

independent of the policy chosen. Given that the instantaneous cost rate is $c \cdot X(t)$, it follows that for every $x \in \mathbb{R}^m$ the designated class $i^*(x)$ in Theorem 2 is simply that class i for which c_i is minimal. In the limiting diffusion control problem, this policy is actually optimal in the *pathwise* sense, meaning that it minimizes total cost incurred up to time t for all t simultaneously (with probability* 1). In terms of the original model, this is the *greedy heuristic* that minimizes the instantaneous expected cost rate at each point in time, giving highest priority to the class i with largest c_i value, ..., and lowest priority to the class with smallest c_i value. As we shall see in the example that follows, the greedy heuristic is *not* generally optimal, at least in the limiting diffusion control problem.

7 Numerical Solution of a Two Class Model

Armed with the analytical characterization of the optimal policy developed in Section 5, the objective now is to explicitly compute it. This section discusses a simple two-class example, for which the optimal policy is calculated numerically. We make no attempt to present a full exposition of the numerical methods that are used, nor do we describe implementation details; these topics are quite interesting in their own right, but both are well beyond the scope of this paper.

Because the optimal index rule described in Theorem 2 uses the gradient of the value function, clearly the key is to calculate the latter. (All other parameters defining the optimal policy are assumed to be known.) Our starting point is the non-linear partial differential equation (21). The first task in computing the value function is to *linearize* this equation, for instance, by fixing an initial "guess" of the optimal policy. The Hamiltonian (22) is consequently reduced to a linear function of the gradient, and the resulting equation can now be solved numerically, leading to a solution V^0 for the value function. This, in turn, is used to derive a revised estimate of the optimal policy via Theorem 2, and the process repeats itself. These *policy iterations* (see, e.g., [5]) generate a sequence of policies $\{\pi^k\}$ that, under reasonable conditions, should converge to an optimal policy, that is, $\pi^k \to \pi^*$, in which case the associated sequence of solutions of (21) $\{V^k\}$ should converge to the value function V. (Precise notions of convergence will not be rigorously addressed here.) The only remaining algorithmic detail is the method by which the linearized version of (21) is solved. To this end, we use the so-called *finite element method* (FEM), which is a standard numerical tool in solving partial differential equations; for a general description of the method see, for example, [3], and for recent work involving FEM in the context of numerical solutions to dynamic control problems in queueing systems see [19]. An alternative method which is often used is the Markov chain approach described in [22].

For a numerical example, consider a system with two customer classes and the following parameter values: service rates $\mu_1 = 1$ and $\mu_2 = 1.5$; abandonment rates $\gamma_1 = 0.5$ and $\gamma_2 = 1$; holding cost rates $h_1 = 1$ and $h_2 = 1.5$; abandonment penalties $p_1 = 1$ and $p_2 = 1.5$; interest rate for discounting $\alpha = 0.1$; and asymptotic excess capacity $\theta = 1$. Given the characterization of the optimal policy in Theorem 2, we define

$$\psi(x_1, x_2) := c_2 + (\mu_2 - \gamma_2)\delta_2(x_1, x_2) - c_1 - (\mu_1 - \gamma_1)\delta_1(x_1, x_2) \ ,$$

where $\delta_i(x_1, x_2) = \nabla_i V(x_1, x_2)$ is ith component of the gradient of the value function and $c_i = h_i + p_i\gamma_i$ for $i = 1, 2$. The function ψ is the "test quantity" that determines the optimal action, as follows:

- If $x_1 + x_2 \geq 0$ and $\psi(x_1, x_2) > 0$ then class 1 gets priority; the resulting allocation is $\pi_1(x_1, x_2) = x_1$ and $\pi_2(x_1, x_2) = -x_1$.
- If $x_1 + x_2 \geq 0$ and $\psi(x_1, x_2) \leq 0$ then class 2 gets priority; the resulting allocation is $\pi_1(x_1, x_2) = -x_2$ and $\pi_2(x_1, x_2) = x_2$.
- If $x_1 + x_2 \leq 0$ the system has excess capacity and there are no server allocation decisions to make.

[PLACE FIGURES 2 & 3 HERE]

Figure 2 gives a plot of the test quantity $\psi(x_1, x_2)$ as a function of the state. Recall that the state corresponds roughly to the excess (or shortfall) of customers in the original system, relative to the nominal (fluid scale) allocation of servers. Figure 3 depicts the corresponding optimal index rule, which partitions the state space into three disjoint regions where the optimal action is prescribed. This example shows that the optimal policy derived from our diffusion control problem is generally *not* a static priority policy.

A Proof of the Main results

Proof of Theorem 1: Since the proof is somewhat lengthy and proceeds in several steps, we first sketch briefly the main ingredients. The starting point is to apply a standard truncation idea (see, e.g., [23]). This enables us to study a sequence of quasilinear PDE's with a Dirichlet boundary condition. The results given in Ladyzeskaya and Uralseva [20] describe the required regularity which the PDE should posses for solvability. Essentially, this amounts to the Hamiltonian $\mathcal{H}$ in (22) having proper Lipschitz regularity. We then take a sequence of Dirichlet problems such that the boundary condition vanishes in the limit. The unique solutions to this sequence of truncated problems, denote them by $\{V_n\}$, are smooth and moreover, we show that these functions along with their first and second derivatives constitute an equicontinuous family. Hence they are pre-compact in the class of continuous functions equipped with the topology of uniform convergence on compact sets. Consequently, we can extract a subsequence that converges uniformly on compact with the limit being the sought value function which satisfies the original HJB equation (21). We now proceed with the proof.

