MEAN-FIELD METHODS IN LARGE STOCHASTIC NETWORKS

Eric Friedlander

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Statistics and Operations Research.

Chapel Hill
2018

Approved by:

Amarjit Budhiraja

Vladas Pipiras

Shankar Bhamidi

Vidyadhar Kulkarni

Chuanshu Ji

ABSTRACT

ERIC FRIEDLANDER: Mean-Field Methods in Large Stochastic Networks
(Under the direction of Amarjit Budhiraja)

Analysis of large-scale communication networks (e.g. ad hoc wireless networks, cloud computing systems, server networks etc.) is of great practical interest. The massive size of such networks frequently makes direct analysis intractable. Asymptotic approximations using fluid and diffusion scaling limits provide useful methods for approaching such problems. In this dissertation, I study such approximations in two different settings. In the first, I consider a rate control problem for a weakly interacting particle system. I show that by considering an associated diffusion control problem, one can construct controls which are asymptotically optimal for the finite particle system control problem. In the second, I consider a class of load balancing mechanisms in a large cloud-storage network that uses a Maximum Distance Separable coding scheme to store a large set of files. Fluid and diffusion approximations are developed for this system and the long-time behavior of the network is studied.

To my father.

ACKNOWLEDGEMENTS

I would like to thank all the people without whom this dissertation would not be possible.

First, I would like to express my deepest gratitude and admiration for my advisor, Professor Amarjit Budhiraja, as I can't even imagine having a better mentor. I am forever indebted to him for his patience, guidance, teaching, training, and career advice. Throughout my graduate career here at UNC, Amarjit has helped to provide ideas when I've been stuck and proofread both the grammar and technical details of my work. Later on, he provided invaluable career advice and helped to keep me calm while I was on the job market. Without his guiding hand I would not be as successful (or sane) as I am now. I feel lucky to have had such a wonderful advisor. Finally, I would like to thank him for all of the funding he has provided me throughout my tenure at UNC. This financial support allowed me to fully focus on my work, as well as attend a variety of conferences and workshops integral to my development as a researcher.

I would also like to thank the other members of my thesis committee. Professors Shankar Bhamidi and Vladas Pipiras have served as wonderful friends, mentors, and teachers during my time here at UNC. I appreciate all the of time you've spent working on projects with me, teaching me, and (last but not least) joking around with me. The two of you have provided an incredible amount of support, and for that I am incredibly grateful. I very much appreciate all of the guidance given to me by Professor Chuanshu Ji. In particular, I am very thankful for all of the teaching advice given to me by Professor Ji along with Professor Ed Carlstein and Dr. Robin Cunningham. I am forever indebted to Professor Vidyadhar Kulkarni for his advice aand suggestions as a member of my thesis committee. In addition, I feel that Professor Kulkarni's jovial nature and youthful exuberance is integral to creating the wonderful culture in the STOR department that has made my time here so great. While not part of my thesis committee I would also like to thank Professor Steve Marron for all of his invaluable guidance working on projects unrelated to my thesis work. Finally, I would like to thank all of my teachers without whom I have never been able to complete this dissertation.

I would be remiss if I did not acknowledge the immense contributions of the department's unsung heroes. Alison Kieber, Christine Keat, Sam Radel, and Linda Stutts have provided invaluable help, guidance, support, and friendship throughout my time at UNC. I would like to thank Christine for helping to organize and send out all of my recommendation letters. In addition, Christine was instrumental in helping to sort out any class/scheduling related problems I ever had. This was no small undertaking and I would certainly have been unable to do it myself. Thank you to Sam and Linda for helping to organize all things business/financial related. I appreciate all of Alison's help with organizing everything from textbooks to classrooms. Your help has been invaluable throughout my time here. Finally, I would like to thank all of you for the time you spent chatting with me. I feel that we've had some wonderful conversations over the years and I will truly miss spending time with all of you.

I would like to take a moment to acknowledge my undergraduate advisor Professor Rudy Guerra and his wife Nancy Guerra. It is because of Rudy that I even ended up in the field of statistics and I can honestly say that I would be a much different person if not for his advice, guidance, and friendship. Rudy remains one of the most important people in my life and one of the small group of people whose advice I constantly solicit and trust. It is unlikely that I would have made it to this point without his help. In addition, Nancy is one of the most loving and uplifting people I have ever had the pleasure to meet. While I have not tasted any of her amazing cooking in some years, I always enjoy having the opportunity to chat with her over the phone and catch up. It never fails to bring a smile to my face.

I would like to thank all of my friends. Thank you:

- Mike Lamm for spending half of your weekends with me and being such a great friend.
- James Wilson and Rosie Scott for all of your advice, friendship, and fun times.
- Jimmy Jin, Kelly Bodwin, and John Palowitch for your advice as older students.
- Donqing Yu, Eunjee Lee, Samopriya Basu, Kevin O'Connor, and Sepehr Moravej for being such wonderful officemates.
- Jon Williams and Iain Carmichael for all the fun times and good discussions we've had.

- Meilei Jiang and Suman Chakraborty for the homework help, discussions, and fun experiences we've had over the past five years.
- Dylan Glotzer and Lindsey Comer for being such great friends. Also for performing my marriage.
- Adam Waterbury, Michael Conroy, Candice Crilly, Jack Prothero, Aditya Bhalaram, Carson Mosso, Mark He, and Mike Perlmutter for all the fun we've had together.
- Ruoyu Wu for the insightful discussions and talks that we've had.
- Sayan Banerjee for joking around with me and for the useful discussions we've had.

Last but not least, I'd like to thank my family. I am eternally grateful for my brilliant, caring, and beautiful wife Maria. I appreciate all of the support you've given me over the years. Of course I would not be here without a little help from my mother. She has always been incredibly supportive throughout my life and there is no way I would have made it this far without her encouragement. Finally, I'd like to thank my brother Adam, my aunt Melissa, and my grandmother Maria. I am extremely lucky to have such a supportive family and am grateful for all that you've done for me over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

JSQ	Join-the-Shortest Queue
JIQ	Join-the-Idle Queue
MDS	Maximum Distance Separable
LLN	Law of Large Numbers
CLT	Central Limit Theorem
ODE	Ordinary Differential Equation
SDE	Stochastic Differential Equation
FIFO	First-In-First-Out
RCLL	Right continuous functions (stochastic processes) with left limits
$\mathbb{N}$	Set of natural numbers
$\mathbb{N}_0$	Set of non-negative integers
$\mathbb{R}$	Set of real numbers
$\mathbb{R}_+$	Set of non-negative real numbers
$\mathbb{R}^d$	Set of d-dimensional real vectors
$\mathbb{Z}$	Set of integers
$\mathcal{B}(\mathbb{S})$	Borel σ-field on a topological space $\mathbb{S}$
$\mathbb{C}_b(\mathbb{S})$	The collection of bounded continuous functions from space $\mathbb{S}$ to $\mathbb{R}$
$\mathbb{C}([0,T]:\mathbb{S})$	The space of continuous functions from $[0,T]$ to $\mathbb{S}$
$\mathbb{C}(\mathbb{S}_1:\mathbb{S}_2)$	The space of continuous functions from $\mathbb{S}_1$ to $\mathbb{S}_2$
$\mathbb{C}_b(\mathbb{S}_1:\mathbb{S}_2)$	The space of bounded continuous functions from $\mathbb{S}_1$ to $\mathbb{S}_2$
$\mathbb{C}^k(\mathbb{R}^d)$	The space of continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$ whose first k derivatives exist and are continuous
$\mathbb{C}_c^k(\mathbb{R}^d)$	The subset of $\mathbb{C}^k(\mathbb{R}^d)$ consisting of functions with compact support
$\mathbb{C}^\infty(\mathbb{R}^d)$	The space of continuous functions from $\mathbb{R}^d$ to $\mathbb{R}$ where all derivatives exist and are continuous
$\mathbb{C}^{1,2}([0,T]\times\mathbb{R}^d)$	The space of continuous functions from $[0,T]\times\mathbb{R}^d$ to $\mathbb{R}$ once continuously differentiable in the time coordinate and twice continuously differentiable in the space coordinate
$\mathbb{D}([0,T]:\mathbb{S})$	The space of right continuous functions with left limits from $[0,T]$ to $\mathbb{S}$
$\lvert A\rvert$	Cardinality of set A

$B(r)$	The L^1 ball of radius r in $\mathbb{R}^d$ centered at the origin
$\|\cdot\|$	Euclidean norm in $\mathbb{R}^d$
$\|f\|_\infty$	The sup norm of a real-valued function f (i.e. $\sup_{x\in\mathbb{S}} \|f(x)\|$)
Sp A	Linear span of a set A
$\mathbb{M}^{m\times n}(\mathbb{S})$	The space of $m \times n$-dimensional matrices whose entries take values is $\mathbb{S}$
$\mathrm{Tr}(M)$	The trace of a matrix M
$\mathbf{1}$	Vector of 1's
I	Identity matrix or operator (depending on context)
$\mathcal{M}(\mathbb{S})$	The space of locally finite measures on Polish space $\mathbb{S}$
$\boldsymbol{\ell}_2$	The space of square summable real valued sequences
$\boldsymbol{\ell}_1$	The space of absolutely summable real valued sequences
$\|\cdot\|_2$	The norm of $\boldsymbol{\ell}_2$
$\|\cdot\|_1$	The norm of $\boldsymbol{\ell}_1$
$\langle x, y\rangle_2$	The inner product between $x, y \in \boldsymbol{\ell}_2$
$\|A\|_{\mathrm{HS}}$	The Hilbert-Schmidt norm of Hilbert-Schmidt operator A
$\mathcal{M}_T^2(\mathbb{H})$	Space of continuous, square integrable martingales $M \equiv \{M(t)\}_{0\le t\le T}$ taking values in the Hilbert space $\mathbb{H}$ with $M(0) = 0$
$(a)_+$	The positive part of a
$X_n \Rightarrow X$	Weak convergence of random variables with values in some metric space

CHAPTER 1
Introduction

"*All models are wrong but some are useful!*" - George Box

Recent technological advances in the telecommunications industry have led to a boom in the use of distributed processor networks, cloud-based marketplaces and storage networks, and mobile and sensor networks. The ubiquity of such systems has prompted a significant amount of research into useful models for these systems (Antunes et al., 2008; Bonald et al., 2004; Gibbens et al., 1990; Gupta and Kumar, 2003; Ganesh et al., 2003). One of the main difficulties in such modeling arises from the large size of the systems being considered. Models for such systems typically take the form of continuous or discrete time Markov chains and the networks of interest frequently have events (e.g. job arrivals, purchases, file requests, etc.) which occur at a rate which scales with the size of the network. The resulting processes have jumps which occur extremely quickly, and thus applying standard techniques to analyze these models for large networks becomes intractable. In order to simplify, it is useful to consider asymptotic approximations of such systems under a suitable scaling. Specifically, in many settings, by speeding the system up and scaling the state space, one can establish tractable approximations of the underlying system in the form of ordinary differential equations (ODE) or stochastic differential equations (SDE). These limiting descriptions provide tractable model simplifications for analyzing the underlying system. In this dissertation, I study two different problem settings where such approximation methods can be developed.

1.1 Summary of Thesis

1.1.1 Diffusion Approximations for Controlled Weakly Interacting Large Finite State Systems with Simultaneous Jumps

We study a pure jump, weakly interacting, Markovian particle system in which jump rates can be dynamically modulated by a controller. The stochastic system of interest describes the state evolution of a collection of n particles where each particle's state takes values in a finite set $\mathbb{X}$. By a weak interaction we mean that the jump rates for a typical particle depend on the states of the remaining particles through the empirical distribution of particle states. System dynamics will allow for multiple particles to change states simultaneously, but there will be a fixed finite number of jump types. Such jump-Markov processes have been proposed as models for ad hoc wireless networks (Antunes et al., 2008) of the following form. Consider a system of n finite capacity servers (particles/nodes). Jobs of K different types, each with their own capacity requirement, arrive at each node at rate $\lambda_k, k = 1, \ldots, K$ and are admitted if there is enough available capacity. All the jobs in the system of type k have exponential residence time with mean τ_k^{-1}. After an exponential holding time with mean γ_k^{-1} a job of type k will attempt to switch to another server which is chosen uniformly at random, and is admitted if there is available capacity, otherwise the job is lost. The state of a particle describes the number of various types of jobs being processed at the server. Under conditions, by classical results, the stochastic process of particle state empirical measures converges to the solution of a d-dimensional ODE (cf. (Kurtz, 1970)), where $d = |\mathbb{X}|$. This ODE captures the nominal behavior of the system over time as n becomes large.

Taking a different perspective, the analysis of such ODE is a natural starting point for system design. By studying the mapping between system parameters and solution sets of the ODE one can identify parameter values that lead to desirable system behavior over time, at least in the law of large number limit as determined by the solution of the ODE. However, even when the system has been designed to reproduce a certain targeted nominal behavior the actual stochastic process of interacting particles may deviate significantly from the behavior determined by the ODE. It then becomes of interest to study dynamic control algorithms that modulate controllable system parameters to nudge the stochastic process closer to its

desired nominal behavior. In general, adjusting system parameters incurs a cost and thus there is a trade off between this and the cost for deviating from the nominal behavior. A natural approach for analyzing this trade off is through an optimal stochastic control formulation where the controller seeks to minimize a suitable cost function which accounts for both types of costs noted above.

The goal of this work is to develop a systematic stochastic control framework for studying optimal regulation of large, weakly interacting, pure jump Markov processes that arise from problems in communication networks. Since the jump rates in the system are of $\mathcal{O}(n)$, and in a typical system n is large, an exact analysis of this control problem becomes computationally intractable and thus one seeks a suitable approximate approach. The basic idea is to consider a sequence of networks indexed by n such that the given physical system is embedded in this sequence for some fixed large value of n. A suitable asymptotic model, as $n \to \infty$, is used as a surrogate for the control problem in the n-th network. The asymptotic model taken here is based on diffusion approximations which give the limit behavior of fluctuations of the empirical measure process from its law of large number (LLN) limit. In an uncontrolled setting, such diffusion limits can be derived from classical martingale problem techniques (Kurtz, 1971; Joffe and Métivier, 1986) that are also the starting point here for developing an asymptotic framework for the study of the optimal stochastic control problem. Diffusion approximation methods have been used extensively in stochastic network theory, in particular they have been very useful in the study of critically loaded stochastic processing networks (see (Kushner, 2013; Harrison, 1988; Atar and Shifrin, 2014; Bell and Williams, 2001; Dai and Lin, 2008; Whitt, 2002; Budhiraja and Ghosh, 2012; Budhiraja et al., 2011) and references therein). In this context, diffusion processes arise as approximations for a fixed number of centered renewal processes with rates approaching infinity. Limit theorems and the scaling regime considered in these works (number of nodes is fixed, traffic intensity approaches 1) is quite different from the one where the number of nodes (particles) approaches infinity that is considered here. In communication systems that motivate the study of such interacting processes, jumps correspond to either an admission of a job to one of the n nodes in the system, transfer of a job from one node to another node, or the completion/rejection of a job (and thus exit from the system). We consider a formulation in which controls can make "small" adjustments to the rate values in

order to nudge the system toward its nominal state. Specifically, the overall rate of jumps in the system is $\mathcal{O}(n)$ whereas the allowable rate controls will be $\mathcal{O}(\sqrt{n})$. Although the magnitude of control becomes negligible compared to the overall rate as n becomes large, in the diffusion scaling such a control can lead to an appreciable improvement in performance (see Section 3.6 for some numerical results). In the LLN limit the controlled and uncontrolled systems both converge to the same nominal behavior as expected, but the diffusion limit of the two systems will in general differ in the drift coefficient. In particular, under suitable feedback controls the centered and normalized controlled process will converge to a diffusion with a nonlinear (in state) drift term whereas the uncontrolled process will converge to a time inhomogeneous Gauss-Markov process. In terms of cost, one can consider various types of criteria, but for simplicity we restrict ourselves to a finite time horizon cost where the running cost is a sum of two terms. The first term is a continuous function, with at most polynomial growth, of the state of the centered and normalized empirical measure, and the second is a finite convex function of the (normalized) control.

Rather than attempting to look for an optimal control for the stochastic control problem for a fixed value of n, i.e. for the n-th system, we instead focus on the more tractable goal of asymptotic optimality. More precisely, we are interested in constructing a sequence of control policies (indexed by n) such that the cost associated with the n-th system under the n-th control policy converges to the smallest possible value as $n \to \infty$. Analogous notions of asymptotic optimality are routinely used in heavy traffic analysis of queuing networks (Kushner, 2013; Harrison, 1988; Atar and Shifrin, 2014; Bell and Williams, 2001; Dai and Lin, 2008; Budhiraja and Ghosh, 2012; Budhiraja et al., 2011), but in the current work they are introduced in a very different asymptotic regime. The key ingredient in the approach is to formulate and analyze a closely related stochastic control problem for diffusion processes. Roughly speaking, the state process in the diffusion control problem is the asymptotic analogue of the centered and normalized empirical measure process as $n \to \infty$. The control enters in the drift of the diffusion process whereas the diffusion coefficient is a non-random function of time. Our main result, Theorem 2, shows that the diffusion control problem is a good approximation of the control problem for the n-th system, when n is sufficiently large. Specifically, this theorem says that the value function associated with the control problem for the n-th system converges to the

value function of the limit diffusion control problem. In addition, the result states that for any $\varepsilon > 0$, there exists an ε-optimal continuous feedback control and that the cost incurred from using such a control in the prelimit system will converge to the cost in the limit. What this means is that instead of solving the original control problem, one can solve a diffusion control problem. Using that solution in the original system will yield a near optimal solution if n is large.

In Section 3.6, we will illustrate our approach through a numerical example. This example is the controlled analogue of a model introduced in (Antunes et al., 2008), and one can approach more general forms of this model along similar lines. The running cost function we consider is quadratic in the normalized state and control processes. The corresponding limit diffusion control problem in this case becomes the classical stochastic linear quadratic regulator (LQR) with time dependent coefficients (see (Fleming and Rishel, 1976)). The optimal feedback control for the diffusion control problem can be given explicitly by solving a suitable Riccati equation. Our numerical results show that implementation of the control policy based on the optimal feedback control for the limit LQR to a system with $n = 10,000$ leads to an improvement of up to 15.5% on the cost for the uncontrolled system.