Step 1. We apply the aforementioned truncation argument, and consider properties of the "truncated problem". Fix $n \in \mathbb{N}$ and let $B(0,n) = \{y : \|y\| \le n\}$. Fix a policy $\pi \in \Pi$ and an initial condition $X(0) = x \in B(0,n)$. We will be considering the diffusion X which solves (18) "killed" at the boundary of $B(0,n)$. Set $T_n^\pi = \inf\{t \ge 0 : X(t) \in \partial B(0,n)\}$, where for a set S we let ∂S denote its boundary. Where no ambiguity arises, we use $T_n := T_n^\pi$ for brevity. Let

$$\mathcal{J}_n(x,\pi) := \mathbb{E}_x^\pi \int_0^{T_n} e^{-\alpha t} c(X(t), \pi(X(t)))dt$$

and set

$$V_n(x) := \inf_{\pi \in \Pi} \mathcal{J}_n(x,\pi) \quad .$$

Lemma 1 *$V_n(x)$ is the unique bounded solution in $C^2(\mathbb{R}^m)$ of*

$$\begin{aligned} &\frac{1}{2}\sum_{i=1}^m \sigma_i^2 \frac{\partial^2 W(x_i)}{\partial x_i^2} + \mathcal{H}(x, \nabla W(x)) - \alpha W(x) = 0 \\ &W(x) = 0; \qquad x \in \partial B(0,n) \quad , \end{aligned} \tag{26}$$

where $\mathcal{H}(\cdot,\cdot)$ is defined in (22).

The proof, which is based on existence results for quasilinear PDE's is deferred to Appendix B.

Fix $r > 0$, and set $B(0,r) = \{y : \|y\| \le r\}$; the ball of radius r in $\mathbb{R}^m$. Then, for all $n \ge \lfloor r \rfloor + 1$ we have by the standard interior estimates of Ladyzeskaya and Uralseva [20, p. 298-300] that $\|\nabla V_n(x)\| \le C_1$, for all $x \in B(0,r)$, where C_1 is a constant depending on r but independent of n. A similar estimate holds for $V_n(x)$, which we make explicit using the following argument. First, note that $V_n(x) \le V(x)$, and the latter can be bounded using Fubini's theorem as follows: $V(x) \le C \int_0^\infty \mathbb{E}_x^\pi[\|X(t)\|]$ for some constant C independent of n. We now appeal to Lemma 2 which asserts that $\mathbb{E}_x^\pi\|X(t)\| \le C(1+\|x\|)(1+t)$, thus we have $V(x) \le C_2(1+\|x\|)$ and this implies the uniform bound on $V_n(x)$.

Step 2. We consider a sequence of truncated problems, and their limit. The results stated so far, imply that $\{V_n\}$ and $\{\nabla V_n\}$ are bounded uniformly on compact sets, independent of n. Since V_n satisfies the HJB equation associated with the truncated problem, and the Hamiltonian is Lipschitz (as established in the proof of Lemma 1), it follows that $\{\Delta_\sigma V_n\}$ is also bounded on compact sets, independent of n. Here $\Delta_\sigma(\cdot)$ denotes the second order operator in the HJB equation, i.e., the Laplacian operator, with weights σ_i^2, $i = 1, 2, \ldots, m$. Since $\{\Delta_\sigma V_n\}$ and $\{\nabla V_n\}$ are uniformly bounded on $B(0,r)$ it follows that both V_n and ∇V_n are Hölder continuous, in the ball $B(0,r)$, uniformly in n. Again, since the Hamiltonian is Lipschitz in its arguments, and since V_n satisfies the PDE with boundary conditions, it must be that $\Delta_\sigma V_n$ is also Hölder continuous uniformly in n. Hence, the families $\{V_n\}, \{\nabla V_n\}$ and $\{\Delta_\sigma V_n\}$ are equicontinuous and bounded. Consequently, Arzela-Ascoli (see, e.g., Rudin [25]) establishes that for the former there exist convergent subsequences, denoted for brevity still by subscript n. Standard results concerning interchange of derivatives and limits establish the existence of a $V \in C^1$ such that $V_n \to V$, $\nabla V_n \to \nabla V$, and $\Delta_\sigma V_n \to \Delta_\sigma V$ uniformly on $B(0,r)$. Standard PDE arguments (cf. [20]) then give the improved smoothness of V. Now, V_n satisfies the HJB equation with boundary condition and $V_n \to V$ uniformly on $B(0,r)$. Since the Hamiltonian (22) is Lipschitz (see the proof of Lemma 1) we can "pass" the above limits "through" the truncated HJB equation to establish that V satisfies the original HJB partial differential equation equation on $B(0,r)$. Since r was arbitrary, V must satisfy the original HJB equation (21) in $\mathbb{R}^m$. Now, observe that by definition of V_n and V, monotone convergence implies

$$V_n(x) \uparrow V(x) = \inf_{\pi \in \Pi} \mathbb{E}_x^\pi \int_0^\infty e^{-\alpha t} c(X(t), \pi(X(t))) dt \quad .$$

Thus the proposed limit V is the value function of the original control problem. That V is finite for all x follows from the bound established above, namely, $V(x) \leq C(1 + \|x\|)$.

Step 3. The main task here is to to apply a verification argument for functions in the class $\mathcal{C}^2$. Fix $W \in \mathcal{C}^2$, and a policy $\pi \in \Pi$. Now, application of the Itô differential rule to $\exp(-\alpha t)W(X(t))$ gives (see, e.g., the verification proof in [11, p. 173])

$$W(x) \leq \mathcal{J}(x, \pi) + \liminf_{t \to \infty} e^{-\alpha t} \mathbb{E}_x^\pi \left[W(X(t)) \right] \quad . \tag{27}$$

The following result is proved in Appendix B. (We note that the original proof of this lemma contained an error that was pointed out to us by Rami Atar; the current proof is similar to one in [2] but was developed independently.)

Lemma 2 *For any $x \in \mathbb{R}^m$ and policy $\pi \in \Pi$*

$$\mathbb{E}_x^\pi \left[\|X(t)\| \right] \leq C(1 + \|x\|)(1 + t)$$

for some positive finite constant C.

Consequently, using $W(x) \leq C(1 + \|x\|)$ we have that the last term on the right hand side of (27) converges to zero. Thus, we have $W(x) \leq \mathcal{J}(x, \pi)$, and since $\pi \in \Pi$ was arbitrarily chosen, we have $W(x) \leq V(x)$ where V is the value function. On the other hand, the optimal policy π^* described in Theorem 2 satisfies

$$\pi^*(x) \in \operatorname{argmin} \left\{ b(x, u) \cdot \nabla W(x) + c(x, u) : \; u \in \mathbb{U}(x) \right\}$$

for all $x \in \mathbb{R}^m$, and is measurable. Thus for X solving (18) under the policy π^* we have

$$\pi^*(X(t)) \in \operatorname{argmin} \{b(X(t), u) \cdot \nabla W(X(t)) + c(X(t), u) : \ u \in \mathbb{U}(X(t))\}$$

almost surely for all t. Applying Itô's differential rule as before we have that $W(x) = \mathcal{J}(x, \pi^*)$. Thus, $W(x) \geq V(x)$, and together with the previous bound establishes that W is the value function, and π^* is an optimal policy. This concludes the proof. ■

Proof of Theorem 2: In the proof of Theorem 1 we established that $V(x) \leq \mathcal{J}(x, \pi)$ for all $\pi \in \Pi$. It was also shown there that for a solution $W \in \mathcal{C}^2$ of the HJB equation (21), and policy π^* satisfying

$$\pi^*(x) \in \operatorname{argmin} \{b(x, u) \cdot \nabla W(x) + c(x, u) : \ u \in \mathbb{U}(x)\}$$

for all $x \in \mathbb{R}^m$, we have $V(x) = \mathcal{J}(x, \pi^*)$. It is easy to check that the explicit expression for the policy given in Theorem 2 is obtained by substituting the drift function b and feasible control set $\mathbb{U}$ into the above optimization problem. The mapping defined by $\pi : \mathbb{R}^m \mapsto \mathbb{R}^m$ is clearly measurable (thus, we do not need to invoke a measurable selection theorem). Therefore, $U^*(t) = \pi^*(X(t))$ is by construction an optimal control. The existence of a weak solution of the SDE (18) under the policy π^* follows from [17, Proposition 5.3.6]. Uniqueness in law follows from an application of the Girsanov transformation as in [17, Propositions 5.3.10], given that we can show that for all $T < \infty$, $\int_0^T \|b(X(t), \pi(X(t)))\|^2 dt$ is finite, $\mathbb{P}_x$ almost surely. This, in turn, can be verified by applying Itô's lemma to the process X^2, using the crude bound $\|\pi(x)\| \leq C\|x\|$ and finally applying Gronwall's inequality. (The reader may also consult [17, Problem 5.3.15] for further details.) Thus, π^* defines an admissible stationary Markov policy, the associated (optimal) diffusion is well defined, and its law is unique which suffices for defining the value function. This concludes the proof. ■

B Technical Proofs and Auxiliary Results

Proof of Proposition 2: Since the proof is rather technical, we first describe the main idea which essentially relies on a localization argument. We first suitably truncate the drift function $b(x, u)$ in (18) to $b^n(x, u)$ which is bounded. We then consider the stochastic differential equation with drift b^n, the "truncated version" of the original problem, displayed below in (28). Existence and uniqueness of a strong solution to the truncated version (28) is due to the drift being bounded (see, e.g., [17, Proposition 5.5.17] and Veretennikov [28] for the multidimensional analogue). We then restrict attention to a region in $\mathbb{R}^m$ such that the truncated drift is equal to the original drift (i.e., $b^n = b$ for x values in that region). The solution of the original problem (18) is then given as the solution to the truncated problem up to a random "exit time" of the aforementioned region. That is, the process solving (18) is defined as the solution to the truncated problem (28), up until the time it exits the bounded domain where the truncated drift b^n is equal to the original b. Finally, we prove that this exit time occurs eventually (a.s.) after time T, for any $T < \infty$, thus proving the existence of a strong solution to the original problems (18) on $[0, \infty)$.

Step 1. For each $n \geq 1$ we define the truncated version of (18) to be

$$\begin{aligned} dX^n(t) &= b^n(X^n(t), \pi(X^n(t)))dt + \Sigma dW(t) \\ X^n(0) &= x \quad , \end{aligned} \tag{28}$$

with the drift coefficient truncated coordinate-wise as follows

$$b_i^n(x_i, u_i) = \begin{cases} b_i(x_i, u_i) & \text{if } |b_i(x_i, u_i)| \leq n \\ n \text{ sign}(b_i(x_i, u_i)) & \text{if } |b_i(x_i, u_i)| > n \end{cases}$$

where $\text{sign}(x) = 1$ if $x \geq 0$ and -1 if $x < 0$. Fix $T < \infty$, $\pi \in \Pi$, and note that for all $x \in \mathbb{R}^m$ the truncated problem (28) admits a unique strong solution $X^n = (X^n(t) : t \in [0, T])$ (see, e.g., [17, Proposition 5.5.17]). Let us now define a region in $\mathbb{R}^m$ such that $b^n(x, u) = b(x, u)$, where $b(x, u)$ is the drift in the original problem (18). To this end, note that by definition of the drift function b and the feasible control set $\mathbb{U}$, there exists a positive constant $R < \infty$ such that $|b(x, u)| \leq R(1 + \|x\|)$ for all $u \in \mathbb{U}(x)$. Put $R_n := n/R - 1$, and consider the ball of radius R_n centered at the origin, $B(0, R_n)$. By construction, $\|b(x, u)\| \leq n$ for all $x \in B(0, R_n)$ and $u \in \mathbb{U}(x)$. Fix $n, \ell \in \mathbb{N}$ such that $n \geq \ell$, $x \in B(0, R_\ell)$, $\pi \in \Pi$, and let