1.1.2 Load Balancing Mechanisms in Cloud Storage Systems

In the world of cloud-based computing, large data centers are often used for file storage. These data centers consist of large networks of servers that are used to store even larger sets of files. In order to improve reliability and retrieval speed, these files are often "coded". By coded, we mean that the file is broken down into smaller pieces which are stored on multiple servers. Consider the situation in which there are four servers and one file. One can store the entire file on one server but in such a configuration the file would be inaccessible if that server were to fail. In order to improve reliability, one can replicate the file across all four servers but such a method would require much more memory. Suppose we instead split the file into halves, A and B, and then store $A,\ B,\ A+B,\ A-B$ in each of the four servers, respectively. Then the original file can be constructed from any two pieces. One can extend this idea to the case where equally sized pieces of a file are stored across L servers and any k pieces can reconstruct the original file.

This can be accomplished using the Maximum Distance Separable (MDS) code with parameters (L, k) (Lin and Costello, 2004). The MDS code greatly improves reliability since $L-k+1$ servers must fail before the file becomes irretrievable, while only requiring enough total memory to store L/k files. Given a coding scheme, one can consider load balancing mechanisms to improve file retrieval speed. In (Li et al., 2016), two routing schemes, called Batch Sampling (BS) and Redundant Request with Killing (RRK), are considered. In BS routing, incoming jobs are routed to the k shortest queues containing the file being requested, while in RRK routing jobs are routed to all servers containing the requested file and then removed from the queue (killed) once k pieces of the file have been returned. The paper (Li et al., 2016) formally calculates the steady state ($T \to \infty$) queue length distribution in the large system limit ($n \to \infty$) and gives simulation results for different values of L and k in both routing schemes. In this work we focus primarily on BS routing.

We are interested in developing a rigorous limit theory for such load balancing schemes for systems with MDS coding as n becomes large. Specifically, we establish law of large numbers and diffusion approximations for such systems under an appropriate scaling, as $n \to \infty$. Such limit theorems provide useful model simplifications that can then be employed for approximate simulation of the large and complex n-server systems (see Section 4.5 for some numerical results). These limit theorems are also the first steps towards making rigorous the program initiated in (Li et al., 2016) of developing steady state approximations for such systems, with provable convergence properties as n becomes large.

We consider a system with n servers on which $I(n)$ files are stored using MDS coding with parameters (L, k). A key assumption to our analysis is that the files are stored such that each combination of L servers has exactly c files. We further assume that jobs arrive in the system at rate $n\lambda$ and request a file uniformly at random. This is another simplifying assumption on our model that roughly says that all files are in equal demand. These structural assumptions imply a convenient exchangeability property of the system which allows for the use of certain mean-field approximation techniques. A single file request spawns k jobs which are then routed into the k shortest queues within the set of L servers containing the file being requested. Each server processes the jobs in their queue at rate k according to the first-in-first-out (FIFO) discipline and processing times are mutually independent. Regarding each server as a "particle", the

above formulation describes an interacting particle system with *simultaneous jumps*. Note that the symmetry structure introduced above implies that every time a file request arrives, it leads to a selection of L servers uniformly at random (from which the k servers with shortest queues are chosen). In particular this says that the well studied "Power-of-d" routing scheme (also known as the "supermarket model") is a special case of the scheme considered here on taking $L = d$ and $k = 1$. Direct analysis of such large and complex n-server systems is challenging even by simulation methods as frequently the servers in networks of interest number in the hundreds of thousands with arrival rates of file requests of similar order. The goal of this work is to develop suitable approximate approaches to such systems.

Limit theorems of the form studied in this work can be used for model simplification and for computing approximations for performance measures, e.g. through simulation methods. Direct simulation of the underlying n-server system would in general be prohibitively expensive for large n since the jumps in the system occur at rate proportional to n. The asymptotic approximations given in this work (cf. Theorems 10 and 14) allow a system manager to simulate performance metrics for the system at a coarser scale via numerical ODE solvers or Euler discretizations for SDE (see Section 4.5 for an example). Although the systems considered here are required to satisfy certain symmetry conditions (all files are equally sized and all jobs are in equal demand), the simplified models given by the limiting ODE and SDE give useful qualitative insights into the behavior of large storage networks employing these types of coding schemes.

The results obtained here are useful in analyzing the long-time behavior of such systems as well, e.g. in providing information on the rate at which the queue lengths decay in steady-state and how such a decay is impacted by different values of L and k. We show that the ODE system that determines the LLN behavior of the occupancy measure process has a unique fixed point $\bar{u}$ which is stable. Namely, starting from an arbitrary initial condition, the solution to the ODE converges to this fixed point as $t \to \infty$. We also show that the queue length distribution corresponding to the fixed point has tails which decay super-exponentially extending this well known property of the supermarket model (i.e. $k = 1$) to a general $k < L$. We give explicit upper and lower bounds (cf. Theorem 11) on these tails which are sharp in the sense that they coincide when $k = 1$. Finally, in Theorem 13, we prove an important interchange of limit property. In

(Li et al., 2016), it has been shown that the queue length process Q^n for the n-server system is positive recurrent and, thus, has a unique invariant probability measure. This then implies that the occupancy measure process has a unique invariant distribution. In this work we show that this invariant measure converges to $\delta_{\bar{u}}$ in probability, as $n \to \infty$. Roughly speaking, this result says that the limits $n \to \infty$ and $t \to \infty$ can be interchanged and, in particular, the fixed point of the ODE is a good approximation for the steady state behavior of the occupancy process for large n.

1.2 Notation

The following notation will be used. We will use $\{X_t\}$ and $\{X(t)\}$ interchangeably for stochastic processes. The space of probability measures on a Polish space $\mathbb{S}$, equipped with the topology of weak convergence, will be denoted by $\mathcal{P}(\mathbb{S})$. When $\mathbb{S} = \mathbb{N}_0$ we will metrize $\mathcal{P}(\mathbb{S})$ with the metric d_0 defined as

$$d_0(\mu, \nu) \doteq \sum_{j=0}^{\infty} \frac{|\mu(j) - \nu(j)|}{2^j}, \ \mu, \nu \in \mathcal{P}(\mathbb{N}_0).$$

For $\mathbb{S}$ valued random variables X, X_n, $n \geq 1$, convergence in distribution of X_n to X as $n \to \infty$ will be denoted as $X_n \Rightarrow X$. The Borel σ-field on a Polish space $\mathbb{S}$ will be denoted as $\mathcal{B}(\mathbb{S})$. The space of functions that are right continuous with left limits (RCLL) from $[0, T]$ to $\mathbb{S}$ will be denoted as $\mathbb{D}([0, T] : \mathbb{S})$ and equipped with the usual Skorohod topology. Similarly $\mathbb{C}([0, T] : \mathbb{S})$ will be the space of continuous functions from $[0, T]$ to $\mathbb{S}$, equipped with the uniform topology.

We will usually denote by $\kappa, \kappa_1, \kappa_2, \cdots$, the constants that appear in various estimates within a proof. The values of these constants may change from one proof to another. Cardinality of a finite set A will be denoted as $|A|$. We will denote by $B(r)$ the L^1 ball of radius r centered at the origin in some Euclidean space $\mathbb{R}^d$. The Euclidean norm of a d-dimensional vector or a $d \times d$ matrix will be denoted as $\|\cdot\|$. The linear span of a set $A \subset \mathbb{R}^d$ will be denoted as $\mathrm{Sp}A$. The space of continuous (resp. continuous and bounded) functions from metric space $\mathbb{S}_1$ to $\mathbb{S}_2$ will be denoted as $\mathbb{C}(\mathbb{S}_1 : \mathbb{S}_2)$ (resp. $\mathbb{C}_b(\mathbb{S}_1 : \mathbb{S}_2)$). When $\mathbb{S}_2 = \mathbb{R}$ we sometimes abbreviate this notation and write $\mathbb{C}(\mathbb{S}_1)$ and $\mathbb{C}_b(\mathbb{S}_1)$. For a bounded function $f : \mathbb{S} \to \mathbb{R}$, $\|f\|_\infty \doteq \sup_{x \in \mathbb{S}} |f(x)|$. The space of real valued continuous functions defined on $\mathbb{R}^d$ whose first $k \in \mathbb{N}$ (resp. all) derivatives exist

and are continuous will be denoted $\mathbb{C}^k(\mathbb{R}^d)$ (resp. $\mathbb{C}^\infty(\mathbb{R}^d)$). We denote the subset of $\mathbb{C}^k(\mathbb{R}^d)$ of functions with compact support as $\mathbb{C}^k_c(\mathbb{R}^d)$. Similarly $\mathbb{C}^{1,2}([0,T]\times\mathbb{R}^d)$ denotes the space of functions from $(0,T)\times\mathbb{R}^d$ to $\mathbb{R}$ that are once continuously differentiable in the time coordinate, twice continuously differentiable in the space coordinate, and are such that the function and its derivatives can be continuously extended to $[0,T]\times\mathbb{R}^d$. The space of $m\times n$-dimensional matrices whose entries take values in a set $\mathbb{S}$ will be denoted $\mathbb{M}^{m\times n}(\mathbb{S})$. For $M\in\mathbb{M}^{m\times n}(\mathbb{S})$, $M_{i,j}$ will the denote that entry of M which is in the i-th row and j-th column. The transpose of a matrix M will be denoted as M' and trace of a square matrix M will be denoted as $\mathrm{Tr}(M)$. $\mathbf{1}$ and I will denote the vector of 1's and the identity matrix, respectively, the dimension of which will be context dependent. For a Polish space $\mathbb{S}$ we denote by $\mathcal{M}(\mathbb{S})$ the space of all locally finite measures on $\mathbb{S}$. This space will be equipped with the usual vague topology, namely, the weakest topology such that for every $f\in\mathbb{C}_b(\mathbb{S})$ with compact support,

$$\nu\mapsto\int_{\mathbb{S}}f(u)\nu(du),\ \nu\in\mathcal{M}(\mathbb{S}),$$

is continuous.

Let $\boldsymbol{\ell}_2=\{(a_j)_{j=0}^\infty|\sum_{j=0}^\infty a_j^2<\infty\}$ be the space of square summable real sequences. This space is a Hilbert space with inner product

$$\langle x,y\rangle_2=\sum_{j=0}^\infty x_jy_j.$$

We denote the corresponding norm as $\|\cdot\|_2$. Similarly, $\boldsymbol{\ell}_1=\{(a_j)_{j=0}^\infty|\sum_{j=0}^\infty|a_j|<\infty\}$ and $\|\cdot\|_1$ is the norm on this Banach space. The Hilbert-Schmidt norm of a Hilbert-Schmidt operator A on $\boldsymbol{\ell}_2$ will be denoted $\|A\|_{\mathrm{HS}}$ (cf. Appendix B). We denote by I the identity operator. For a Hilbert Space $\mathbb{H}$, $\mathcal{M}_T^2(\mathbb{H})$ will denote the space of all $\mathbb{H}$-valued continuous, square integrable martingales $M\equiv\{M(t)\}_{0\le t\le T}$, such that $M(0)=0$. For a real number a, $(a)_+$ will denote the positive part of a.

CHAPTER 2
Background and Preliminaries

This chapter contains an introduction to some models used for a variety of communication networks as well as some background on the techniques used to analyze them. In addition, we review some of the related literature on communication networks and weakly interacting particle systems. In Section 2.1, we present an overview of some of the relevant existing work on weakly interacting particle systems in communication networks. These works describe some of the ways in which weakly interacting particle systems are used in modeling communication networks and why such models are useful. Specifically, weakly interacting particle systems suggest simpler models through mean field approximations under certain symmetry conditions on the system. The mean field techniques described are particularly useful for load balancing problems. Section 2.2 is devoted to an overview of the existing relevant work in this area. In Section 2.3, we present a LLN result which can be used to approximate the dynamics of a given system through a set of ODE. Section 2.4 provides an introduction to methods for analyzing the deviations around the LLN. Namely, we provide some of the basic approaches to proving various Central Limit Theorems (CLT) of interest. These approximation techniques allow us to analyze communication networks whose large size make this analysis otherwise intractable. Section 2.5 provides an outline of topics in this dissertation

2.1 Weakly Interacting Particle Systems and Communication Networks

Weakly interacting particle systems are frequently used to model a variety of communication networks (e.g. large server networks, ad-hoc wireless networks, etc.). A typical model will consist of a set of n particles (or nodes), each taking values in some state space (in this dissertation we mainly consider discrete state spaces). The evolution of these particles is described in terms of a Markov process. Roughly speaking, a particle system is "weakly interacting" if

the evolution of a typical particle's state only depends on its own current state and the current empirical measure of the states of all particles in the system. In other words, the dynamics of a given particle only depends on the total number of particles in each state and not on the state of any individual particle (other than itself). This property, together with certain natural symmetry conditions, implies the exchangeability of the system that makes these networks well-suited for mean field approximations. More specifically, if one views the evolution of the system through the empirical measure of particle states, then, under conditions, the system can be approximated by a deterministic evolution equation in the space of probability measures, referred to as the Mckean-Vlasov equation. Under conditions, one can also establish a CLT that says that the appropriately scaled fluctuations from this nominal deterministic evolution equation converge to a Gaussian process which is described through a linear SDE (Shiga and Tanaka, 1985; Sznitman, 1991). These fluid and diffusion approximations can be used to analyze useful properties regarding the system (e.g. performance measures, stability, etc.). Below we discuss several examples of such networks that have been studied in the literature.

In (Gibbens et al., 1990), the authors study a routing scheme used in telecommunication networks called Random Alternative Routing. The authors analyze how this method performs on calls routed along the edges of a complete graph with n nodes. In this setting, the "particles" are the links between the nodes rather than the nodes themselves. Each link can handle a fixed, maximum number of calls at a given time. A natural way to view the state of the system is through the available capacity at each link. It is assumed that calls arrive at each link as a Poisson process and are routed as follows. If a call attempts to use a link which does not have available capacity, *two* more links are chosen uniformly at random and, if there is available capacity at both, the call is routed through that path, otherwise the call is lost. The authors derive a LLN approximation for the system (as $n \to \infty$) and show that the limit ODE has exactly two fixed points. In addition, a diffusion approximation is presented and used to explore the tunneling behavior between the two stable points.

The paper (Hunt and Kurtz, 1994) presents a method of analyzing large loss networks. Specifically, the authors consider a network with J links. Each link j has C_j "circuits". In relation to (Gibbens et al., 1990), a circuit is analogous to a unit of capacity. The paper presents a LLN limit for the system. The form of scaling in this paper is different than the one considered

in (Gibbens et al., 1990), namely the paper considers the limit as the arrival/departure rates and the available capacity go to infinity rather than as the number of links in the network approach infinity.

In (Antunes et al., 2008), a general mathematical model for a class of communication networks is studied. Consider a collection of particles (or nodes) each with a finite amount of space (or capacity). Different types of jobs, each with its own capacity requirement, arrive from the outside and are accepted only if there is sufficient capacity to meet the job's requirement. After an exponential holding time, jobs can either move to another particle or leave the system. The evolution of the available capacity at all nodes in the system can be described through a high dimensional pure jump Markov process. Similar to (Gibbens et al., 1990), the authors present a LLN approximation of the empirical measure process associated with the system and then show that the resulting deterministic system of ODEs has multiple stable points.

Weakly interacting particle systems are a tractable class of models since they can be often approximated by simpler mean field models. One such approximation result, which is closely related to the LLN results described above, has been established in (Graham, 2000) that studies a class of routing schemes for large queuing networks. A sequence of infinite collections of sequences of $\mathcal{S}$-valued random variables is said to be $\mathcal{Q}$-chaotic, where $\mathcal{Q}$ is a probability measure on $\mathcal{S}$, if the joint probability law of any subcollection of sequences of k random variables converges to $\mathcal{Q}^{\otimes k}$ for all $k \geq 1$. Namely, the collection of random variables is asymptotically i.i.d. with probability law $\mathcal{Q}$. Consider a collection of n servers which process jobs at rate μ from their own infinite buffer queues. Jobs arrive in the system at rate λn and each job is immediately routed to the shortest of d randomly chosen queues. It is shown in (Graham, 2000) that, under exchangeability conditions and independence of initial conditions, this system has a "Propagation of Chaos" (initial independence [i.e. chaos of particle states] propagates to later time instants) property. Namely the queue length processes viewed as a collection of $\mathbb{D}(\mathbb{R}_+ : \mathbb{N}_0)$-valued random variables are $\mathcal{Q}$-chaotic for an appropriate probability measure $\mathcal{Q}$ on $\mathbb{D}(\mathbb{R}_+ : \mathbb{N}_0)$ where $\mathbb{D}(\mathbb{R}_+ : \mathbb{N}_0)$ is the space of right continuous functions with left limits from $\mathbb{R}_+$ to $\mathbb{N}_0$ equipped with the usual Skorohod topology.

In (Graham and Robert, 2009) an extension of chaoticity described in the previous paragraph is presented for multi-class systems. Suppose that, if instead of the full collection being

exchangeable, the random variables can be divided into K classes such that there is exchangeability within each class. Stated formally, a system is said to be $\mathcal{Q}_1 \otimes \cdots \otimes \mathcal{Q}_K$-multi-chaotic if the joint probability law of any collection of Km variables, such that m are selected from each exchangeable group, converges to $\mathcal{Q}_1^{\otimes m} \otimes \cdots \otimes \mathcal{Q}_K^{\otimes m}$. The authors establish such a multi-chaoticity property for a class of queueing systems. Using this result they then analyze a model for data transmission.

2.2 Load Balancing

Due to the need to properly design and maintain distributed processor networks and cloud-based storage systems, mechanisms for an efficient allocation of jobs or file requests in such networks has garnered quite a bit of attention in recent years. A typical model of interest is described in terms of asystem of n processors or servers each maintaining its own FIFO queue. A stream of jobs or file requests enters the system and are routed by a centralized dispatcher into one or more of the queues. Ultimately the goal is to study how different routing schemes impact various performance metrics of interest (e.g. mean delay time, queue length distribution, etc.). This class of problems associated with different types of routing schemes is frequently referred to as load balancing. In general, the large size of such networks precludes a direct analysis of such systems so the performance is typically studied in a suitable asymptotic regime. In many settings, by appropriately scaling the system and taking limits (e.g. as the number of servers n tends to infinity), one can establish fluid or diffusion approximations for the desired performance metrics. I now give a brief review of some relevant work but refer the interested reader to (van der Boor et al., 2017) for a more in depth exposition.