$$T_n^\ell := \inf\{t \geq 0 : X^n(t) \in \partial B(0, R_\ell)\} \tag{29}$$

denote the exit time of X^n from the ball $B(0, R_\ell)$. Define $X(t) := X^n(t)$ for $0 \leq t \leq T_n^\ell$. Note that $X(t)$ is well defined since up to the exit time T_n^ℓ, $X^n(t)$ is the unique strong solution to (28). Moreover, note that b^n is equal to b^ℓ in $B(0, R_\ell)$, so $X^n(t) = X^\ell(t)$ (a.s.) until the exit time from $B(0, R_\ell)$, and $T_\ell^\ell = T_n^\ell$ (a.s.) for all $\ell \leq n$. But note that $b^n(X^n(t), \pi(X^n(t))) \leq n$ (a.s.) for $0 \leq t \leq T_n := T_n^n$. Thus, we have constructed a strong solution, $X(t)$, to the original problem (18), up to the "explosion time" $T_\infty := \lim_{n\to\infty} T_n$. (Note that $T_n \uparrow \infty$ as $n \to \infty$ so the limit T_∞ exists, almost surely.)

Step 2. Our main objective here is to prove that $T_n > T$, eventually, $\mathbb{P}_x$-almost surely. First, note that $\{T_n < T\} \subseteq \{\sup_{t\in[0,T]} \|X^n(t)\| \geq R_n\}$. Now, we take the function $g(x) = \log(1 + \|x\|^2)$ so that

$$\begin{aligned} \nabla g(x) &= \frac{2x}{1 + \|x\|^2} \\ \frac{1}{2} \sum_{i=1}^m \sigma_i^2 \frac{\partial^2 g(x)}{\partial x_i^2} &\leq C_1 m. \end{aligned} \tag{30}$$

Here, and in what follows C_i are constants independent of n. By Itô's formula we have

$$g(X^n(t)) = g(x) + \int_0^t \mathcal{A}^n g(X^n(s))ds + \underbrace{\int_0^t \nabla g(X^n(s))\Sigma dW(s)}_{M_n(t)} \tag{31}$$

where $\mathcal{A}^n$ is the generator corresponding to the truncated problem (28). But note that by (30) and the fact that $\|b(x,u)\| \le R(1+\|x\|)$ we have that

$$\begin{aligned} \mathcal{A}^n g(x) &= \frac{1}{2}\sum_{i=1}^m \sigma_i^2 \frac{\partial^2 g(x)}{\partial x_i^2} + \nabla g(x) \cdot b^n(x, \pi(x)) \\ &\le C_1 m + C_2 \frac{\|x\|^2}{1+\|x\|^2} \le C_3 \ . \end{aligned} \tag{32}$$

Now, the quadratic variation of the stochastic integral $M_n(t)$ is bounded almost surely, $\langle M_n\rangle(t) \le C_1(1+t)$, which follows from the boundedness of ∇g. Thus, we also have $\mathbb{E}M_n^2(t) \le C_1(1+t)$, and therefore $M_n = (M_n(t) : t \in [0,T])$ is an L_2-martingale (with respect to the Brownian filtration $\mathcal{F}_t$). Moreover, note that for all $\kappa > 0$ we have $\sup_n \mathbb{E}\exp\{\kappa\langle M_n\rangle(t)\} < \infty$ for all $t \ge 0$. Thus, Novikov's condition holds (see, e.g., [17, Proposition 3.5.12]) and therefore

$$\tilde{M}_n(t,\kappa) := \exp\{\kappa M_n(t) - \frac{\kappa^2}{2}\langle M_n\rangle(t)\}$$

is a martingale, in particular, $\mathbb{E}\tilde{M}_n(T,\kappa) = 1$. Going back to (31) we have

$$\begin{aligned} \log(1+\|X^n(t)\|^2) &\le \log(1+\|x\|^2) + C_1 T + M_n(t) - (1/2)\langle M_n\rangle(t) + (1/2)\langle M_n\rangle(t) \\ &\le \log(1+\|x\|^2) + C_2(1+T) + M_n(t) - (1/2)\langle M_n\rangle(t), \end{aligned}$$

almost surely for all $t \in [0,T]$. Exponentiating both sides, taking the supremum over $t \in [0,T]$ and taking expectations we have

$$\mathbb{E}_x \sup_{t\in[0,T]} \|X^n(t)\|^2 \le C_3(1+\|x\|^2)e^{C_4 T}\mathbb{E}\left[\sup_{t\in[0,T]} \exp\{M_n(t) - (1/2)\langle M_n\rangle(t)\}\right] \ .$$

We now bound the expectation on the right hand side as follows. First, by Jensen's inequality

$$\mathbb{E}\left[\sup_{t\in[0,T]} \exp\{M_n(t) - (1/2)\langle M_n\rangle(t)\}\right] \le \left(\mathbb{E}\left[\sup_{t\in[0,T]} \exp\{M_n(t) - (1/2)\langle M_n\rangle(t)\}\right]^2\right)^{1/2} ,$$

and for the term on the right hand side we have,

$$\begin{aligned} \mathbb{E}\left[\sup_{t\in[0,T]} \left(\exp\{M_n(t) - (1/2)\langle M_n\rangle(t)\}\right)^2\right] &\le C_1\mathbb{E}\exp\{2M_n(T) - \langle M_n\rangle(T)\} \\ &\le C_2\mathbb{E}\exp\{2M_n(T) - 2\langle M_n\rangle(T)\}\exp\{C_3 T\} \\ &= C_2 e^{C_3 T}. \end{aligned}$$

The first step above uses the L^2-maximal inequality for martingales, the second step follows by adding and subtracting $\langle M_n \rangle(T)$ in the exponent and using the previous bounds on this process, and the third step follows on noting that $\exp\{2M_n(T) - 2\langle M_n\rangle(T)\}$ is a martingale (indeed, the exponential martingale $\tilde{M}_n(t,\kappa)$ with $\kappa = 2$). Combining these arguments we have that

$$\sup_{t\in[0,T]} \|X^n(t)\|^2 \leq Y_n^2$$

with

$$Y_n^2 := C_1(1+\|x\|^2)\exp\{C_2 T\} \sup_{t\in[0,T]} \exp\{M_n(t) - (1/2)\langle M_n\rangle(t)\}$$

independent of t. Also, note that $\mathbb{E}Y_n^2 \leq C_1(1+\|x\|^2)\exp\{C_2 T\}$ independent of t and n.

Step 3. To finish the proof, let

$$A_n = \left\{\omega : \sup_{t\in[0,T]} \|X^n(t,\omega)\| \geq R_n\right\}$$

and note that

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) \leq \sum_{n=1}^{\infty} \mathbb{P}(Y_n \geq R_n)$$

and the latter is finite by the above bounds on $\mathbb{E}Y_n^2$ which assert that $\mathbb{E}Y_n^2 < C$, independent of n. Thus, an application of Borel-Cantelli yields that $\|X^n(t)\| < R_n$, for all but finitely many n, $\mathbb{P}_x$-almost surely for all $t \in [0,T]$. But, by definition of the exit time T_n, this implies that $T_n \geq T$, eventually, $\mathbb{P}_x$-almost surely. Since $T_n \uparrow \infty$ as $n \to \infty$, it follows that $\mathbb{P}_x(T_\infty < T) = 0$. Since T was arbitrary, this proves that the strong solution to (18), $X(t)$, can be constructed on $[0,\infty)$. This concludes the proof. ∎

Proof of Lemma 1: We appeal to the general existence theory in Ladyzenskaya and Uralseva, specifically, their Theorem 8.3 [20, p. 301] lists the conditions that one needs to verify for the differential operator in the HJB equation, so that a smooth solution exists. In particular, the HJB equation in our case is uniformly elliptic and the obvious solvability condition $\alpha > 0$ holds. In addition, the Dirichlet boundary condition is regular (recall, it is zero), and involves a smooth boundary surface (the sphere of radius n in $\mathbb{R}^m$). The only condition that requires verification concerns the regularity of the Hamiltonian $\mathcal{H}(x,p)$ defined in (22). First we prove that $\mathcal{H}(x,\cdot)$ is Lipschitz continuous. Fix $p, q \in \mathbb{R}^m$, and $x \in B(0,n)$. Then

$$\begin{aligned} |\mathcal{H}(x,p) - \mathcal{H}(x,q)| &\leq \|p-q\| \sup_{u\in\mathbb{U}(x)} \|b(x,u)\| \\ &\leq C_2(1+\|x\|)\|p-q\| \end{aligned}$$

which follows from the structure of the feasible control set $\mathbb{U}(x)$. We now show that $\mathcal{H}(\cdot,p)$ is Lipschitz continuous. To facilitate the following derivation, we make a change-of-variable and set

$w_i = (x_i - u_i)/((e \cdot x) \vee 0)$, for $i = 1, \ldots, m$, where $e = (1, \ldots, 1) \in \mathbb{R}^m$. Thus, $u \leq x$ implies $w \geq 0$, and $(e \cdot u) = (e \cdot x) \wedge 0$ implies $e \cdot w = 1$. With this parameterization we have that $b(x, w) = -\theta\mu - \mu x + (\mu - \gamma)(w((e \cdot x) \vee 0))$, and $c(x, w) = c \cdot w((e \cdot x) \vee 0)$. The key observation is that both $b(x, w)$ and $c(x, w)$ are Lipschitz continuous, uniformly in x and w. (This follows upon observing that $f(x) = (e \cdot x) \vee 0$ is Lipschitz continuous.) Now, fix $p \in \mathbb{R}^m$, and $x, y \in B(0, n)$. For any $\delta > 0$ it is possible to choose a $u^* \in \mathbb{U}(y)$ such that $b(y, u^*) \cdot p + c(y, u^*) \leq \mathcal{H}(y, p) + \delta$. Alternatively, $b(y, w^*) \cdot p + c(y, w^*) \leq \mathcal{H}(y, p) + \delta$. Note that $\mathcal{H}(x, p) \leq b(x, w^*) \cdot p + c(x, w^*)$ is obvious since the Hamiltonian is the infimum over all feasible w. Consequently, we have that

$$\begin{aligned} \mathcal{H}(x, p) - \mathcal{H}(y, p) &\leq b(x, w^*) \cdot p + c(x, w^*) - (b(y, w^*) \cdot p + c(y, w^*) - \delta) \\ &\leq C(1 + \|p\|)\|x - y\| + \delta, \end{aligned}$$

where C is a constant that is independent of x, y and p. The reverse inequality is obtained in an analogous manner. Finally, since $\delta > 0$ is arbitrary we may set it to zero to obtain the desired result $|\mathcal{H}(x, p) - \mathcal{H}(y, p)| \leq C(1 + \|p\|)\|x - y\|$. Thus, $\mathcal{H}(\cdot, p)$ is Lipchitz. Since x, y are restricted to $B(0, n)$, and the differential operator is uniformly elliptic, we also have that the gradient of the solution is bounded (cf. [20, p. 296]), which implies boundedness of p, q. This, and the local Lipschitz result above suffice to verify the remaining regularity conditions in [20, Theorem 8.3]. Thus, by the cited theorem we have existence of a C^2 solution to the PDE. Standard arguments yield that V_n, the value function for the truncated problem, is the unique bounded solution of the truncated HJB equation; for details see, e.g., Fleming and Soner [11, pp. 172-174]. We omit the details; more general arguments appear in the verification step (Step 3) used in the proof of Theorem 1, and the proof of existence of optimal Markov policies (Theorem 2). This concludes the proof. ■