The simplest load balancing scheme is random routing. Namely, when a job arrives in the system, the dispatcher sends it to a server which is chosen uniformly at random. Consider the expectation of the empirical measure of queue lengths π^n under the stationary distribution. It can be shown that as $n \to \infty$, if the traffic intensity λ (i.e. the ratio of arrival and departure rates) is less than one, the limiting expectation, which is a deterministic measure on $\mathbb{N}_0$ denoted as ν, has an exponentially decaying tail (i.e. $\nu[k, \infty) \sim \lambda^k$). The paper (Graham, 2000), discussed in Section 2.1, and the papers (Vvedenskaya et al., 1996; Mitzenmacher, 2001) first

analyzed the so-called Power-of-d routing scheme (also known as the Supermarket Model). Under this scheme, at each instant of job arrival, the dispatcher polls d randomly chosen servers and routes the jobs to the server with the shortest queue. The paper (Graham, 2000) establishes a functional law of large numbers for π^n on $\mathbb{D}([0,T]:\mathcal{S})$ in the Power-of-d routing scheme using characterization results for nonlinear martingale problems. In (Graham, 2000; Vvedenskaya et al., 1996; Mitzenmacher, 2001), it is shown that for $d \geq 2$ the corresponding measure ν has tails which decay hyperexponentially, namely $\nu[k,\infty) \sim \lambda^{(d^k-1)/(d-1)}$, which is a vast improvement over the exponential rate for the setting where jobs are routed to servers uniformly at random.

In (Eschenfeldt and Gamarnik, 2015), the authors consider another routing scheme known as Join-the-Shortest Queue (JSQ) in which incoming jobs are simply routed into the shortest available queue. This scheme corresponds to the Power-of-d upon taking $d = n$. However, since d scales with n the asymptotic analysis is quite different. The authors establish fluid and diffusion approximations for the empirical measure JSQ routing policy in the large-system limit under a heavy-traffic scaling. It is shown that probability of having a queue of length larger than one converges to zero as $n \to \infty$. Furthermore, the diffusion limit can be characterized through a two-dimensional diffusion. It follows from this theorem that JSQ produces, asymptotically, the minimal possible wait time. Namely, as the number of servers increases to infinity and the traffic intensity approaches criticality, the proportion of servers with two of more jobs goes to zero and thus all jobs which enter the system are routed to empty servers. The excellent performance of JSQ is counterbalanced by an extremely high overhead cost. The dispatcher must query every server each time a jobs arrives which may be costly in large networks in which jobs are arriving extremely rapidly.

A different class of methods, known as pull based routing schemes have also been studied. Here the dispatcher routes jobs based on information which it receives from the individual servers. One basic example of such a method is the Join-the-Idle-Queue (JIQ) routing scheme. In JIQ, each server notifies the dispatcher when it is empty. The dispatcher then routes incoming jobs to empty servers or, if there are no empty servers, to a server according to some other routing policy. In (Mukherjee et al., 2016b) it is shown that in a similar asymptotic regime (i.e. heavy traffic and large n) JIQ produces the same diffusion limit as JSQ. The difference

between push and pull based schemes is subtle, but by using pull based schemes one can reduce communication overhead while maintaining low wait times. In the JIQ scheme each server needs to communicate to the dispatcher when its buffer is empty. In practice this implies that the number of communications between the dispatcher and the server is of the same order as the total number of arrivals in the system and thus, in terms of communication costs JIQ and JSQ are not very different. The authors of (Mukherjee et al., 2016b) also establish a useful interchange of limits property showing that the steady-state behavior of the n-server system converges to the unique fixed point of the limiting system under a fluid scaling.

While not discussed here we refer the interested reader to (Mitzenmacher, 2001; Bramson et al., 2012; Stolyar, 2015; Graham, 2000; Mukherjee et al., 2016a, 2017) and references therein for further work on load balancing. In the next two sections we summarize the basic LLN and central limit theorems for pure jump Markov processes that are useful for studying asymptotics of weakly interacting particle systems of the form considered in this work.

2.3 Law of Large Numbers

Typically, the first method employed when attempting to describe the evolution of the systems of interest here, as their size becomes large, is to derive a LLN limit for the associated empirical measure. This limit is given in terms of a system of coupled ODE and describes the asymptotic behavior of the system under a fluid scaling. One of the classical works on such limit theory is (Kurtz, 1970) which proves the following result (see Theorem 2.11 therein):

Theorem 1. *Let E be a closed set in $\mathbb{R}^k$ and let, for $n \in \mathbb{N}$, $E_n = E \cap \frac{1}{n}\mathbb{N}_0^k$. Let $\{\mu_n(t)\}_{t\geq 0}$ be a pure jump Markov process with state space E_n and infinitesimal generator $\mathcal{A}_n$, defined as*

$$\mathcal{A}_n f(x) = \lambda_n(x) \int_{E_n} [f(z) - f(x)]\gamma_n(x, dz)$$

where $\lambda_n : E_n \to \mathbb{R}_+$ and γ_n is a transition probability kernel on E_n. Define

$$F_n(x) = \lambda_n(x) \int_{E_n} (z - x)\gamma_n(x, dz). \tag{2.1}$$

Suppose the following conditions are met:

i) There exists a Lipschitz function $F : \mathbb{R}^k \to \mathbb{R}^K$ *such that*

$$\lim_{n\to\infty} \sup_{x\in E_n} |F_n(x) - F(x)| = 0.$$

ii) $\lim_{n\to\infty} \mu_n(0) = x_0$ *for some* $x_0 \in E$.

Let μ *be the solution to the following ODE*

$$\dot{\mu}(s) = F(\mu(s)), \qquad \mu(0) = x_0.$$

Then for every $\delta > 0$ *and* $t > 0$

$$\lim_{n\to\infty} \mathbb{P}\{\sup_{s\le t} |\mu_n(s) - \mu(s)| > \delta\} = 0.$$

This theorem says that for a sequence of K-component jump Markov processes, if the function F_n in (2.1) obtained from the generator of the process converges in a suitable manner, then the sequence of processes converge to the solution of a system of ODE. To see how this result applies to weakly interacting systems note that for a typical sequence of communication networks considered in our work the n-th state process is n-dimensional. In particular the dimension of the state space is increasing with n. In order to arrive at a sequence of processes with a common state space we instead view the system through its empirical measure process which will have a finite state space if each particle's state space is finite. In our work we will usually apply Theorem 1 (or a generalization in the case that the state space is countably infinite) to this empirical measure process. In general, an empirical measure process constructed from an n-dimensional Markov process may not itself be Markovian. However, under the symmetry properties of the models considered in this work, the Markov property of the empirical process will indeed hold which will allow the use of Theorem 1.

2.4 Diffusion Approximations

After obtaining a LLN of the form considered in Section 2.3, it is natural to consider the fluctuations around this limit. More precisely, with μ_n and μ as in Theorem 1, we will be

interested in the asymptotic behavior of the stochastic process V_n defined as

$$V_n(t) = \sqrt{n}(\mu_n(t) - \mu(t)), \quad t \geq 0.$$

Under conditions, this asymptotic behavior can be characterized in terms of a suitable diffusion process. One natural approach for proving such a limit theorem is to describe the evolution of the centered and scaled process V_n through a collection of appropriate time changed Poisson processes (see e.g. (3.6) in Chapter 3). Using this description, one can give a semimartingale representation for V_n of the following form

$$V_n(t) = V_n(0) + \int_0^t A_n(s, V_n(s))ds + M_n(t) + o_p(1)$$

where M_n is a local martingale with respect to a suitable filtration and $A_n : [0, \infty) \times \mathbb{R}^k \to \mathbb{R}^k$ is a suitable map.

The first key step in proving the convergence to a diffusion process is to argue tightness of (V_n, M_n). Next, one needs to argue that every weak limit point (V, M) satisfies a stochastic equation of the form

$$V(t) = \int_0^t a(s, V(s))ds + M(t), \quad M(t) = \int_0^t \sigma(s)dB(s), \tag{2.2}$$

where a, σ are suitable maps, B is a Brownian motion with respect to a suitable filtration and V is a continuous process adapted to the filtration. The final step is to argue the uniqueness of solutions to the stochastic equation (2.2). This progression of arguments can be carried out under quite general conditions on the model (see e.g. (Joffe and Métivier, 1986)).

2.5 Overview & Organization

This dissertation is organized as follows. In Chapter 3, we study a class of control problems for models arising from ad hoc wireless networks that are described through certain weakly interacting particle systems. In a typical setting of interest, a system is designed to produce a desired nominal state trajectory. However, due to various approximations and sources of randomness, the actual system performance may deviate significantly from the desired nominal

state. We consider a formulation where the system manager, by adjusting various rates, can nudge the actual stochastic system closer to the desired nominal state. However, exercising control of rates incurs a cost and one needs to suitably balance this cost with the cost of deviating from the desired behavior. Theory of stochastic control gives a natural framework for analyzing such processes. For large systems, solving such stochastic control problems directly is intractable. In this work, we instead consider an approximate approach. Specifically, we introduce a diffusion control problem which approximates the control problem of interest under a suitable scaling. Such diffusion control problems have been well studied and there exists an extensive literature on numerical methods for finding solutions (see e.g. (Kushner and Dupuis, 2013)). Our main result (Theorem 2) shows how an analysis of this diffusion control problem leads to construction of an asymptotically optimal control policy for the system of interest. A paper (Budhiraja et al., 2018) has appeared in the *Annals of Applied Probability*.

In Chapter 4, we study a class of load balancing policies for a large cloud-storage system. In such networks, files are often "coded" across servers to increase both reliability and retrieval speed. By coded, we mean that the file is broken down into smaller pieces which are stored on multiple servers. In the model considered here, we consider a Maximum Distance Separable (MDS) code. Namely, the files chunks are distributed across a set of L servers such that any subset of size k is sufficient for reconstructing the original file. In this work we are interested in developing a rigorous limit theory for such load balancing schemes for systems with MDS coding as the number of servers becomes large and the arrival rate of file requests approaches infinity. Specifically, we establish law of large numbers and diffusion approximations (cf. Theorem 10 and Theorem 14) for such systems under an appropriate scaling, as $n \to \infty$. Such limit theorems provide useful model simplifications that can then be employed for approximate simulation of the large and complex n-server systems. We also study the long-time behavior of the system under fluid scaling. In Theorem 13, we establish a useful interchange of limits property. Namely, that the steady-state distribution of the finite system converges to a dirac measure concentrated at the fixed point of the limiting ODE. This fixed point will be the probability measure representing the distribution on queue lengths in the network in the "steady state". We provide explicit upper and lower bounds on the tail decay of this fixed point (cf. Theorem

11). The results of this work have been submitted for publication ((Budhiraja and Friedlander, 2017; Friedlander, 2018)).

CHAPTER 3

Diffusion Approximations for Controlled Weakly Interacting Large Finite State Systems with Simultaneous Jumps

In this chapter we study a pure jump, weakly interacting, Markovian particle system in which jump rates can be dynamically modulated by a controller. The stochastic system of interest describes the state evolution of a collection of n particles where each particle's state takes values in a finite set $\mathbb{X}$. In many applications, the jump rate of such a system scales with n and thus, for large n, the system jumps extremely quickly. Constructing and implementing a control policy in such a system is intractable. The goal of this chapter is to provide an approximate method for constructing asymptotically optimal control policies. In our main result, Theorem 2, we show that, when n is sufficiently large, one can solve an associated diffusion control problem instead of the control problem for the n-particle system. Specifically, this theorem says that the value function associated with the control problem for the n-th system converges to the value function of the limit diffusion control problem. The key ingredients in the proof are Theorems 3, 7, and 9. Theorem 3 gives the lower bound, namely it shows that the value function of the n-th system, asymptotically as $n \to \infty$, is bounded below by the value function of the diffusion control problem. The key steps in the proof are to establish suitable tightness properties of the sequence of scaled state and control processes and the characterization of the weak limit points. For the first step it is convenient to work with the relaxed control formulation (cf. (Kushner, 2013; Borkar, 1989)) through which one can view controls as elements of a tractable Polish space. The second step proceeds via classical martingale problem techniques (cf. (Stroock and Varadhan, 2007; Ethier and Kurtz, 2009; Joffe and Métivier, 1986)). Theorems 7 and 9 give the main steps needed for the complementary upper bound. For this bound, the main idea is to show that for any fixed $\varepsilon > 0$, there exists an ε-optimal *continuous* feedback control for the diffusion control problem (Theorem 9), and that any such feedback control can be used to construct a sequence of control policies for the interacting

particle system such that the associated costs converge to the cost under the feedback policy for the diffusion control problem (Theorem 7). We begin, in Theorem 8, by arguing that for the diffusion control problem the infimum over all admissible controls is the same as that over the class of feedback controls. Proof of this proceeds via certain conditioning arguments and PDE characterization results (cf. (Borkar, 1989)) that allow the construction of a feedback control associated with any given admissible control such that the cost corresponding to the feedback control is no larger than that of the given admissible control. The result says that one can find an ε-optimal control in the space of feedback controls. Although any such control corresponds to a natural collection of control policies for the sequence of n-particle systems, in order to prove the convergence of associated costs, which once more is based on martingale problem methods, we require additional regularity properties of the feedback control. The key step is Theorem 9 that shows that for any feedback control g there exists a sequence of continuous feedback controls $\{g_n\}$ for the limit diffusion control problem such that the associated sequence of controlled diffusions converge weakly to the diffusion under the feedback control g. The proof requires some estimates based on an application of Girsanov's theorem which, in turn, relies on the non-degeneracy of the diffusion coefficient. Although the controlled diffusion that describes the asymptotic model is degenerate, we show that there is an equivalent formulation in terms of a $(d-1)$-dimensional controlled diffusion which is uniformly non-degenerate under suitable assumptions. This equivalent representation, in addition to providing a feedback control of the desired form, is also key in proving weak uniqueness for SDE describing limit state processes associated with feedback controls.

The chapter is organized as follows. Section 3.1 presents the precise system of weakly interacting pure jump processes considered here. We will also present key assumptions and the main result of this chapter. Sections 3.1.1 and 3.1.2 describe the uncontrolled and controlled systems, respectively. Assumptions which ensure convergence of the system to its fluid limit are introduced for both cases. Section 3.1.2 also introduces the cost criteria that is considered in this chapter. Section 3.1.3 presents the diffusion control problem that formally corresponds to the limit as $n \to \infty$ of the control problem for the n-th system. The section also introduces the key non-degeneracy assumption (Condition 3.1.5) that is needed in order to obtain weak uniqueness of SDE with feedback controls and existence of near optimal continuous feedback controls. We

also introduce our main assumptions on the controlled rate functions (Conditions 3.1.3 and 3.1.4). In Section 3.1.4 we present our main result, namely Theorem 2. In order to validate the results of this chapter, we present a numerical example in Section 3.6. This example is the controlled analogue of a model introduced in (Antunes et al., 2008). The running cost function we consider is quadratic in the normalized state and control processes. The corresponding limit diffusion control problem in this case becomes the classical stochastic LQR with time dependent coefficients (see (Fleming and Rishel, 1976)). The remainder of the chapter is devoted to proof of Theorem 2. In Section 3.2 we present a key tightness result which is used both in the proof of the upper and lower bound. In Section 3.3 (see Theorem 3) we prove the lower bound that was discussed earlier. In preparation for the proof of the upper bound, we introduce the class of feedback controls in Section 3.4. Sections 3.4.1 and 3.4.2 describe such controls for the prelimit system and the limit diffusion model, respectively. Section 3.4.3 constructs a sequence of prelimit control policies from an arbitrary continuous feedback control for the diffusion control problem such that the cost for the particle systems under the sequence of control policies converges to the cost of the corresponding controlled diffusion. Finally in Section 3.5, we show that the infimum of the cost for the limit diffusion over all admissible controls is the same as that over the class of feedback controls and that there exist continuous feedback controls which are ε-optimal. The results from sections 3.3, 3.4, and 3.5, (namely Theorems 3, 7, and 9) together give our main result, Theorem 2.

3.1 Problem Formulation and Main Results

In this section we will describe the basic control problem of interest and give a precise mathematical formulation. We begin by introducing the uncontrolled pure jump Markov process in Section 3.1.1 and recall a classical law of large numbers result for such systems. Section 3.1.2 will present the controlled system that we study and also our cost criteria. In Section 3.1.3 we will introduce our main assumptions on the controlled rate matrices and based on these assumptions introduce a control problem for diffusion processes that can formally be regarded as the limit of control problems considered in Section 3.1.2. Finally, in Section 3.1.4 we present our main result. This result says in particular that a suitable near optimal diffusion control

can be used to construct a sequence of control policies for the particle system in Section 3.1.2 that are asymptotically near optimal. For a numerical example that illustrates the application of the result, we refer the reader to Section 3.6 where we present a model from communication networks that is a controlled version of some models introduced in (Antunes et al., 2008) and which falls within the framework considered here.