Proof of Lemma 2: It suffices to seek upper and lower bounds on the individual coordinates of the controlled diffusion, in terms of "nice" processes which yield the requisite growth rate stated in the Lemma. To this end, let $\mathcal{I}_\mu = \{i \in \{1, \ldots, m\} : \gamma_i \leq \mu_i\}$, and $\mathcal{I}_\gamma = \{i \in \{1, \ldots, m\} : \gamma_i > \mu_i\}$ (allowing for the possibility that one of the two sets is empty). Note that for $i \in \mathcal{I}_\mu$ we have for all $x \in \mathbb{R}^m$ and admissible $u \in \mathbb{U}(x)$ that the drift of the controlled diffusion is bounded as follows

$$\begin{aligned} b_i(x, u) &= -\theta\mu_i - \gamma_i(x_i - u_i) - \mu_i u_i \\ &= -\theta\mu_i - \gamma_i x_i - (\mu_i - \gamma_i)u_i \\ &\geq -\theta\mu_i - \mu_i x_i \ , \end{aligned} \tag{33}$$

where the inequality is a consequence of $u_i \leq x_i$ for all admissible $u \in \mathbb{U}(x)$. Similarly, we have for $i \in \mathcal{I}_\gamma$ that $b_i(x, u) \leq -\theta\mu_i - \mu_i x_i$. Now, fix a complete filtered probability space and an m-dimensional standard Brownian motion $W = (W(t) : t \geq 0)$ where $W(t) = (W_1(t), \ldots, W_m(t))$.

Fix $x \in \mathbb{R}^m$ and let $Y_i = (Y_i(t) : t \geq 0)$ be the solution of the stochastic differential equation

$$dY_i(t) = -(\theta\mu_i + \mu_i Y_i(t))dt + \sigma_i dW_i(t),$$

with $Y_i(0) = x_i$, for $i = 1, \ldots, m$. Fix $\pi \in \Pi$ and let X be a solution of the controlled diffusion (18). Fix $i \in \mathcal{I}_\mu$, assuming this set is non-empty, and let $\Delta_i(t) = Y_i(t) - X_i(t)$. Then, we have that

$$\Delta_i(t) = \int_0^t \left[-\theta\mu_i - \mu_i Y_i(s) - b_i(X(s), \pi(X(s)))\right] ds \ .$$

Note that the drift ordering in (33) is policy independent and moreover does not depend on $\{x_j : j \neq i\}$. Also, Y_i has linear drift which does not depend on π. In this setting we can replicate the comparison proof in [17, Proposition 5.2.18] yielding $(\Delta(t))^+ \leq \mu_i \int_0^t (\Delta(s))^+ ds$, $\mathbb{P}_x$-almost surely for all $t \geq 0$. Gronwall's inequality implies that $(\Delta(t))^+ = 0$, thus $Y_i(t) \leq X_i(t)$, $\mathbb{P}_x$-almost surely for all $t \geq 0$. The same argument yields $Y_i(t) \geq X_i(t)$ for $i \in \mathcal{I}_\gamma$, $\mathbb{P}_x$-almost surely for all $t \geq 0$.

Observe that for any $x \in \mathbb{R}^m$ and $u \in \mathbb{U}(x)$, we have

$$\begin{aligned}
\Delta b(x,u) &:= -\sum_{i \in \mathcal{I}_\mu} \frac{b_i(x,u)}{\mu_i} + \sum_{i \in \mathcal{I}_\gamma} \frac{b_i(x,u)}{(\gamma_i - \mu_i)} \\
&= \sum_{i \in \mathcal{I}_\mu} \frac{\gamma_i}{\mu_i}(x_i - u_i) + \sum_{i \in \mathcal{I}_\mu} u_i - \sum_{i \in \mathcal{I}_\gamma} \frac{\gamma_i}{(\gamma_i - \mu_i)} x_i + \sum_{i \in \mathcal{I}_\gamma} u_i \\
&\overset{\text{(a)}}{\geq} -\sum_{i \in \mathcal{I}_\gamma} \frac{\gamma_i}{(\gamma_i - \mu_i)} x_i + \sum_{i=1}^m u_i \\
&\overset{\text{(b)}}{=} \begin{cases} -\sum_{i \in \mathcal{I}_\gamma} \frac{\gamma_i}{(\gamma_i - \mu_i)} x_i & \sum_{i=1}^m x_i \geq 0 \\ -\sum_{i \in \mathcal{I}_\gamma} \frac{\mu_i}{(\gamma_i - \mu_i)} x_i + \sum_{i \in \mathcal{I}_\mu} x_i & \sum_{i=1}^m x_i < 0 \end{cases} ,
\end{aligned} \tag{34}$$

where (a) follows from $x_i \geq u_i$, and (b) follows from the fact that $\sum_{i=1}^m u_i = (\sum_{i=1}^m x_i) \wedge 0$. Now, fix a policy $\pi \in \Pi$, and an index $j \in \mathcal{I}_\gamma$, assuming this set is non-empty. Then, for some finite positive constants $C_1, \ldots, C_4$ we have