3.1.1 Weakly Interacting Jump Markov Process

Fix $T \in (0, \infty)$. All stochastic processes in this chapter will be considered on the time horizon $[0, T]$. Consider a system of n particles where the state of each particle takes values in the set $\mathbb{X} = \{1, \ldots, d\}$. The evolution of the system is described by an n-dimensional pure jump Markov process $\mathbf{X}_n(t) = \{X_n^1(t), \ldots, X_n^n(t)\}$ where $X_n^i(t)$ represents the state of particle i at time t. The system allows multiple particles to change state at a given time, but restricts such jumps to K transition types; in particular the k-th transition type can only affect at most n_k particles, $k \in \mathbf{K} \doteq \{1, \ldots, K\}$. The jump intensity is state dependent, however the state dependence is of the following specific form: Denoting for $x \in \mathbb{X}^n$, the probability measure $\{\frac{1}{n}\sum_{i=1}^n 1_{\{x_i\}}(m)\}_{m\in\mathbb{X}}$ on $\mathbb{X}$ by $\{\zeta_n^m(x)\}_{m\in\mathbb{X}}$, the jump intensity at the instant t is a function of $\zeta_n(X_n(t))$. The set of jumps and the corresponding transition rates can be described in terms of the subset $\mathbb{M}_n$ of $\mathbb{M}^{d\times d}(\mathbb{N}_0)$ consisting of all matrices with zeroes on the diagonal and with sum of all entries at most n, as follows. To any $k \in \mathbf{K}$ we associate a map $\Psi_n^k : \mathcal{P}(\mathbb{X}) \times \mathbb{M}_n \to \mathbb{R}_+$ such that for $x \in \mathbb{X}^n$, $\Psi_n^k(\zeta_n(x), \Theta) = 0$ if

$$\sum_{i,j} \Theta_{i,j} > n_k \text{ or } \sum_{j=1}^d \Theta_{i,j} > n\zeta_n^i(x), \ i = 1, \ldots, d. \tag{3.1}$$

Roughly speaking, $\Psi_n^k(\zeta_n(x), \Theta)$ will give the rate of type k jumps (associated with Θ) when the system is in state $x \in \mathbb{X}^n$. A type k jump associated with $\Theta \in \mathbb{M}_n$ corresponds to Θ_{ij} particles simultaneously jumping from state i to state j, for all $i \neq j$ and $i, j = 1, \ldots, d$. Thus the first inequality in (3.1) says that at most n_k particles change states under a jump of type k, while the second inequality says that a jump of type k can occur only when there are enough particles to participate in it. In terms of Ψ_n^k the overall rate of jumps of type k associated with

Θ, when the system is in state $x \in \mathbb{X}^n$, is given as

$$\Psi_n^k(\zeta_n(x), \Theta) \prod_{m=1}^{d} \binom{n\zeta_n^m(x)}{\sum_{j=1}^d \Theta_{m,j}} \binom{\sum_{j=1}^d \Theta_{m,j}}{\Theta_{m,1}, \ldots, \Theta_{m,d}}$$

and such a jump takes a state $x \in \mathbb{X}^n$ to a state $\tilde{x} \in \mathbb{X}^n$ where

$$n\zeta_n^m(\tilde{x}) = n\zeta_n^m(x) + \sum_{i=1}^{d} \Theta_{i,m} - \sum_{j=1}^{d} \Theta_{m,j},\ m = 1, \ldots, d.$$

A more convenient description of this system is given through the pure jump Markov process $\{\mu_n(t)\}$ where $\mu_n(t) \doteq \zeta_n(X_n(t))$ represents the empirical measure of the particle states. We will identify the space of probability measures, $\mathcal{P}(\mathbb{X})$, with the d-dimensional simplex, $\mathcal{S} \doteq \{(x_1, \ldots, x_d) \in \mathbb{R}_+^d | \sum_{i=1}^d x_i = 1\}$. Similarly, we will identify $\mathcal{P}_n(\mathbb{X})$, the space of all $\mu \in \mathcal{P}(\mathbb{X})$ such that $\mu\{j\} \in \frac{1}{n}\mathbb{N}$ for all $j \in \mathbb{X}$, with $\mathcal{S}_n = \mathcal{S} \cap \frac{1}{n}\mathbb{N}^d$. Let, for $k \in \mathbf{K}$,

$$\Delta^k \doteq \left\{ (I, J) \in \mathbb{N}_0^d \times \mathbb{N}_0^d : \quad \sum_{x \in \mathbb{X}} I_x = \sum_{x \in \mathbb{X}} J_x \leq n_k, \quad \sum_{x \in \mathbb{X}} |J_x - I_x| > 0 \right\},$$

and for $\nu = (I, J) \in \Delta^k$ let

$$\Phi(\nu) = \Phi(I, J) \doteq \left\{ \Theta \in \mathbb{M}_n \Big| \sum_{j=1}^{d} \Theta_{i,j} = I_i, \sum_{i=1}^{d} \Theta_{i,j} = J_j,\ i, j = 1, \cdots, d \right\}.$$

The jumps of $\{\mu_n(t)\}$ are described as follows. For each $k \in \mathbf{K}$ and $\nu = (I, J) \in \Delta^k$ the empirical measure jumps from $r \mapsto r + \frac{1}{n} e_\nu$ with rate

$$\bar{\Gamma}_n^k(r, \nu) \doteq \sum_{\Theta \in \Phi(\nu)} \Psi_n^k(r, \Theta) \prod_{m=1}^{d} \binom{nr^m}{\sum_{j=1}^d \Theta_{m,j}} \binom{\sum_{j=1}^d \Theta_{m,j}}{\Theta_{m,1}, \ldots, \Theta_{m,d}}$$

where $r = (r^m)_{m=1}^d \in \mathcal{S}_n$, $e_\nu \doteq \sum_{x \in \mathbb{X}} (J_x - I_x) e_x$ and e_x is the unit vector in $\mathbb{R}^d$ with 1 at the x-th coordinate and 0 everywhere else. Thus a jump associated with $k \in \mathbf{K}$ and $\nu \in \Delta^k$ corresponds to I_x particles in state x, $x \in \mathbb{X}$, simultaneously jumping to new states such that J_y of the particles end up in state y, $y \in \mathbb{X}$. A succinct description of the evolution of the Markov process

$\mu_n(t)$ is through its infinitesimal generator which is given as

$$\bar{L}^n f(r) = \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \bar{\Gamma}_n^k(r,\nu)\left[f\left(r+\frac{1}{n}e_\nu\right) - f(r)\right], \qquad r \in \mathcal{S}_n. \tag{3.2}$$

We will make the following assumption on the asymptotic behavior of the rates.

Condition 3.1.1. For all $k \in \mathbf{K}$ and $\nu \in \Delta^k$ there exists a Lipschitz function $r \mapsto \Gamma^k(r,\nu)$ on $\mathcal{S}$ such that

$$\limsup_{n\to\infty} \sup_{r\in\mathcal{S}_n} \left|\frac{1}{n}\bar{\Gamma}_n^k(r,\nu) - \Gamma^k(r,\nu)\right| = 0 \tag{3.3}$$

We now present a classical law of large numbers result that characterizes the limit, $\mu(t)$, of the pure jump Markov process $\mu_n(t)$ as $n \to \infty$. For a proof we refer the reader to Theorem 2.11 of (Kurtz, 1970).

Proposition 1. *Define,*

$$F(r) \doteq \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma^k(r,\nu)e_\nu, \ r \in \mathcal{S}. \tag{3.4}$$

Suppose that $\mu_n(0) \to \mu_0$ in probability and Condition 3.1.1 holds, then $\mu_n(t) \to \mu(t)$ uniformly on $[0,T]$, in probability, where $\mu(t)$ is the unique solution of the ODE

$$\dot{\mu}(t) = F(\mu(t)), \ \mu(0) = \mu_0. \tag{3.5}$$

3.1.2 Controlled System

In this chapter we will study a controlled version of the Markov process introduced in Section 3.1.1. Roughly speaking, control action will allow perturbations of the rate function $\bar{\Gamma}_n^k$ that are of $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$. The goal of the controller is to minimize a suitable finite time horizon cost. A precise mathematical formulation is as follows. Let

$$\ell \doteq \sum_{k\in\mathbf{K}} |\Delta^k|, \tag{3.6}$$

Λ be a compact convex subset of $\mathbb{R}^\ell$, and $\Lambda_n = \frac{1}{\sqrt{n}}\Lambda$ for $n \in \mathbb{N}$. Λ_n will be the control set in the n-th system. Let $\{\Gamma_n^k(r,u,\nu) : r \in \mathcal{S}_n, u \in \Lambda_n, k \in \mathbf{K}, \nu \in \Delta^k\}$ be a collection of non-negative real numbers. More precisely, $(r,u) \mapsto \Gamma_n^k(r,u,\nu)$ is a map from $\mathcal{S}_n \times \Lambda_n$ to $\mathbb{R}_+$ for each $n \in \mathbb{N},\ k \in \mathbf{K},\ \nu \in \Delta^k$. These correspond to the controlled rates in the n-th system. We now introduce the controlled stochastic processes associated with such controlled rates.

Fix $n \in \mathbb{N}$ and let $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$ be a probability space on which are defined unit rate mutually independent Poisson processes $\{\mathcal{N}_{k,\nu},\ k \in \mathbf{K},\ \nu \in \Delta^k\}$. The processes $\{\mathcal{N}_{k,\nu}\}$ will be used to describe the stream of jumps corresponding to $k \in \mathbf{K},\ \nu \in \Delta^k$. Let U^n be a Λ_n-valued measurable process representing the rate control in the system. Under control U^n the state process $\mu_n(\cdot)$ is given by the following equation:

$$\mu_n(t) = \mu_n(0) + \frac{1}{n}\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} e_\nu \mathcal{N}_{k,\nu}\left(\int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu)ds\right). \tag{3.7}$$

In order for such a control to be admissible it should satisfy suitable non-anticipative properties. More precisely, U^n is said to be an admissible control if, with some filtration $\{\mathcal{F}_t^n\}$ on $(\Omega^n, \mathcal{F}^n, \mathbb{P}^n)$, U^n is $\{\mathcal{F}_t^n\}$-progressively measurable, μ_n is $\{\mathcal{F}_t^n\}$-adapted, and $\{M_{k,\nu}^n, k \in \mathbf{K}, \nu \in \Delta^k\}$ defined below are $\{\mathcal{F}_t^n\}$-martingales

$$M_{k,\nu}^n(t) \doteq \frac{1}{n}\left(\mathcal{N}_{k,\nu}\left(\int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu)ds\right) - \int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu)ds\right) \tag{3.8}$$

with quadratic variation processes $\langle M_{k,\nu}^n, M_{k',\nu'}^n\rangle_t = \delta_{(k,\nu),(k',\nu')}\frac{1}{n^2}\int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu)ds$ where $\delta_{\alpha,\alpha'}$ equals 1 if $\alpha = \alpha'$ and 0 otherwise. We note that in general such a filtration will depend on the control. We denote the set of all such admissible controls as $\mathcal{A}_n$.

For a $U^n \in \mathcal{A}_n$, define the process

$$V_n(s) = \sqrt{n}(\mu_n(s) - \mu(s)) \tag{3.9}$$

where, as above, μ_n is the state process under control U^n. We consider a cost that is a function of the suitably normalized control action and the centered and normalized state of the system given through the process $\{V_n(\cdot)\}$. Specifically, we consider for $n \in \mathbb{N},\ x_n \in \mathcal{S}_n$ a "finite time

horizon cost" associated with an admissible control $U^n \in \mathcal{A}_n$ and initial condition x_n as,

$$J_n(U^n, v_n) \doteq \mathbb{E} \int_0^T (k_1(V_n(s)) + k_2(\sqrt{n}U^n(s)))ds \tag{3.10}$$

where $v_n = \sqrt{n}(x_n - \mu_0)$, $k_2 \in \mathbb{C}(\Lambda)$ is a nonnegative convex function, and $k_1 \in \mathbb{C}(\mathbb{R}^d)$ is a nonnegative function with at most polynomial growth. I.e. there exists a $p > 1$ and $C_{k_1} \in (0, \infty)$ such that $k_1(x) \leq C_{k_1}(1 + \|x\|^p)$ for all $x \in \mathbb{R}^d$. Define the corresponding value function to be

$$R_n(v_n) \doteq \inf_{U^n \in \mathcal{A}_n} J_n(U^n, v_n).$$

Computing an optimal control for the above problem for a given n is, in general, challenging and computationally intensive. It is therefore of interest to consider approximate approaches. In the next section we introduce some conditions on the controlled rate matrices that will suggest a natural diffusion approximation for this control problem.

3.1.3 Diffusion Control Problem

We now introduce our main assumptions on the controlled rate matrices. The first two conditions make precise the requirement that controlled rates are $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$ perturbations of the nominal values given through $\{\Gamma^k,\ k \in \mathbf{K}\}$. In particular, the first condition will ensure that the controlled pure jump Markov process will converge to the same limit as the uncontrolled process μ_n in Section 3.1.1 under the law of large number scaling.

Condition 3.1.2. With $\{\Gamma^k(r, \nu),\ k \in \mathbf{K},\ \nu \in \Delta^k,\ r \in \mathcal{S}\}$ as in Condition 3.1.1

$$\limsup_{n\to\infty} \sup_{r\in\mathcal{S}_n} \sup_{u\in\Lambda_n} \left|\frac{1}{n}\Gamma_n^k(r, u, \nu) - \Gamma^k(r, \nu)\right| = 0. \tag{3.11}$$

We next introduce a strengthening of Condition 3.1.2 that will play a key role in the proof of tightness of the sequence $\{V_n\}$ of controlled state processes.

Condition 3.1.3. There exists a $C_1 \in (0, \infty)$ such that for every $n \in \mathbb{N}$

$$\sup_{u\in\Lambda_n} \sup_{\xi\in\mathcal{S}_n(y)} \sqrt{n}\left|\frac{1}{n}\Gamma_n^k\left(\frac{1}{\sqrt{n}}y + \xi, u, \nu\right) - \Gamma^k(\xi, \nu)\right| \leq C_1(1 + \|y\|) \tag{3.12}$$

for all $k \in \mathbf{K},\ \nu \in \Delta^k$, and $y \in B(2\sqrt{n}) \subset \mathbb{R}^d$ where $\mathcal{S}_n(y) = \{\xi \in \mathcal{S} : \frac{1}{\sqrt{n}} y + \xi \in \mathcal{S}_n\}$.

Taking $y = 0$ in (3.12) we see that Condition 3.1.3 implies that there exists a $C_2 \in (0, \infty)$ such that

$$\sup_{n \geq 1} \sup_{r \in \mathcal{S}_n} \sup_{u \in \Lambda_n} \frac{1}{n} \Gamma_n^k(r, u, \nu) \leq C_2 \tag{3.13}$$

for all $k \in \mathbf{K},\ \nu \in \Delta^k$. Note also that Condition 3.1.3 implies Condition 3.1.2.

The next condition will identify the drift term in our limit diffusion control problem. Note that any $u \in \Lambda$ (or Λ_n) can be indexed by $k \in \mathbf{K}$ and $\nu \in \Delta^k$ and we will denote the corresponding entry by $u_{k,\nu}$.

Condition 3.1.4. There exist, for each $k \in \mathbf{K}, \nu \in \Delta^k$, bounded functions $h_1^k(\nu, \cdot) : \mathcal{S} \to \mathbb{R}$ and $h_2^k(\nu, \cdot) : \mathcal{S} \to \mathbb{R}^d$ such that for $u \in \Lambda$, $\xi \in \mathcal{S}$, $y \in \mathbb{R}^d$, with

$$H^k(y, \xi, u, \nu) \doteq h_1^k(\nu, \xi) u_{k,\nu} + h_2^k(\nu, \xi) \cdot y,$$

we have for all compact $A \subset \mathbb{R}^d$,

$$\limsup_{n \to \infty} \sup_{u \in \Lambda} \sup_{y \in A} \sup_{\xi \in \mathcal{S}_n(y)} |\beta_k^n(y, \xi, u, \nu)| = 0 \tag{3.14}$$

where for $n \in \mathbb{N}, k \in \mathbf{K}$, and $\nu \in \Delta^k$, we define $\beta_k^n(\cdot, \cdot, \cdot, \nu) : \mathbb{R}^d \times \mathcal{S} \times \Lambda \to \mathbb{R}$ as

$$\beta_k^n(y, \xi, u, \nu) \doteq \sqrt{n} \left(\frac{1}{n} \Gamma_n^k \left(\frac{1}{\sqrt{n}} y + \xi, \frac{1}{\sqrt{n}} u, \nu \right) - \Gamma^k(\xi, \nu) \right) - H^k(y, \xi, u, \nu),$$

if $\xi \in \mathcal{S}_n(y)$ and 0 otherwise.

Define $\eta : [0, T] \times \mathbb{R}^\ell \to \mathbb{R}^d$ and $\beta : [0, T] \to \mathbb{R}^{d \times d}$ as

$$\eta(t, u) \doteq \sum_{k \in \mathbf{K}} \sum_{\nu \in \Delta^k} \left(h_1^k(\nu, \mu(t)) u_{k,\nu} \right) e_\nu \text{ and } \beta(t) \doteq \sum_{k \in \mathbf{K}} \sum_{\nu \in \Delta^k} e_\nu [h_2^k(\nu, \mu(t))]' \tag{3.15}$$

Note that

$$\sum_{k \in \mathbf{K}} \sum_{\nu \in \Delta^k} H^k(y, \mu(t), u, \nu) e_\nu = \eta(t, u) + \beta(t) y,\ \ t \in [0, T], y \in \mathbb{R}^d. \tag{3.16}$$

Let $a : [0,T] \to \mathbb{R}^{d\times d}$ be defined as

$$a(t) \doteq \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} (\Gamma^k(\mu(t),\nu))e_\nu e_\nu'.$$

The $d \times d$ matrix $a(t)$ will be the square of the diffusion coefficient for the limit controlled diffusion process. Note that $a(t)$ is a singular matrix since $e_\nu \cdot \mathbf{1} = 0$ for all $k \in \mathbf{K}$ and $\nu \in \Delta^k$. Let $Q = [q_1 \dots q_d]$, $q_k \in \mathbb{R}^d$, be a $d\times d$ orthogonal matrix (i.e $QQ' = Q'Q = I$) such that $q_d = \frac{1}{\sqrt{d}}\mathbf{1}$. Then, in view of the above observation,

$$Q'a(t)Q = \begin{pmatrix} \alpha(t) & 0 \\ 0 & 0 \end{pmatrix} \tag{3.17}$$

where $\alpha(\cdot)$ is a Lipschitz, nonnegative definite, $(d-1)\times(d-1)$ matrix valued function. Let $\alpha^{1/2}(t)$ be the symmetric square root of $\alpha(t)$. Since $t \mapsto \alpha(t)$ is continuous so is $t \mapsto \alpha^{1/2}(t)$ (see e.g. (Chen and Huan, 1997)). Define

$$\sigma(t) \doteq Q\begin{bmatrix} \alpha^{1/2}(t) & 0 \\ 0 & 0 \end{bmatrix} Q'. \tag{3.18}$$

The main goal of this paper is to show that an optimal control problem for certain diffusion processes can be used to construct asymptotically near optimal control policies for the sequence of controlled systems in Section 3.1.2. We now introduce this diffusion control problem. Let $(\Omega,\mathcal{F},\mathbb{P},\{\mathcal{F}_t\})$ be a filtered probability space with a d-dimensional $\{\mathcal{F}_t\}$-Brownian motion $\{W_t\}$. We refer to $(\Omega,\mathcal{F},\mathbb{P},\{\mathcal{F}_t\},\{W_t\})$ as a system and denote it by Ξ. Denote the collection of $\mathcal{F}_t$-progressively measurable, Λ valued processes as $\mathcal{A}(\Xi)$. This collection will represent the set of admissible controls for the diffusion control problem. The initial condition v_0 for our controlled diffusion process will lie in the set $\mathbb{V}_{d-1} = \{x \in \mathbb{R}^d | x \cdot \mathbf{1} = 0\}$. For $U \in \mathcal{A}(\Xi)$ and $v_0 \in \mathbb{V}_{d-1}$, let V be the unique pathwise solution of

$$V(t) = v_0 + \int_0^t \eta(s,U(s))ds + \int_0^t \beta(s)V(s)ds + \int_0^t \sigma(s)dW(s) \tag{3.19}$$

where η, β are as introduced in (3.15) and σ is as in (3.18). Define the cost associated with $U \in \mathcal{A}(\Xi)$ and $v_0 \in \mathbb{V}_{d-1}$ as

$$J(U, v_0) \doteq \mathbb{E} \int_0^T (k_1(V(s)) + k_2(U(s)))ds. \tag{3.20}$$

The value function associated with the above diffusion control problem is

$$R(v_0) \doteq \inf_{\Xi} \inf_{U \in \mathcal{A}(\Xi)} J(U, v_0),$$

where the outside infimum is taken over all possible systems Ξ.

Although the matrix $\sigma(t)$ is singular for each t, the following condition will ensure that the dynamics of V restricted to a certain $(d-1)$-dimensional subspace is non-degenerate.

Condition 3.1.5. There exists a $\Delta^* \subset \cup_{k \in \mathbf{K}} \Delta^k$ such that $\mathrm{Sp}\{e_\nu : \nu \in \Delta^*\}$ equals $\mathbb{V}_{d-1}$, and for every $\nu \in \Delta^*$ there is a $k_\nu \in \mathbf{K}$ such that $\nu \in \Delta^{k_\nu}$ and

$$\kappa(T) \doteq \inf_{\nu \in \Delta^*} \inf_{0 \le t \le T} \Gamma^{k_\nu}(\mu(t), \nu) > 0.$$

The following lemma shows that under Condition 3.1.5, α is uniformly non-degenerate on compact sets.

Lemma 1. *Under Condition 3.1.5, $\{\alpha(t) : t \in [0, T]\}$ is a uniformly positive definite collection, namely, there exists a $C(T) \in (0, \infty)$ such that $x'\alpha(t)x \ge C(T)\|x\|^2$ for all $x \in \mathbb{R}^{d-1}$ and $0 \le t \le T$.*

Proof. We first show that the matrix $G = \sum_{\nu \in \Delta^*} e_\nu e_\nu'$ satisfies, for some $C_G \in (0, \infty)$,

$$\xi' G \xi \ge C_G \|\xi\|^2 \tag{3.21}$$

for all $\xi \in \mathbb{V}_{d-1}$. For this it satisfies to check that for any nonzero $\xi \in \mathbb{V}_{d-1}$, $\xi' G \xi > 0$.

Suppose for some nonzero $\xi \in \mathbb{V}_{d-1}$, $\xi' G \xi = 0$. Since $\xi' G \xi = \sum_{\xi \in \Delta^*} |\xi \cdot e_\nu|^2$ and $\mathrm{Sp}\{e_\nu : \nu \in \Delta^*\} = \mathbb{V}_{d-1}$, we must have $\xi \perp \mathbb{V}_{d-1}$. But by assumption ξ is a nonzero element of $\mathbb{V}_{d-1}$ which is a contradiction. This proves (3.21).

Now for $x \in \mathbb{R}^{d-1}$, letting $\hat{x} = \binom{x}{0} \in \mathbb{R}^d$,

$$x'\alpha(t)x = \hat{x}'Q'a(t)Q\hat{x} = (Q\hat{x})'a(t)(Q\hat{x}).$$

Since $\mathbf{1} = \sqrt{d}q_d$ and $\hat{x}_d = 0$,

$$Q\hat{x} \cdot \mathbf{1} = (q_1\hat{x}_1 + \cdots + q_d\hat{x}_d) \cdot \mathbf{1} = (q_1x_1 + \cdots + q_{d-1}x_{d-1}) \cdot \mathbf{1} = 0.$$

Thus $y = Q\hat{x} \in \mathbb{V}_{d-1}$, and consequently for $t \in [0,T]$,

$$\begin{aligned} y'a(t)y &= \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} (\Gamma^k(\mu(t),\nu))y'e_\nu e_\nu'y \\ &\geq \sum_{\nu\in\Delta^*} (\Gamma^{k(\nu)}(\mu(t),\nu))y'e_\nu e_\nu'y \geq \kappa(T)y'Gy \geq \kappa(T)C_G\|y\|^2. \end{aligned}$$

Thus

$$x'\alpha(t)x \geq \kappa(T)C_G\|Q\hat{x}\|^2 = \kappa(T)C_G\|\hat{x}\|^2 = \kappa(T)C_G\|x\|^2 \tag{3.22}$$

and the result follows. □

Since $t \mapsto \alpha(t)$ is Lipschitz, it follows from Lemma 1 that under Condition 3.1.5, $t \mapsto \alpha^{1/2}(t)$ is Lipschitz as well (see Theorem 5.2.2 in (Stroock and Varadhan, 2007)). Note from (3.22), that $x'\alpha^{1/2}(t)x \geq (\kappa(T)C_G)^{1/2}\|x\|^2$ for all $x \in \mathbb{R}^{d\times d}$ and $t \in [0,T]$. In particular

$$\sup_{0\leq t\leq T} \|\alpha^{-1/2}(t)\| < \infty. \tag{3.23}$$

3.1.4 Main Result

We now present the main result of this chapter. In Section 3.4 we will show that for every measurable function $g : [0,T] \times \mathbb{R}^d \to \Lambda$ there exists a system Ξ and a $U_g \in \mathcal{A}(\Xi)$ such that the corresponding controlled diffusion process is a (time inhomogeneous) Markov process with

generator

$$\mathcal{L}_g f(t,x) \doteq \nabla f(x) \cdot [\eta(t, g(t,x)) + \beta(t)x] + \frac{1}{2}\operatorname{Tr}(\sigma(t) D^2 f(x) \sigma'(t)), \quad f \in \mathbb{C}_c^\infty(\mathbb{R}^d) \tag{3.24}$$

where ∇ and D^2 are the gradient and the Hessian operators, respectively. Furthermore, as we will describe in Section 3.4, such a g also defines a control U_g^n in the n-th system, under which the state process μ_n^g is a time inhomogeneous Markov process (see (3.49)). We refer to U_g and U_g^n as the feedback controls associated with g for the diffusion control problem and the n-th controlled system, respectively. The following is the main result of this chapter. It says the following three things: (i) The value functions of the n-particle control problem converge to that of the diffusion control problem as $n \to \infty$; (ii) For every $\varepsilon > 0$, there exists a continuous ε-optimal feedback control for the diffusion control problem; (iii) A near optimal continuous feedback control for the diffusion control problem can be used to construct a sequence of asymptotically near optimal controls for the systems indexed by n.

Theorem 2. *Suppose Conditions 3.1.3, 3.1.4, and 3.1.5 hold. Let $x_n \in \mathcal{S}_n$ be such that $v_n = \sqrt{n}(x_n - x_0) \to v_0$ as $n \to \infty$. Then*

(i) $R_n(v_n) \to R(v_0)$ as $n \to \infty$.

(ii) For every $\varepsilon > 0$, there is a continuous $g_\varepsilon : [0,T] \times \mathbb{R}^d \to \Lambda$ such that

$$J(U_{g_\varepsilon}, v_0) \le R(v_0) + \varepsilon.$$

(iii) For any continuous $g : [0,T] \times \mathbb{R}^d \to \Lambda$, $J_n(U_g^n, v_n) \to J(U_g, v_0)$ as $n \to \infty$. In particular, with g_ε as in (ii),

$$R(v_0) = \lim_{n\to\infty} R_n(v_n) \le \lim_{n\to\infty} J_n(U_{g_\varepsilon}^n, v_n) \le R(v_0) + \varepsilon.$$

Proof. The above result will be proved in three parts. First in Theorem 3 we will show that for all v_n, v_0 as in the statement,

$$\liminf_{n\to\infty} R_n(v_n) \ge R(v_0).$$

Next, Theorem 7 shows the first statement in (iii). Finally in Theorem 9 we prove part (ii) of the theorem.

Combining the above results we see that for each $\varepsilon > 0$

$$\limsup_{n\to\infty} R_n(v_n) \le \lim_{n\to\infty} J(U^n_{g_\varepsilon}, v_n) = J(U_{g_\varepsilon}, v_0) \le R(v_0) + \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary it follows immediately that $\limsup_{n\to\infty} R_n(v_n) \le R(v_0)$ completing the proof of part (i) and also the second statement in (iii). □

Proof of Theorems 3, 7, and 9 are given in Sections 3.3, 3.4, and 3.5, respectively. Section 3.6 of the paper will present an example that is a controlled analogue of systems introduced in (Antunes et al., 2008) as models for ad hoc wireless networks. We will verify Conditions 3.1.3-3.1.5 for this example and describe how results from Theorem 2 can be used to construct a sequence of asymptotically near optimal control policies.

3.2 Tightness

In this section we prove a tightness result which will be needed in the proofs of Theorems 4 and 7. For $u \in \Lambda_n$, $k \in \mathbf{K}$ and $\nu \in \Delta^k$, we extend the map $r \to \Gamma^k_n(r, u, \nu)$ to all of $\mathbb{R}^d$ by setting $\Gamma^k_n(r, u, \nu) = 0$ if $r \notin \mathcal{S}_n$.

For $U^n \in \mathcal{A}_n$ define V_n by (3.9) where μ_n is the controlled Markov process corresponding to the system under control U^n given as in (3.7). Define $\gamma_n : [0,T] \times \mathbb{R}^d \to \mathbb{R}^d$ as $\gamma_n(t,x) \doteq \mu(t) + \frac{1}{\sqrt{n}}x$, for $x \in \mathbb{R}^d$, $t \in [0,T]$ and for $\phi \in \mathbb{C}^2(\mathbb{R}^d)$, $s \in [0,T]$, $u \in \Lambda_n$, and $y \in \mathbb{R}^d$ define

$$\mathcal{L}^n_u(\phi, s, y) \doteq \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma^k_n(\gamma_n(s,y), u, \nu)\left[\phi\left(y + \frac{1}{\sqrt{n}}e_\nu\right) - \phi(y)\right] - \sqrt{n}F(\mu(s))\nabla\phi(y). \quad (3.25)$$

For $i = 1, \ldots, d$ define $\phi^i(y) \doteq y_i$ and denote the i-th coordinate of e_ν and F by e^i_ν and F^i, respectively. Let

$$\begin{aligned} b^{i,u}_n(s,y) \doteq \mathcal{L}^n_u(\phi^i, s, y) &= \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma^k_n(\gamma_n(s,y), u, \nu)\frac{1}{\sqrt{n}}e^i_\nu - \sqrt{n}F^i(\mu(s)) \\ &= \sqrt{n}\sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e^i_\nu\left(\frac{1}{n}\Gamma^k_n(\gamma_n(s,y), u, \nu) - \Gamma^k(\mu(s), \nu)\right) \end{aligned}$$

where the second equality follows from the definition of F in Proposition 1. Also, for $i,j = 1,\dots,d$ let,

$$\begin{aligned}
a_n^{i,j,u}(s,y) &\doteq \mathcal{L}_u^n(\phi^i\phi^j,s,y) - y_i b_n^{j,u}(s,y) - y_j b_n^{i,u}(s,y) \\
&= \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} \Gamma_n^k\left(\gamma_n(s,y),u,\nu\right)\left(\frac{y_i}{\sqrt{n}}e_\nu^j + \frac{y_j}{\sqrt{n}}e_\nu^i + \frac{1}{n}e_\nu^i e_\nu^j\right) \\
&\qquad - y_i\sqrt{n}F^j(\mu(s)) - y_j\sqrt{n}F^i(\mu(s)) - y_i b_n^{j,u}(s,y) - y_j b_n^{i,u}(s,y) \\
&= \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} \Gamma_n^k\left(\gamma_n(s,y)),u,\nu\right)\frac{1}{n}e_\nu^i e_\nu^j.
\end{aligned}$$

We write $b_n^u = (b_n^{1,u},\dots,b_n^{d,u})$ and $a_n^u = (a_n^{i,j,u})_{i,j=1,\dots,d}$.

Let

$$n_{\mathbf{K}} \doteq 2\max_{k\in\mathbf{K}} n_k. \tag{3.26}$$

The following Lemma gives a key bound needed for tightness.

Lemma 2. *Suppose Condition 3.1.3 holds. Then there exists $C_3 \in (0,\infty)$ such that for every $n\in\mathbb{N}$ and $t\in[0,T]$*

$$(\|b_n^{U^n(t)}(t,V_n(t))\|^2 + \mathrm{Tr}(a_n^{U^n(t)}(t,V_n(t)))) \le C_3(1+\|V_n(t)\|^2)$$

almost everywhere for every $U^n\in\mathcal{A}_n$.

Proof. It follows from (3.13) that for $y\in B(2\sqrt{n})$ such that $\mu(t)\in\mathcal{S}_n(y)$, $u\in\Lambda_n$, and $i=1,\dots,d$

$$a_n^{i,i,u}(t,y) = \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} \Gamma_n^k\left(\gamma_n(t,y),u,\nu\right)\frac{1}{n}e_\nu^i e_\nu^i \le \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} C_2 e_\nu^i e_\nu^i \le C_2\ell n_{\mathbf{K}}^2,$$

and from Condition 3.1.3

$$\begin{aligned} b_n^{i,u}(t,y)^2 &= \left(\sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu^i \sqrt{n} \left(\frac{1}{n} \Gamma_n^k \left(\gamma_n(t,y), u, \nu \right) - \Gamma^k(\mu(t), \nu) \right) \right)^2 \\ &\le \left(\sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} |e_\nu^i| C_1 (1 + \|y\|) \right)^2 \\ &\le \left(C_1 \ell n_{\mathbf{K}} (1 + \|y\|) \right)^2 \\ &\le 2 C_1^2 \ell^2 n_{\mathbf{K}}^2 (1 + \|y\|^2). \end{aligned}$$

The result now follows on noting that $V_n(t) \in B(2\sqrt{n})$ and $\mu(t) \in \mathcal{S}_n(V_n(t))$ a.s. □

For $n \ge 1$ and $\phi \in \mathbb{C}^2(\mathbb{R}^d)$, let $\psi_n \in \mathbb{C}^{1,2}([0,T] \times \mathbb{R}^d)$ be defined as

$$\psi_n(t,y) \doteq \phi(\sqrt{n}(y - \mu(t))), \ t \in [0,T], \ y \in \mathbb{R}^d.$$

Note that $\phi(x) = \psi_n(t, \gamma_n(t,x))$. Using (3.7) and Dynkin's formula,

$$\begin{aligned} \phi(V_n(t)) &= \psi_n(t, \mu_n(t)) \\ &= \psi_n(0, \mu_n(0)) + \int_0^t L_{U^n(s)}^n \psi_n(s, \mu_n(s)) ds + \int_0^t \frac{\partial}{\partial s} \psi_n(s, \mu_n(s)) ds + M_t^{n,\phi} \end{aligned} \tag{3.27}$$

where $M_t^{n,\phi}$ is a locally square-integrable martingale and for $u \in \Lambda_n, (s,r) \in [0,T] \times \mathbb{R}^d$,

$$\begin{aligned} L_u^n \psi_n(s,r) &\doteq \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma_n^k(r,u,\nu) \left[\psi_n \left(s, r + \frac{1}{n} e_\nu \right) - \psi_n(s,r) \right] \\ &= \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma_n^k(r,u,\nu) \left[\phi \left(\sqrt{n}(r - \mu(s)) + \frac{1}{\sqrt{n}} e_\nu \right) - \phi(\sqrt{n}(r - \mu(s))) \right]. \end{aligned}$$

Also, since $\dot{\mu}(t) = F(\mu(t))$,

$$\frac{\partial}{\partial s} \psi_n(s,r) = -\sqrt{n} F(\mu(s)) \cdot \nabla \phi(\sqrt{n}(r - \mu(s))).$$

This shows that the process V_n is a $\mathcal{D}$-semimartingale in the sense of Definition 3.1.1 of (Joffe and Métivier, 1986) with increasing function $A(t) = t$ and the associated mapping $\mathbf{L}^n : \mathbb{C}^2(\mathbb{R}^d) \times$

$\mathbb{R}^d \times [0,T] \times \Omega^n \to \mathbb{R}$ (in the notation of (Joffe and Métivier, 1986)) defined as

$$\mathbf{L}^n(\phi, y, t, \omega) \doteq \mathcal{L}^n_{U^n(t,\omega)}(\phi, t, y),$$

where $\mathcal{L}^n_u$ is defined as in (3.25). Furthermore,

$$\mathbf{b}^n_i(y,t,\omega) \doteq b^{i,U^n(t,\omega)}(t,y), \;\; \mathbf{a}^n_{ij}(y,t,\omega) \doteq a^{i,U^n(t,\omega)}_n(t,y),$$

are the local coefficients of first and second order of the semimartingale V_n in the sense of Definition 3.1.2 of (Joffe and Métivier, 1986). In particular, equation (3.27) combined with (3.25) implies that

$$M^n_t \doteq V_n(t) - V_n(0) - \int_0^t \mathbf{b}^n(V_n(s), s, \omega) ds \tag{3.28}$$

is a d-dimensional locally square-integrable martingale.