$$\begin{aligned}
\frac{X_j(t)}{(\gamma_j - \mu_j)} &= \sum_{i \in \mathcal{I}_\mu} \frac{X_i(t)}{\mu_i} - \sum_{i \in \mathcal{I}_\gamma, i \neq j} \frac{X_i(t)}{(\gamma_i - \mu_i)} - \sum_{i \in \mathcal{I}_\mu} \frac{X_i(t)}{\mu_i} + \sum_{i \in \mathcal{I}_\gamma} \frac{X_i(t)}{(\gamma_i - \mu_i)} \\
&\overset{\text{(c)}}{\geq} -C_1 \sum_{i=1}^m |Y_i(t)| - C_2 t + \int_0^t \Delta b(X(s), \pi(X(s))) ds - \sum_{i \in \mathcal{I}_\mu} \sigma_i W_i(t) + \sum_{i \in \mathcal{I}_\gamma} \sigma_i W_i(t) \\
&\overset{\text{(d)}}{\geq} -C_1 \sum_{i=1}^m |Y_i(t)| - C_2 t - C_3 \int_0^t \sum_{i=1}^m |Y_i(s)| ds - C_4 \sum_{i=1}^m |W_i(t)| \ ,
\end{aligned} \tag{35}$$

where (c) follows from $X_i(t) \geq -|Y_i(t)|$ for $i \in \mathcal{I}_\mu$ and $X_i(t) \leq |Y_i(t)|$ for $i \in \mathcal{I}_\gamma$; and, (d) follows from (34). For $i \in \mathcal{I}_\gamma$ we already have the upper bound $X_i \leq Y_i$, where the processes $\{Y_i\}$ are stable Ornstein-Uhlenbeck processes. In particular, $\mathbb{E}_x|Y_i(t)| \leq C_5(1 + \|x\|)$ for some finite positive constant independent of t. Consequently, using (35) we have that for all $i \in \mathcal{I}_\gamma$

$\mathbb{E}_x^\pi|X_i(t)| \leq C(1+\|x\|)(1+t)$ for some finite positive constant C independent of x, t and the policy π. For $i \in \mathcal{I}_\mu$ we can derive a lower bound on $-X_i(t)/\mu_i$ by essentially repeating the above arguments verbatim. This, in turn, yields an upper bound on $X_i(t)$ that together with the lower bound $X_i \geq Y_i$ establishes that $\mathbb{E}_x^\pi|X_i(t)| \leq C(1+\|x\|)(1+t)$ for all $i \in \mathcal{I}_\mu$. This concludes the proof. ■

Acknowledgements. We are indebted to George Papanicolaou for helpful discussions concerning the existence of smooth solutions to the Hamilton-Jacobi-Bellman equation, and Sunil Kumar for introducing us to the finite element numerical solution method. We would also like to thank Rami Atar for his careful reading and helpful comments, and Avi Mandelbaum for providing us with preprints of his work with R. Atar and M. Reiman. Finally, we wish to thank two anonymous referees for their careful reading and helpful suggestions.

References

[1] R. Atar, A. Mandelbaum and M.I Reiman. A Brownian control problem for a simple queueing system in the Halfin-Whitt regime. Technical report #1297, Dept. of Electrical Engineering, Technion, 2001.

[2] R. Atar, A. Mandelbaum and M.I Reiman. Scheduling a multi-class queue with many exponential servers: Asymptotic optimality in heavy-traffic. Preprint, July, 2002.

[3] E.B. Becker, G.C. Carey and J.T. Oden. *Finite elements: An introduction*. Prentice Hall, 1981.

[4] A. Bensoussan. *Stochastic Control by Functional Analysis Methods*. Studies in Mathematics and its Applications, North-Holland Publishing Co., Amsterdam-New York, 1982

[5] D.P. Bertsekas. *Dynamic Programming and Optimal Control*. Athena Scientific, 1995.

[6] S. Borst, A. Mandelbaum, and M. Reiman. Dimensioning of Large Call Centers. *preprint*, 2000.

[7] S. Browne and W. Whitt. Piecewise - linear diffusion processes. Chapter 18 in Advances in Queueing, CRC Press, 463-480, 1995.

[8] A. Das and R. Srikant. Diffusion approximations for a single node accessed by congestion controlled sources. *IEEE Transactions on Automatic Control*, **45**: 1783-1799, 2000.

[9] P. Fleming and B. Simon. Heavy traffic approximations for a system of infinite servers with load balancing. *Probab. Engrg. Inform. Sci.*, **13**: 251–273, 1999.

[10] P. Fleming, A. Stolyar, and B. Simon. Heavy traffic limit for a mobile phone system loss model. Proceedings of the Nashville ACM Telecommunications Conference, 1995.

[11] W.H. Fleming and H.M. Soner *Controlled Markov Processes and Viscosity Solutions*. Applications of Mathematics, Springer-Verlag, Berlin-New York, 1993.

[12] O. Garnett, A. Mandelbaum and M. Reiman. Designing a call center with impatient customers. *Manufacturing and Service Operations Management*, **4:**208-227, 2002.

[13] S. Halfin and W. Whitt. Heavy-traffic limits for queues with many exponential servers. *Op. Res.*, **29**: 567-588, 1981.

[14] J.M. Harrison and J.A. Van Mieghem. Dynamic Control of Brownian Networks: State Space Collapse and Equivalent Workload Formulations. *Ann. Appl. Probab.*, **7**:747-771, 1997.

[15] J.M. Harrison. Heavy traffic analysis of a system with parallel servers: asymptotic optimality of discrete-review policies. *Ann. Appl. Probab.*, **8:**822–848, 1998.

[16] J.M. Harrison and M.J. López. Heavy traffic resource pooling in parallel-server systems. *QUESTA*, **33:**339–368, 1999.

[17] I. Karatzas and S.E. Shreve. *Brownian Motion and Stochastic Calculus*. Springer, 1991.