Definition 3.1. For $x \in \mathbb{D}([0,T] : \mathbb{R}^d)$ let $j_T(x) \doteq \sup_{0<t\leq T} \|x(t) - x(t-)\|$ be the maximum jump size of x. We say a tight collection of $\mathbb{D}([0,T] : \mathbb{R}^d)$-valued random variables $\{X_n\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight if $j_T(X_n) \Rightarrow 0$.

If X_n, X are $\mathbb{D}([0,T] : \mathbb{R}^d)$-valued random variables and $X_n \Rightarrow X$ then $\mathbb{P}(X \in \mathbb{C}([0,T] : \mathbb{R}^d)) = 1$ if and only if $\{X_n\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight (Billingsley, 1999). Using Lemma 2, the following Proposition follows directly from Lemma 3.2.2 and Proposition 3.2.3 of (Joffe and Métivier, 1986).

Proposition 2. *Suppose Condition 3.1.3 holds. Define for $n \in \mathbb{N}$, V_n through (3.9), where μ_n is defined as in (3.7) for some $U^n \in \mathcal{A}_n$. Suppose $V_n(0) = v_n \in \mathbb{R}^d$ and $\sup_n \|v_n\| < \infty$. Then*

$$\sup_{n\geq 1} \mathbb{E} \sup_{0\leq t\leq T} \|V_n(t)\|^2 < \infty$$

and the sequence $\{V_n\}_{n\geq 1}$ is a tight collection of $\mathbb{D}([0,T] : \mathbb{R}^d)$-valued random variables. Furthermore the sequence is $\mathbb{C}$-tight.

Proof. Since $\mathbf{b}^n$ and $\mathbf{a}^n$ are the local coefficients of the semimartingale V_n, the moment bound is immediate from the properties of b_n^u and a_n^u established in Lemma 2 upon using Lemma 3.2.2 of (Joffe and Métivier, 1986). Using this moment bound and Lemma 2 once again, tightness follows from verifying Aldous' tightness criteria (cf. Theorem 2.2.2 in (Joffe and Métivier, 1986)) as in Proposition 3.2.3 of (Joffe and Métivier, 1986). Also note that $\{V_n\}$ is $\mathbb{C}$-tight because $j_T(V_n) \leq \frac{1}{\sqrt{n}} \ell^{1/2} n_{\mathbf{K}}$ where ℓ and $n_{\mathbf{K}}$ are as in (3.6) and (3.26), respectively. □

Remark 3.2.1. Proposition 2 in particular says that under Condition 3.1.3 μ_n converges to μ in $\mathbb{D}([0,T] : \mathbb{R}^d)$.

3.3 Lower Bound

In this section we prove the following result.

Theorem 3. *Suppose Conditions 3.1.3, 3.1.4, and 3.1.5 hold. Let v_n, v_0 be as in the statement of Theorem 2. Then*

$$\liminf_{n\to\infty} R_n(v_n) \geq R(v_0).$$

A key ingredient in the proof of Theorem 3 will be Theorem 4 which is presented below. In order to formulate this we first begin with some notation. Note that the local martingale M^n in (3.28) takes the following explicit form.

$$M^n(t) = \sqrt{n} \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu M^n_{k,\nu}(t),\ t \in [0,T], \tag{3.29}$$

where $M^n_{k,\nu}$ is as defined in (3.8). To see this, denote the right side of (3.29) as $\tilde{M}^n(t)$ and then, using (3.7), we can write

$$\mu_n(t) = \mu_n(0) + \frac{1}{n} \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu \int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu) ds + \frac{1}{\sqrt{n}} \tilde{M}^n(t).$$

From this and recalling the definition of μ from (3.7) and of H^k from Condition 3.1.4, we have the following representation for V_n in terms of $\tilde{M}^n$

$$\begin{aligned} V_n(t) &= \sqrt{n}(\mu_n(t) - \mu(t)) \\ &= v_n + \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu \int_0^t \sqrt{n}\left(\frac{1}{n}\Gamma_n^k(\mu_n(s), U^n(s), \nu) - \Gamma^k(\mu(s), \nu)\right) ds + \tilde{M}^n(t) \\ &= v_n + \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu \int_0^t H^k(V_n(s), \mu(s), \sqrt{n}U^n(s), \nu) ds + \int_0^t \vartheta^n(s) ds + \tilde{M}^n(t) \end{aligned} \tag{3.30}$$

where the error term ϑ^n is given as

$$\vartheta^n(s) = \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \vartheta^n_{k,\nu}(s), \quad \vartheta^n_{k,\nu}(s) = e_\nu \beta^n_k(V_n(s), \mu(s), \sqrt{n}U^n(s), \nu),$$

and β^n_k is as in Condition 3.1.4. This proves (3.29).

Note that ϑ^n can be estimated as

$$\|\vartheta^n(s)\| \le \theta^n(V_n(s)). \tag{3.31}$$

where for $y \in \mathbb{R}^d$

$$\theta^n(y) \doteq (\ell)^{1/2} n_{\mathbf{K}} \sup_{\xi\in\mathcal{S}_n(y)} \sup_{u\in\Lambda} \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} |\beta^n_k(y, \xi, u, \nu)|,$$

with ℓ and $n_{\mathbf{K}}$ as in (3.6) and (3.26), respectively. Condition 3.1.4 then implies

$$\sup_{y\in A} \theta^n(y) \to 0, \text{ as } n \to \infty \tag{3.32}$$

for all compact A. The above estimate will allow us to estimate the error term ϑ^n in (3.30).

In order to have suitable tightness properties of the control processes it will be convenient to introduce the following collection of random measures. Define $\mathcal{M}([0,T]\times\Lambda)$ valued random variables m^n as

$$m^n(A \times B) = \int_A 1_B(\sqrt{n}U^n(s)) ds. \tag{3.33}$$

Note that m^n can be disintegrated as $m_s^n(du)ds$, where $m_s^n(du) = \delta_{\sqrt{n}U^n(s)}(du)$ and δ_x is the Dirac measure at the point x. Then for $s \in [0,T]$,

$$H^k(V_n(s), \mu(s), \sqrt{n}U^n(s), \nu) = \int_\Lambda h_1^k(\nu, \mu(s))u_{k,\nu} m_s^n(du) + h_2^k(\nu, \mu(s)) \cdot V_n(s).$$

Thus the state equation (3.30) can be rewritten as

$$\begin{aligned} V_n(t) = v_n + \int_0^t \vartheta^n(s)ds + M^n(t) + \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu \int_0^t \int_\Lambda h_1^k(\nu, \mu(s))u_{k,\nu} m_s^n(du)ds \\ + \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} e_\nu \int_0^t h_2^k(\nu, \mu(s)) \cdot V_n(s)ds. \end{aligned} \tag{3.34}$$

Recall from Section 1.2 that $\mathcal{M}([0,T]\times\Lambda)$ is the space of all finite measures on $[0,T]\times\Lambda$ equipped with the usual weak convergence topology.

Theorem 4. *Suppose Conditions 3.1.3, 3.1.4, and 3.1.5 hold and let v_n, v_0 be as in Theorem 2. Then:*

(i) $Y^n = \{V_n, M^n, m^n, \int_0^\cdot \vartheta^n(s)ds\}_{n\geq 1}$ *is a tight collection of* $\mathbb{D}([0,T]:\mathbb{R}^{2d}) \times \mathcal{M}([0,T]\times\Lambda) \times \mathbb{C}([0,T]:\mathbb{R}^d)$ *valued random variables.*

(ii) $\int_0^\cdot \vartheta^n(s)ds$ *converges to 0 in probability in* $\mathbb{C}([0,T]:\mathbb{R}^d)$.

(iii) $(V_n, M^n)_{n\geq 1}$ *is* $\mathbb{C}$*-tight.*

(iv) Suppose $\{Y^n\}$ *converges weakly along a subsequence to* $Y = (V, M, m, 0)$ *defined on a probability space* $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$. *Then,* $\mathbb{P}^*$ *a.s., the first marginal of* m *is the Lebesgue measure on* $[0,T]$. *Disintegrating* m *as*

$$m(A\times B) = \int_A m_t(B)dt, \quad A \in \mathcal{B}([0,T]),\ B \in \mathcal{B}(\Lambda),$$

define

$$U_{k,\nu}(t) \doteq \int_\Lambda u_{k,\nu} m_t(du), \quad t \in [0,T],\ k \in \mathbf{K},\ \nu \in \Delta^k. \tag{3.35}$$

Let $\{B_d(t)\}$ be a one-dimensional standard Brownian motion given on $(\Omega^*, \mathcal{F}^*, \mathbb{P}^*)$ that is independent of Y. Let $\mathcal{G}_t^\circ = \sigma\{B_d(s), V(s), M(s), m([0,s]\times B) : s \le t, B \in \mathcal{B}(\Lambda)\}$ and $\mathcal{G}_t$ be the $\mathbb{P}^*$-completion of $\mathcal{G}_t^\circ$. Then there is a d-dimensional $\{\mathcal{G}_t\}$-Brownian motion $\{W(t)\}, W = (W_1, \ldots, W_d)$ such that the following equation is satisfied

$$\begin{aligned} V(t) &= v_0 + \int_0^t \sigma(s)dW(s) + \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} e_\nu \int_0^t \int_\Lambda h_1^k(\nu,\mu(s))U_{k,\nu}(s)ds \\ &\quad + \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} e_\nu \int_0^t h_2^k(\nu,\mu(s))\cdot V(s)ds \\ &= v_0 + \int_0^t \eta(s,U(s))ds + \int_0^t \beta(s)V(s)ds + \int_0^t \sigma(s)dW(s). \end{aligned} \tag{3.36}$$

Proof. Tightness of $\{m^n\}$ as $\mathcal{M}([0,T]\times\Lambda)$-valued random variables is immediate since $m^n([0,T]\times\Lambda) = T$ for all n and Λ is a compact set. $\mathbb{C}$-tightness of $\{V_n\}$ was proved in Proposition 2.

In order to verify the tightness of $\{M^n\}_{n\ge1}$, we will use Theorem 2.3.2 of (Joffe and Métivier, 1986) (see Theorem 15 in Appendix). According to this theorem it suffices to verify conditions [A] and [T_1], given in Theorem 15, for the sequence of quadratic variation processes, $\{\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} n\langle M_{k,\nu}^n\rangle\}_{n\ge1}$. Note that

$$\sum_{k\in\mathbf{K}}\sum_{\nu\Delta^k} n\langle M_{k,\nu}^n\rangle(t) = \frac{1}{n}\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\int_0^t \Gamma_n^k(\mu_n(s),U^n(s),\nu)ds.$$

Condition [A] and [T_1] are now immediate on noting that Condition 3.1.3 implies (see (3.13))

$$\frac{1}{n}\Gamma_n^k(\mu_n(s),U^n(s),\nu) \le C_2$$

almost surely for all n, k, ν, and s. Furthermore $\{M^n\}$ is $\mathbb{C}$-tight because $j_T(M^n) \le \frac{1}{\sqrt{n}}\ell^{1/2}n_{\mathbf{K}}$.

Finally, from (3.31), for $\delta > 0$ we have that

$$\mathbb{P}\left[\sup_{0\le s\le T}\left\|\int_0^s \vartheta^n(u)du\right\| > \delta\right] \le \mathbb{P}\left[\sup_{0\le s\le T}\theta^n(V_n(s)) > \frac{\delta}{T}\right].$$

Since $\{V_n\}$ is $\mathbb{C}$-tight for every $\varepsilon > 0$, there exists some $\kappa_1 < \infty$ such that

$$\mathbb{P}\left[\sup_{0\le s\le T} \|V_n(s)\| > \kappa_1\right] \le \varepsilon$$

for all $n \in \mathbb{N}$. Recalling (3.32) we see that there exists an $n_0 > 0$ such that

$$\sup_{y:\|y\|\le\kappa_1} \theta^n(y) \le \frac{\delta}{T}$$

for all $n \ge n_0$. Thus for all $n \ge n_0$

$$\begin{aligned}&\mathbb{P}\left[\sup_{0\le s\le T}\left\|\int_0^s \vartheta^n(u)du\right\| > \delta\right]\\ &\quad\le \mathbb{P}\left[\sup_{0\le s\le T}\theta^n(V_n(s)) > \frac{\delta}{T}, \sup_{0\le s\le T}\|V_n(s)\| \le \kappa_1\right] + \mathbb{P}\left[\sup_{0\le s\le T}\|V_n(s)\| > \kappa_1\right] \le \varepsilon.\end{aligned}$$

Since $\varepsilon > 0$ is arbitrary we conclude that $\{\int_0^{\cdot} \vartheta^n(s)ds\}$ converges to 0 in probability in $\mathbb{C}([0,T]:\mathbb{R}^d)$. This concludes the proof of $(i), (ii)$ and (iii).

Consider now (iv). Let Y be as in the statement of the theorem, namely Y^n converges weakly along a subsequence to $Y = (V, M, m, 0)$. The property that the last component of Y must be 0 is a consequence of (ii). For notational convenience we label the subsequence once more by $\{n\}$. Recall the orthogonal matrix $Q = [q_1\ q_2\ \dots\ q_d]$ and function $a : [0,T] \to \mathbb{R}^{d\times d}$ defined in Section 3.1.3 as well as the function $\alpha^{1/2} : [0,T] \to \mathbb{R}^{(d-1)\times(d-1)}$ introduced above (3.18). Define $(d-1)$- and 1-dimensional processes $\hat{M}^n$ and R^n, respectively, as

$$\begin{pmatrix}\hat{M}^n(t)\\ R^n(t)\end{pmatrix} = Q'M^n(t). \tag{3.37}$$

Note that

$$R^n(t) = q_d'M^n(t) = \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\frac{1}{\sqrt{d}}\mathbf{1}'e_\nu M^n_{k,\nu}(t) = 0$$

since $\mathbf{1}'e_\nu = 0$ for all $k \in \mathbf{K}, \nu \in \Delta^k$.

We now show that M is a $\{\mathcal{G}_t\}$-martingale. The Burkholder-Davis-Gundy inequality (see Theorem IV.48 of (Protter, 2005)) implies that there exists $\kappa_2 \in (0,\infty)$ such that for $i = 1,\ldots,d$

$$\begin{aligned}\sup_{n\in\mathbb{N}}\mathbb{E}\sup_{0\le t\le T}(M_i^n(t))^4 &\le \sup_{n\in\mathbb{N}}\kappa_2 n^2\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\mathbb{E}[M_{k,\nu}^n]_T^2\\ &= \sup_{n\in\mathbb{N}}\kappa_2\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\mathbb{E}\left(\frac{1}{n}\mathcal{N}_{k,\nu}\left(\int_0^T I_n^k(\mu_n(s),U^n(s),\nu)ds\right)\right)^2 \qquad (3.38)\\ &\le \sup_{n\in\mathbb{N}}\kappa_2\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\mathbb{E}\left(\frac{1}{n}\mathcal{N}_{k,\nu}(nTC_2)\right)^2<\infty,\end{aligned}$$

where the first inequality on the last line is from (3.13). Thus $\{\sup_{0\le t\le T}\|M^n(t)\|^2\}_{n\ge1}$ is uniformly integrable. Let $k\in\mathbb{N}$ and $\mathcal{H}:(\mathbb{R}^d\times\mathbb{R}^d\times\mathbb{R})^k\to\mathbb{R}$ be a bounded and continuous function. For $0\le s\le t\le T$ and $0\le t_1\le\ldots\le t_k\le s$ we let $\xi_i^n=(V_n(t_i),M^n(t_i),m_i^n(f))$ and $\xi_i=(V(t_i),M(t_i),m_i(f))$ where $m_i^n(f)=\int_{\Lambda\times[0,t_i]}f(u)m_s^n(du)ds$, $m_i(f)=\int_{\Lambda\times[0,t_i]}f(u)m_s(du)ds$ and $f\in\mathbb{C}_b(\Lambda)$. Then

$$\mathbb{E}^*\mathcal{H}(\xi_1,\ldots,\xi_k)[M(t)-M(s)]=\lim_{n\to\infty}\mathbb{E}\mathcal{H}(\xi_1^n,\ldots,\xi_k^n)[M^n(t)-M^n(s)]=0$$

where the first equality follows from the uniform integrability property noted above, and the second equality is a consequence of the martingale property of M^n (which is a consequence of (3.38)). Combining this with the fact that B_d is a Brownian motion independent of Y, it follows that M is a $\{\mathcal{G}_t\}$-martingale.

We now define the process which will converge to the Brownian motion driving the limit diffusion. Recall that the matrix $\alpha^{1/2}$ is invertible and the property (3.23). Define the $(d-1)$-dimensional processes $B^n(t)=(B_i^n(t))_{i=1}^{d-1}$ as

$$B_i^n(t)=\sum_{j=1}^{d-1}\int_0^t\alpha_{ij}^{-1/2}(s)d\hat{M}_j^n(s),$$

where $\hat{M}^n$ is as in (3.37). Since M^n is a $\{\mathcal{F}_t^n\}$-martingale, both $\hat{M}^n$ and B^n are $\{\mathcal{F}_t^n\}$-martingales as well. From the estimate in (3.38) it follows that $\{\sup_{0\le t\le T}\|B^n(t)\|^2\}_{n\ge1}$ is uniformly integrable. Also note that for integers $1\le i,j\le d-1$, the cross quadratic varia-

tion of B_i^n and B_j^n can be expressed as

$$\langle B_i^n, B_j^n\rangle(t) = \sum_{m_1=1}^{d-1}\sum_{m_2=1}^{d-1}\int_0^t \alpha_{im_1}^{-1/2}(s)\alpha_{jm_2}^{-1/2}(s)d\langle \hat{M}_{m_1}^n, \hat{M}_{m_2}^n\rangle(s).$$

Note that for all $t \in [0,T]$

$$\langle \hat{M}_{m_1}^n, \hat{M}_{m_2}^n\rangle(t) = \langle q'_{m_1}M^n, q'_{m_2}M^n\rangle(t) = \sum_{m_3=1}^{d}\sum_{m_4=1}^{d} q_{m_3m_1}q_{m_4m_2}\langle M_{m_3}^n, M_{m_4}^n\rangle(t)$$

where

$$\langle M_{m_3}^n, M_{m_4}^n\rangle(t) = \sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k} e_\nu^{m_3}e_\nu^{m_4}\frac{1}{n}\int_0^t \Gamma_n^k(\mu_n(s), U^n(s), \nu)ds.$$

Thus

$$\begin{aligned}\langle B_i^n, B_j^n\rangle(t) &= \int_0^t \left(Q'\sigma(s)^{-1}\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\frac{1}{n}\left(\Gamma_n^k(\mu_n(s), U^n(s), \nu)e_\nu e_\nu'\right)(\sigma(s)')^{-1}Q\right)_{ij} ds \\ &= \int_0^t \left(Q'\sigma(s)^{-1}a(s)(\sigma(s)')^{-1}Q\right)_{ij} ds + \varepsilon_{ij}^n(t) = tI_{ij} + \varepsilon_{ij}^n(t)\end{aligned} \tag{3.39}$$

where I is the $d\times d$ identity matrix,

$$\varepsilon^n(t) = \int_0^t Q'\sigma(s)^{-1}\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\left(\frac{1}{n}\Gamma_n^k(\mu_n(s), U^n(s), \nu) - \Gamma^k(\mu(s), \nu)\right)e_\nu e_\nu'(\sigma(s)')^{-1}Qds$$

and ε_{ij}^n is the (i,j)-th coordinate of ε^n. From Condition 3.1.3 and (3.23) we have that $\mathbb{E}\|\varepsilon^n(t)\| \to 0$ for all t as $n\to\infty$.