[18] F.P. Kelly and C.N. Laws. Dynamic routing in open queueing networks: Brownian models, cut constraints and resource pooling. *Queueing Systems Theory Appl.*, **13**:47-86, 1993.

[19] S. Kumar and M. Muthukumar. A numerical method for solving singular Brownian control problems. *Working paper*, Stanford GSB, 2001.

[20] O.A. Ladyzhenskaya and N.N. Urlaseva. *Linear and Quasilinear Elliptic Equations*. Academic press, NY, 1968.

[21] N.V. Krylov. *Controlled Diffusion Processes*. Applications of mathematics, no. 14, Springer-Verlag, 1980.

[22] H.J. Kushner and P. Dupuis. *Numerical Methods for Stochastic Control Problems in Continuous Time*. Springer-Verlag, NY, 1992.

[23] H.J. Kushner. Optimal discounted stochastic control for diffusion processes. *SIAM J. Control*, **5**:520–531, 1967.

[24] A. Puhalskii and M. Reiman. The multiclass $GI/PH/N$ queue in the Halfin-Whitt regime. *Adv. in Appl. Probab.*, **32**:564–595, 2000.

[25] W. Rudin. *Real and Complex Analysis*. McGraw Hill, 1987.

[26] L.I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley, 1999.

[27] J. A. Van Mieghem. Dynamic scheduling with convex delay costs: the generalized $c\mu$ rule. *Ann. Appl. Probab.*, **5**:809–833, 1995.

[28] A.Y. Veretennikov. Strong solutions and explicit formulas for solutions of stochastic integral equations. *Mat. Sb.*, **111(153):**434–452, 1980.

[29] W. Whitt. Understanding the efficiency of multi-server service systems. *Manag. Sci.*, **38**:708-723, 1992.

[30] R.J. Williams. On dynamic scheduling of a parallel server system with complete resource pooling, in *Analysis of communication networks: Call centers, traffic and performance*, Ed. D.R. McDonald and S.R. Turner, Fields Institute communication series, AMS publications, 2000.

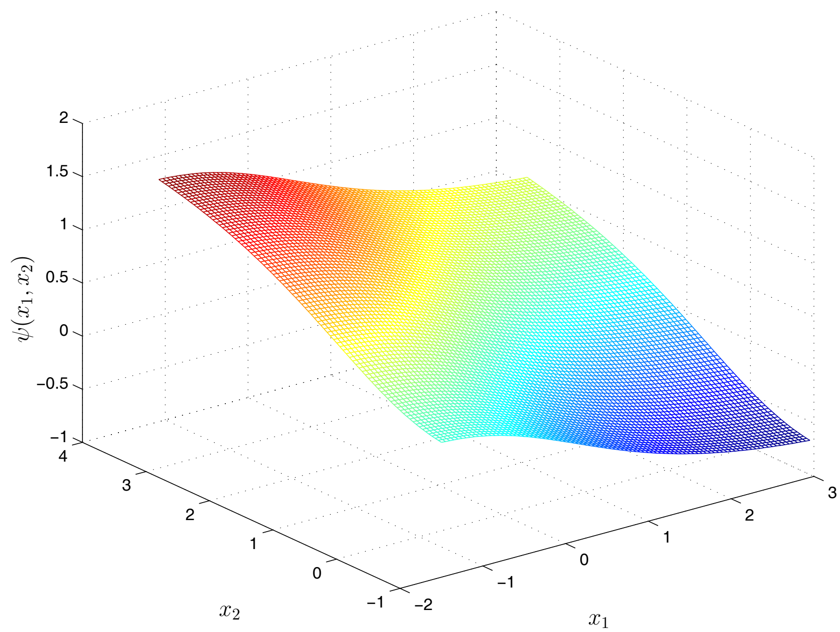

Figure 2: The test quantity $\psi(x_1, x_2)$ as a function of state.

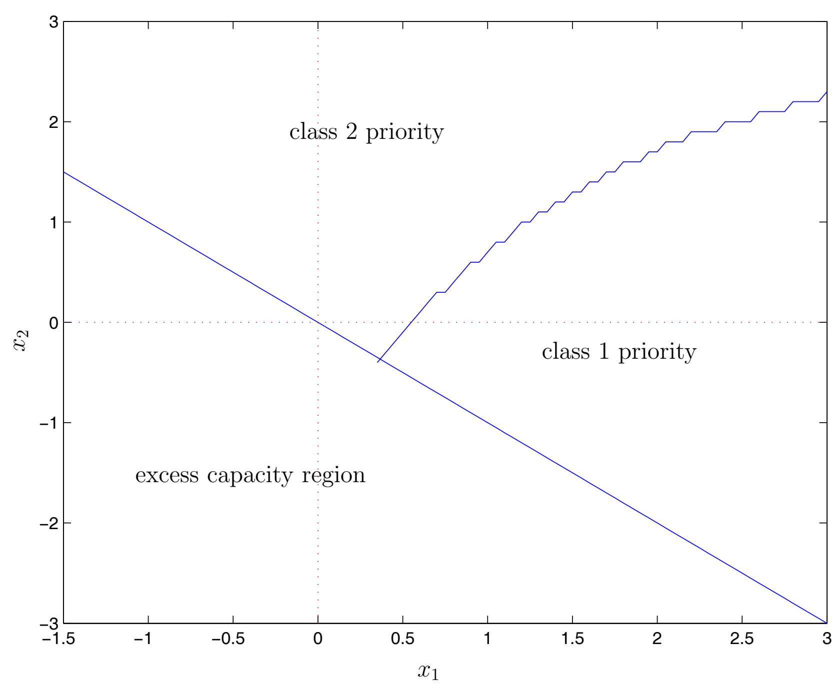

Figure 3: Priority regions for the optimal index policy $i^*(x_1, x_2)$.