Also it is easy to see that (cf. Theorem 2.2 of (Kurtz and Protter, 1991)) $B^n(\cdot) \Rightarrow \int_0^\cdot \alpha^{-1/2}(s)d\hat{M}(s) \doteq B(\cdot)$ in $\mathbb{D}([0,T] : \mathbb{R}^{d-1})$, where $\binom{\hat{M}}{0} = Q'M$. Also since $\{\sup_{0\le t\le T}\|B^n(t)\|^2\}_{n\ge1}$ is uniformly integrable, we have from (3.39) that

$$\begin{aligned}&\mathbb{E}^*\left(\mathcal{H}(\xi_1,\ldots,\xi_k)[B(t)B'(t) - B(s)B'(s) - (t-s)I]\right) \\ &\quad= \lim_{n\to\infty}\mathbb{E}\left(\mathcal{H}(\xi_1^n,\ldots,\xi_k^n)[B^n(t)(B^n)'(t) - B^n(s)(B^n)'(s) - (t-s)I]\right) \\ &\quad= \lim_{n\to\infty}\mathbb{E}\left(\mathcal{H}(\xi_1^n,\ldots,\xi_k^n)\varepsilon^n(t)\right) = 0.\end{aligned}$$

Combining this with the fact that B_d is independent of Y we see that B is a $(d-1)$-dimensional continuous $\mathcal{G}_t$-martingale with quadratic variation $\langle B\rangle(t) = tI$ which implies, by Lévy's theorem, that B is a $(d-1)$-dimensional $\{\mathcal{G}_t\}$-Brownian motion. Since B_d is a Brownian motion independent of Y, it follows that $\hat{W} \doteq (B, B_d)'$ is a d-dimensional $\{\mathcal{G}_t\}$-Brownian motion. Also note that

$$\hat{M}(t) = \int_0^t \alpha^{1/2}(s)dB(s). \tag{3.40}$$

The final step of the proof is to show that V is a solution to (3.36) with $W = Q\hat{W}$. Note that since Q is orthogonal, W is a d-dimensional $\{\mathcal{G}_t\}$-Brownian motion as well. From the definition of η and since $e_\nu \cdot \mathbf{1} = 0$ for all $k \in \mathbf{K}, \nu \in \Delta^k$, $Q'\eta$ takes the form

$$Q'\eta(t,u) = \begin{pmatrix} \hat{\eta}(t,u) \\ 0 \end{pmatrix}. \tag{3.41}$$

Similarly, from the expression for β and from (3.18) it follows that

$$Q'\beta(t)Q = \begin{pmatrix} \hat{\beta}(t) & 0 \\ 0 & 0 \end{pmatrix}, \quad Q'\sigma(t)Q = \begin{bmatrix} \alpha^{1/2}(t) & 0 \\ 0 & 0 \end{bmatrix}. \tag{3.42}$$

Also since $V_n \cdot \mathbf{1} = 0$ and $v_0 \cdot \mathbf{1} = 0$, we have

$$Q'V = \begin{bmatrix} \hat{V} \\ 0 \end{bmatrix}, \quad Q'v_0 = \left[\begin{smallmatrix} \hat{v}_0 \\ 0 \end{smallmatrix}\right]. \tag{3.43}$$

We first show that $\hat{V}$ solves the $(d-1)$-dimensional equation

$$\hat{V}(t) = \hat{v}_0 + \int_0^t \hat{\eta}(s,u)m_s(du)ds + \int_0^t \hat{\beta}(s)\hat{V}(s)ds + \int_0^t \alpha^{1/2}(s)dB(s). \tag{3.44}$$

Letting $\left[\begin{smallmatrix} \hat{V}_n \\ 0 \end{smallmatrix}\right] \doteq Q'V_n$ and using (3.34), we have,

$$\hat{V}_n(t) = \hat{v}_n + \int_0^t \hat{\eta}(s,u)m_s^n(du)ds + \int_0^t \hat{\beta}(s)\hat{V}_n(s)ds + \int_0^t \hat{\vartheta}^n(s)ds + \hat{M}^n(t)$$

where $\left[\begin{smallmatrix}\hat{v}_n\\0\end{smallmatrix}\right] = Q'v_n$ and $\left[\begin{smallmatrix}\hat{\vartheta}^n\\0\end{smallmatrix}\right] = Q'\vartheta^n$. Note that $(\hat{V}_n, \hat{M}^n, m^n, \hat{\vartheta}^n) \Rightarrow (\hat{V}, \hat{M}, m, 0)$. Without loss of generality we assume that the convergence holds a.s.

Since $m^n \to m$, we have

$$\int_0^t \int_\Lambda h_1^k(\nu, \mu(s)) u_{k,\nu} m_s^n(du)ds \to \int_0^t \int_\Lambda h_1^k(\nu, \mu(s)) u_{k,\nu} m_s(du)ds$$

and thus

$$\int_0^t \hat{\eta}(s,u) m_s^n(du)ds \to \int_0^t \hat{\eta}(s,u) m_s(du)ds. \tag{3.45}$$

Similarly it follows that

$$\int_0^t \hat{\beta}(s) \cdot \hat{V}_n(s)ds \to \int_0^t \hat{\beta}(s) \cdot \hat{V}(s)ds. \tag{3.46}$$

Combining (3.45) and (3.46) with (3.40) we see that $\hat{V}$ satisfies (3.44). Recalling the relation between $(\hat{v}_0, \hat{V}, \hat{\eta}, \hat{\beta}, \alpha^{1/2})$ and $(v_0, V, \eta, \beta, \sigma)$ we see that $V = Q\left[\begin{smallmatrix}\hat{V}\\0\end{smallmatrix}\right]$ is a solution of (3.36), where $W = Q\hat{W}$. This proves (iv) and thus completes the proof of the theorem. □

We now apply the above result to prove Theorem 3 which shows that the limit of the value of the optimal control problem for the n-th system as $n \to \infty$ can be bounded from below by the value of the control problem for the limit diffusion.

Proof of Theorem 3. Let v_n, v_0 be as in the statement of the theorem. It suffices to show that for any sequence of admissible controls $\{U^n\}$, $\liminf_{n\to\infty} J_n(U^n, v_n) \geq R(v_0)$. Let $U^n \in \mathcal{A}_n$, and m^n be the corresponding relaxed control defined as in (3.33). From the previous theorem we have that $\{(V_n, M^n, m^n, \int_0^\cdot \vartheta^n(s)ds)\}_{n\geq 1}$ is tight and thus every subsequence (also denoted with the index n) has a further subsequence $\{(V_{n_\ell}, M^{n_\ell}, m^{n_\ell}, \int_0^\cdot \vartheta^{n_\ell}(s)ds)\}$ such that

$$(V_{n_\ell}, M^{n_\ell}, m^{n_\ell}, \int_0^\cdot \vartheta^{n_\ell}(s)ds) \Rightarrow (V, M, m, 0).$$

Furthermore, equation (3.36) holds for the limit point $(V, M, m, 0)$ with a $\{\mathcal{G}_t\}$-Brownian motion W where $\{\mathcal{G}_t\}$ is as in the statement of Theorem 4 and $U_{k,\nu}$ are defined as in (3.35). It follows

from Fatou's Lemma that

$$\liminf_{\ell\to\infty} \mathbb{E}\int_0^T k_1(V_{n_\ell}(s))ds \geq \mathbb{E}^*\int_0^T \int_\Lambda k_1(V(s))ds.$$

Another application of Fatou's Lemma shows

$$\begin{aligned}\liminf_{\ell\to\infty} \mathbb{E}\int_0^T \int_\Lambda k_2(u)m_s^{n_\ell}(du)ds &\geq \mathbb{E}^*\int_0^T \int_\Lambda k_2(u)m_s(du)ds \\ &\geq \mathbb{E}^*\int_0^T k_2(U(s))ds\end{aligned}$$

where the second inequality follows on using Jensen's inequality, the relation (3.35), and the assumed convexity of k_2. Thus

$$\begin{aligned}\liminf_{\ell\to\infty} J_{n_\ell}(U^{n_\ell}, v_n) &= \liminf_{\ell\to\infty} \mathbb{E}\int_0^T (k_1(V_{n_\ell}(s)) + k_2(\sqrt{n_\ell}U^{n_\ell}(s)))ds \\ &\geq \mathbb{E}\int_0^T (k_1(V(s)) + k_2(U(s)))ds \\ &\geq R(v_0),\end{aligned}$$

where the last inequality follows on noting that $U = (U_{k,\nu})_{k\in\mathbf{K},\nu\in\Delta^k} \in \mathcal{A}(\Xi)$ where $\Xi = (\Omega^*, \mathcal{F}^*, \mathbb{P}^*, \{\mathcal{G}_t\})$. This completes the proof of the theorem. □

3.4 Feedback Controls

In this section we will introduce feedback controls, $U_g^n \in \mathcal{A}_n$ and $U_g \in \mathcal{A}(\Xi)$, associated with a measurable map $g : [0,T] \times \mathbb{R}^d \to \Lambda$ and prove that whenever g is continuous and $v_n \to v_0$, we have, under suitable conditions,

$$J_n(U_g^n, v_n) \to J(U_g, v_0). \tag{3.47}$$

In Section 3.4.1 we introduce feedback controls for the n-th system, whereas in Section 3.4.2 we define feedback controls for the limit diffusion. For the latter case we argue, using the non degeneracy of $\alpha(t)$ (under Condition 3.1.5), that there is a unique weak solution of the

corresponding stochastic differential equation. Finally, in Section 3.4.3 we prove the convergence in (3.47) when g is a continuous map.

3.4.1 Feedback Control in the n-th System

Given a measurable function $g : [0,T] \times \mathbb{R}^d \to \Lambda$, define for all $k \in \mathbf{K}, \nu \in \Delta^k$, functions $\Gamma_n^{k,g}(\cdot,\nu) : \mathcal{S}_n \times [0,T] \to \mathbb{R}_+$ as

$$\Gamma_n^{k,g}(r,s,\nu) \doteq \Gamma_n^k\left(r, \frac{1}{\sqrt{n}} g(s, \sqrt{n}(r - \mu(s)), \nu\right). \tag{3.48}$$

As with $u \in \Lambda$, g can be indexed by $k \in \mathbf{K}$ and $\nu \in \Delta^k$ with the corresponding entry denoted as $g_{k,\nu}$. Define μ_n^g through the right side of (3.7) by replacing $U^n(s)$ with

$$U_g^n(s) \doteq \frac{1}{\sqrt{n}} g(s, \sqrt{n}(\mu_n^g(s) - \mu(s))).$$

Then it can be checked that $U_g^n \in \mathcal{A}_n$ and μ_n^g is a time inhomogeneous Markov process with generator

$$L_g^n f(s,r) \doteq \sum_{k=1}^{K} \sum_{\nu \in \Delta^k} \Gamma_n^{k,g}(r,s,\nu) \left[f\left(s, r + \frac{1}{n} e_\nu\right) - f(s,r) \right] \tag{3.49}$$

for $s \in [0,T],\ r \in \mathcal{S}_n,\ f : [0,T] \times \mathcal{S}_n \to \mathbb{R}$.

3.4.2 Diffusion Feedback Control

In this section we introduce feedback controls for the limit diffusion model. Fix $v_0 \in \mathbb{V}_{d-1}$.

Definition 3.2. Let $g : [0,T] \times \mathbb{R}^d \to \Lambda$ be a measurable map. We say that the equation

$$\begin{cases} dV(t) &= \eta(t, g(t, V(t)))dt + \beta(t)V(t)dt + \sigma(t)dW(t) \\ V(0) &= v_0 \end{cases} \tag{3.50}$$

admits a weak solution if there exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ on which is given an $\{\mathcal{F}_t\}$-Wiener process W and an $\mathcal{F}_t$-adapted continuous process V such that for all

$0 \le t \le T$,

$$V(t) = v_0 + \int_0^t \eta(s, g(s, V(s)))ds + \int_0^t \beta(s)V(s)ds + \int_0^t \sigma(s)dW(s)$$

almost surely. We say that (3.50) admits a unique weak solution if whenever there are two sets of such spaces and processes denoted as $(\Omega^i, \mathcal{F}^i, \mathbb{P}^i, \{\mathcal{F}_t^i\}, (W^i, V^i))$, $i = 1, 2$ then the probability law of V^1 is the same as that of V^2.

Given a weak solution V associated with the system $\Xi = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \{W_t\})$ define $U_g \doteq g(\cdot, V(\cdot)) \in \mathcal{A}(\Xi)$. We refer to this control as the feedback control (for the limit diffusion) associated with the map g.

Theorem 5. *Under Condition 3.1.5 there is a unique weak solution of* (3.50).

Proof. Suppose V is a weak solution of (3.50) on some system $\Xi = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \{W_t\})$. Recall the definition of $\hat{V}$, $\hat{\eta}$, and $\hat{\beta}$ from Section 3.3 (cf. (3.41), (3.42), (3.43)). Let $Q'W \doteq \left(\begin{smallmatrix} B \\ W^* \end{smallmatrix}\right)$ and note that B and W^* are independent standard $(d-1)$- and 1-dimensional Brownian motions, respectively. Define $\hat{g} : [0, T] \times \mathbb{R}^{d-1} \to \Lambda$ as $\hat{g}(t, v) = g(t, Q\left(\begin{smallmatrix} v \\ 0 \end{smallmatrix}\right))$ and let $\left(\begin{smallmatrix} \hat{v}_0 \\ 0 \end{smallmatrix}\right) = Q'v_0$. Note that $\hat{V}$ is a solution of the $(d-1)$-dimensional SDE

$$\hat{V}(t) = \hat{v}_0 + \int_0^t \hat{\eta}(s, \hat{g}(s, \hat{V}(s)))ds + \int_0^t \hat{\beta}(s)\hat{V}(s)ds + \int_0^t \alpha^{1/2}(s)dB(s). \tag{3.51}$$

On the other hand if $\hat{V}$ is a solution of the SDE (3.51) on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$, where B is a $(d-1)$-dimensional $\{\mathcal{F}_t\}$ Brownian Motion, then as argued at the end of Theorem 4, by a suitable augmentation of the space with a one-dimensional Brownian motion B_d, $Q\left[\begin{smallmatrix} \hat{V} \\ 0 \end{smallmatrix}\right]$ is a solution of the SDE (3.50), with Brownian motion $W = Q\left[\begin{smallmatrix} B \\ B_d \end{smallmatrix}\right]$. Since from (3.23) $\sup_{v \in \mathbb{R}^d} \int_0^T \|\alpha(s)\|^{-1} \|\hat{\eta}(s, \hat{g}(s, v))\|^2 ds < \infty$, a standard argument using Girsanov's theorem shows that (3.51) has a unique weak solution. From the one-to-one correspondence between solutions of (3.51) and (3.50) noted above it now follows that there is a unique weak solution for (3.50). □

Recall the generator $\mathcal{L}_g$ in (3.24) associated with a measurable map $g : [0, T] \times \mathbb{R}^d \to \Lambda$.

Definition 3.3. Given $v_0 \in \mathbb{V}_{d-1}$, a d-dimensional stochastic process V on some filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ will be called a solution to the martingale problem associated with $(\mathcal{L}_g, v_0)$ if

$$\phi(V(t)) - \phi(v_0) - \int_0^t \mathcal{L}_g \phi(s, V(s)) ds$$

is a martingale for all $\phi \in \mathbb{C}_c^\infty(\mathbb{R}^d)$ and $V(0) = v_0$ almost surely.

The first part of the following result is standard (cf. (Stroock and Varadhan, 2007)) whereas the second part is immediate from Theorem 5.

Theorem 6. *A process V is a weak solution of the SDE (3.50) if and only if it is the solution to the martingale problem for $(\mathcal{L}_g, v_0)$. In particular, under Condition 3.1.5, there is a unique solution to the martingale problem for $(\mathcal{L}_g, v_0)$.*

3.4.3 Convergence Under Continuous Feedback Controls

Let $g : [0,T] \times \mathbb{R}^d \to \Lambda$ be a continuous function and V^g be the unique solution to (3.50) given on some system $\Xi = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \{W_t\})$. Define

$$V_n^g(t) = \sqrt{n}(\mu_n^g(t) - \mu(t)). \tag{3.52}$$

Recall that $U_g(t) = g(t, V^g(t)) \in \mathcal{A}(\Xi)$ and $U_g^n(t) = \frac{1}{\sqrt{n}} g(t, V_n^g(t)) \in \mathcal{A}_n$ are the controls associated with g for the limit diffusion and pre-limit system, respectively. In this section we will show that V_n^g converges in distribution to V^g, in $\mathbb{D}([0,T] : \mathbb{R}^d)$ and that $J_n(U_g^n, v_n)$ converges to $J(U_g, v_0)$. Namely we prove the following result.

Theorem 7. *Suppose Conditions 3.1.3, 3.1.4, and 3.1.5 hold, and let v_n, v_0 be as in Theorem 2, where $x_n = \mu_n^g(0)$. Then as $n \to \infty$:*

(i) V_n^g converges in distribution, in $\mathbb{D}([0,T] : \mathbb{R}^d)$, to V^g where V^g is the unique solution to the martingale problem for $(\mathcal{L}_g, v_0)$.

(ii) $J_n(U_n^g, v_n) \to J(U^g, v_0)$.

Proof. First consider (i). From Proposition 2 we have that $\{V_n^g\}$ is $\mathbb{C}$–tight in $\mathbb{D}([0,T]:\mathbb{R}^d)$. Since g is continuous, the operator $\mathcal{L}_g$ defined in (3.24) maps $\mathbb{C}_c^\infty(\mathbb{R}^d)$ to $\mathbb{C}_b([0,T]\times\mathbb{R}^d)$. In view of this, the tightness of $\{V_n^g\}$, the uniqueness established in Theorem 6, and Theorem 3.3.1 of (Joffe and Métivier, 1986), it suffices to show that for all $\phi\in\mathbb{C}_c^\infty(\mathbb{R}^d)$

$$\lim_{n\to\infty}\int_0^T \mathbb{E}^n|\mathcal{L}_g^n(\phi,s,V_n^g(s))-\mathcal{L}_g\phi(s,V_n^g(s))|ds=0 \tag{3.53}$$

where $\mathcal{L}_g$ is an in (3.24) and $\mathcal{L}_g^n$ is defined by the right side of (3.25), replacing u with $\frac{1}{\sqrt{n}}g(s,\sqrt{n}(s-\mu(s)))$, namely

$$\mathcal{L}_g^n(\phi,s,y)\doteq\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\Gamma_n^{k,g}\left(\gamma_n(s,y),s,\nu\right)\left[\phi\left(y+\frac{1}{\sqrt{n}}e_\nu\right)-\phi(y)\right]-\sqrt{n}F(\mu(s))\nabla\phi(y)$$

for $\phi\in\mathbb{C}_c^\infty(\mathbb{R}^d),\ s\in[0,T],\ y\in\mathbb{R}^d$ where $\Gamma_n^{k,g}$ is as in (3.48) (definition of $\Gamma_n^{k,g}$ is extended to all $r\in\mathbb{R}^d$ on setting $\Gamma_n^{k,g}(r,s,\nu)=0$ if $r\notin\mathcal{S}_n$). We note that Theorem 3.3.1 of (Joffe and Métivier, 1986) considers the setting of time-homogeneous diffusions, however the proof carries over to the setting of time-inhomogeneous generators considered here with minor modifications.

We now fix $\phi\in\mathbb{C}_c^\infty(\mathbb{R}^d)$ and for all $n\in\mathbb{N},\ k\in\mathbf{K},\ \nu\in\Delta^k$ define functions $\varphi_{k,\nu,1}^n:\mathbb{R}^d\to\mathbb{R}_+$, $\varphi_{k,\nu,2}^n:[0,T]\times\mathbb{R}^d\to\mathbb{R}_+$, and $A_j^n:[0,T]\times\mathbb{R}^d\to\mathbb{R}_+$ for $j=1,2,3$, as

$$\begin{aligned}\varphi_{k,\nu,1}^n(y)&\doteq\left|\phi\left(y+\frac{1}{\sqrt{n}}e_\nu\right)-\phi(y)-\frac{1}{\sqrt{n}}e_\nu'\nabla\phi(y)-\frac{1}{2n}e_\nu'D^2\phi(y)e_\nu\right|,\\ \varphi_{k,\nu,2}^n(s,y)&\doteq\left|\beta_k^n(y,\mu(s),g(s,y),\nu)\right|,\end{aligned}$$

and

$$\begin{aligned}A_1^n(s,y)&\doteq\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\Gamma_n^{k,g}\left(\gamma_n(s,y),s,\nu\right)\varphi_{k,\nu,1}^n(y),\\ A_2^n(s,y)&\doteq\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\varphi_{k,\nu,2}^n(s,y)|e_\nu'\nabla\phi(y)|,\\ A_3^n(s,y)&\doteq\frac{1}{2}\sum_{k\in\mathbf{K}}\sum_{\nu\in\Delta^k}\left|\frac{1}{n}\Gamma_n^{k,g}\left(\gamma_n(s,y),s,\nu\right)-\Gamma^k(\mu(s),\nu)\right|\left|e_\nu'D^2\phi(y)e_\nu\right|\end{aligned}$$

for $s \in [0,T]$ and $y \in \mathbb{R}^d$. Note that

$$\mathrm{Tr}(a(t)D^2\phi) = \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma^k(\mu(t),\nu) e_\nu' D^2\phi(y) e_\nu.$$

Adding and subtracting

$$\frac{1}{\sqrt{n}} \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma_n^{k,g}(\gamma_n(s,y),s,\nu) e_\nu' \nabla\phi(y) \text{ and } \frac{1}{2n} \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma_n^{k,g}(\gamma_n(s,y),s,\nu) e_\nu' D^2\phi(y) e_\nu$$

from $\mathcal{L}_g^n(\phi,s,y) - \mathcal{L}_g\phi(s,y)$, the triangle inequality yields

$$|\mathcal{L}_g^n(\phi,s,V_n^g(s)) - \mathcal{L}_g\phi(s,V_n^g(s))| \le A_1^n(s,V_n^g(s)) + A_2^n(s,V_n^g(s)) + A_3^n(s,V_n^g(s)).$$

We now consider the three terms on the right side separately. First consider $A_1^n(s,V_n^g(s))$. It follows from Taylor's theorem and the fact that all derivatives of ϕ are uniformly bounded that there exists $\kappa_1 \in (0,\infty)$ such that,

$$\varphi_{k,\nu,1}^n(V_n^g(s)) \le \frac{1}{6} \max_{\|\alpha\|=3} \sup_{x\in\mathbb{R}^d} \|D^\alpha\phi(x)\| \times \left\| \frac{e_\nu}{\sqrt{n}} \right\|^3 \le \frac{\kappa_1}{n^{3/2}},$$

where the outside maximum is taken over all mixed derivatives of order 3. Then, since $\frac{1}{\sqrt{n}} V_n^g(s) + \mu(s) \in \mathcal{S}_n$, (3.13) implies

$$A_1^n(s,V_n^g(s)) \le \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \Gamma_n^{k,g}\left(\gamma_n(s,V_n^g(s)),s,\nu\right) \frac{\kappa_1}{n^{3/2}} \le \frac{\kappa_2}{\sqrt{n}},$$

for all $s \in [0,T]$ and some $\kappa_2 \in (0,\infty)$. It follows that

$$\int_0^T \mathbb{E}^n |A_1^n(s,V_n^g(s))| ds \to 0 \text{ as } n \to \infty.$$

Now consider $A_2^n(s,V_n^g(s))$. From Condition 3.1.4 it follows that for $\kappa_3 > 0$, $\varepsilon > 0$,

$$\mathbb{P}^n \left[\sup_{0\le s\le T} \|V_n^g(s)\| \le \kappa_3,\ \varphi_{k,\nu,2}^n(s,V_n^g(s)) > \varepsilon \right] \to 0 \text{ as } n \to \infty.$$

Also the $\mathbb{C}$-tightness of $\{V_n^g\}$ implies that

$$\sup_n \mathbb{P}^n \left[\sup_{0 \le s \le T} \|V_n^g(s)\| > \kappa_3 \right] \to 0 \text{ as } \kappa_3 \to \infty.$$

Combining these two observations we see that

$$\varphi_{k,\nu,2}^n(s, V_n^g(s)) \to 0 \text{ in probability as } n \to \infty \text{ for all } s \in [0,T]. \tag{3.54}$$

Next, from Conditions 3.1.3, 3.1.4, and noting that h_1, h_2 are bounded functions, we see that there is a $\kappa_4 \in (0,\infty)$ such that for all $k \in \mathbf{K},\ \nu \in \Delta^k,\ n \ge 1$ and $s \ge 0$

$$\varphi_{k,\nu,2}^n(s, V_n^g(s)) \le \kappa_4(1 + \|V_n^g(s)\|) \text{ a.s.}$$

From Proposition 2,

$$\sup_n \mathbb{E}^n \sup_{t \le T} \|V_n^g(t)\|^2 < \infty. \tag{3.55}$$

Thus $\{\varphi_{k,\nu,2}^n(s, V_n^g(s))\}$ is uniformly integrable over $[0,T] \times \Omega$ and so combining this with (3.54), we have

$$\int_0^T \mathbb{E}^n |\varphi_{k,\nu,2}^n(s, V_n^g(s))| ds \to 0 \text{ as } n \to \infty.$$

Recalling the definition of A_2^n, it follows from the fact that all derivatives of ϕ are uniformly bounded that there exists $\kappa_5 \in (0,\infty)$ such that

$$\int_0^T \mathbb{E}^n |A_2^n(s, V_n^g(s))| ds \le \kappa_5 \sum_{k \in \mathbf{K}} \sum_{\nu \in \Delta^k} \int_0^T \mathbb{E}^n |\varphi_{k,\nu,2}^n(s, V_n^g(s))| ds \to 0 \text{ as } n \to \infty.$$

Finally, consider $A_3^n(s, V_n^g(s))$. It follows from Condition 3.1.3 and the boundedness of the derivatives of ϕ that there exists a $\kappa_6 \in (0,\infty)$ such that,

$$A_3^n(s, V_n^g(s)) \le \kappa_6 \sum_{k=1}^K \sum_{\nu \in \Delta^k} \left| \frac{1}{n} \Gamma_n^{k,g}(\gamma_n(s, V_n^g(s)), s, \nu) - \Gamma^k(\mu(s), \nu) \right| \le \frac{\kappa_6 C_1}{\sqrt{n}} (1 + \|V_n^g(s)\|).$$

Using the moment bound in (3.55) once more, we have that

$$\int_0^T \mathbb{E}|A_3^n(s, V_n^g(s))|ds \to 0.$$

This proves (3.53) and thus completes the proof of part (i).

Now consider (ii). By a similar argument as in Theorem 4

$$V_n^g(t) = v_n + \int_0^t b_n^{U_g^n(s)}(s, V_n^g(s))ds + M^n(t) \text{ for all } n \geq 1$$

where $M^n(t)$ is the local martingale in (3.29), with $M_{k,\nu}^n$ as in (3.8) with U^n replaced by U_g^n. Recall p and C_{k_1} introduced below (3.10). By a similar estimate as in (3.38) there exists $\kappa_7 \in (0, \infty)$ such that

$$\begin{aligned}
\sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\leq t\leq T} \|M_i^n(t)\|^{2p} &\leq \sup_{n\in\mathbb{N}} \kappa_7 n^p \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \mathbb{E}[M_{k,\nu}^n]_T^p \\
&= \sup_{n\in\mathbb{N}} \kappa_7 \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \mathbb{E}\left(\frac{1}{n}\mathcal{N}_{k,\nu}\left(\int_0^T I_n^k(\mu_n(s), U^n(s), \nu)ds\right)\right)^p \\
&\leq \sup_{n\in\mathbb{N}} \kappa_7 \sum_{k\in\mathbf{K}} \sum_{\nu\in\Delta^k} \mathbb{E}\left(\frac{1}{n}\mathcal{N}_{k,\nu}(nTC_2)\right)^p < \infty
\end{aligned} \tag{3.56}$$

where C_2 is as in (3.13). Also, from Lemma 2

$$\|b_n^{U_g^n(s)}(s, V_n^g(s))\|^{2p} \leq \kappa_7(1 + \|V_n^g(s)\|^{2p}). \tag{3.57}$$

Combining these two inequalities implies there exists a $\kappa_8 \in (0, \infty)$ such that

$$\mathbb{E} \sup_{0\leq s\leq t} \|V_n^g(s)\|^{2p} \leq \kappa_8\left(1 + \int_0^t E \sup_{0\leq u\leq s} \|V_n^g(u)\|^{2p}ds\right) \text{ for all } 0 \leq t \leq T.$$

Gronwall's inequality then yields,

$$\sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\leq t\leq T} \|V_n^g(t)\|^{2p} \leq \sup_{n\in\mathbb{N}} \kappa_8 e^{\kappa_8 T} < \infty$$

and thus $\{\sup_{t\leq T} \|V_n^g(t)\|^p\}$ is uniformly integrable. Recalling the definition of J_n in (3.10), it follows from this uniform integrability, part (i) of the theorem, the compactness of Λ, and

growth condition on k_1 (see below (3.10)) that

$$\mathbb{E}\int_0^T (k_1(V_n^g(t)) + k_2(\sqrt{n}U_g^n(t)))dt \to \mathbb{E}\int_0^T (k_1(V^g(t)) + k_2(U_g(t)))dt,$$

upon noting that $\sqrt{n}U_g^n(t) = g(t, V_n^g(t))$, $U_g(t) = g(t, V^g(t))$, and g is continuous. Thus we have shown $J_n(U_g^n, v_n) \to J(U_g, v_0)$ which completes the proof of (ii). □

3.5 Near Optimal Continuous Feedback Controls

In this section we give the final ingredient in the proof of Theorem 2, namely Theorem 9. This result says that for every $v_0 \in \mathbb{V}_{d-1}$ and $\varepsilon > 0$ there is a continuous $g_\varepsilon : [0,T] \times \mathbb{R}^d \to \Lambda$ such that U_{g_ε} is an ε-optimal control for the diffusion control problem, i.e. $J(U_{g_\varepsilon}, v_0) \le R(v_0) + \varepsilon$. Recall from Section 3.1.4 that this result combined with Theorems 3 and 7 proved earlier will complete the proof of Theorem 2. We begin with a result that says that for every $v_0 \in \mathbb{V}_{d-1}$, the infimum of the cost $J(\cdot, v_0)$ over all controls is the same as that over all feedback controls. The proof is similar to Theorem 4.2 in (Borkar, 1989) which considers a time homogeneous setting, and so we only provide a sketch.

Recall that for every measurable $g : [0,T] \times \mathbb{R}^d \to \Lambda$ there is a (feedback) control $U_g \in \mathcal{A}(\Xi)$ on some system Ξ. Denote the family of all such feedback controls as $\mathcal{A}_{fb}$. (This class depends on the initial condition v_0 in (3.50) but we suppress this in the notation). Throughout this section we will assume that Conditions 3.1.3 – 3.1.5 hold.

Theorem 8. *Fix $v_0 \in \mathbb{V}_{d-1}$. Then*

$$R(v_0) = \inf_{U \in \mathcal{A}_{fb}} J(U, v_0).$$

Proof. Suppose $U \in \mathcal{A}(\Xi)$ is an admissible control on a system $\Xi = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \{W_t\})$. As in Section 3.3 (cf. (3.33)) we denote the corresponding relaxed control by m. Let $V(\cdot)$ be the corresponding unique pathwise solution to (3.19). It suffices to show that there exists an admissible feedback control U^* such that $J(U^*, v_0) \le J(U, v_0)$. Define the probability measure

$\nu_{v_0} \in \mathcal{P}([0,T] \times \mathbb{V}_{d-1} \times \Lambda)$ as

$$\int_{[0,T]\times\mathbb{V}_{d-1}\times\Lambda} f(t,x,u) d\nu_{v_0}(t,x,u) = \frac{1}{T}\mathbb{E}\Big[\int_0^T \int_\Lambda f(t,V(t),u) m_t(du)\ dt\Big]$$

for all $f \in \mathbb{C}_b([0,T] \times \mathbb{V}_{d-1} \times \Lambda)$. Disintegrate ν_{v_0} as

$$\nu_{v_0}(dt\ dx\ du) = \beta_{v_0}(dt,dx)\pi(t,x)(du)$$

where $\beta_{v_0} \in \mathcal{P}([0,T] \times \mathbb{V}_{d-1})$ is the marginal distribution of ν_{v_0} on the first two coordinates and $\pi : [0,T] \times \mathbb{V}_{d-1} \to \mathcal{P}(\Lambda)$ is the corresponding regular conditional law. Define $g^* : [0,T] \times \mathbb{R}^d \to \Lambda$ as $g^*(t,x) = \int_\Lambda u\pi(t, \Pi_{\mathbb{V}_{d-1}}(x))(du)$ where $\Pi_{\mathbb{V}_{d-1}} : \mathbb{R}^d \to \mathbb{V}_{d-1}$ is the projection of $\mathbb{R}^d$ onto $\mathbb{V}_{d-1}$. Let U_{g^*} be the feedback control associated with the map g^* given on some system Ξ^* and let V^* be the corresponding state process given as the solution of (3.50) with g replaced by g^*. Let for $t \in [0,T]$, $\pi_t \doteq \pi(t, V^*(t))$. For $(t,z) \in [0,T] \times \mathbb{V}_{d-1}$, $r \in (0,\infty)$ and $\bar{k}_r(v,u) \doteq k_1(v) \wedge r + k_2(u)$ define

$$\begin{aligned}
\phi_r(t,z) &= \mathbb{E}^*\Big[\int_t^T \int_\Lambda \bar{k}_r(V^*(s),u)\pi_s(du)ds \Big| V^*(t) = z\Big], \\
Y_r(t) &= \int_0^t \int_\Lambda \bar{k}_r(V(s),u) m_s(du)ds + \phi_r(t,V(t)).
\end{aligned}$$

It follows using the equivalent description of a weak solution of (3.50) in terms of a $(d-1)$-dimensional SDE with uniformly non-degenerate diffusion coefficient as in the proof of Theorem 5 and classical PDE results (cf. Section III.4.2 of (Bensoussan, 2011)) that ϕ_r solves the equation

$$\int_\Lambda \bar{k}_r(x,u)\pi(t,x)(du) + \frac{\partial}{\partial t}\phi_r(t,x) + (\mathcal{L}_{g^*}\phi_r)(t,x) = 0 \tag{3.58}$$

where $\mathcal{L}_{g^*}$ is the generator for V^* given by the right side of (3.24) with g replaced by g^*. From the Itô-Krylov formula (cf. (Krylov, 2008)) we have

$$\begin{aligned}
\mathbb{E}[Y_r(t)] - \mathbb{E}[Y_r(0)] = \mathbb{E}\int_0^t \Big(\int_\Lambda \bar{k}_r(V(s),u) m_s(du) + \frac{\partial}{\partial t}\phi_r(s,V(s)) \\
+ (\hat{\mathcal{L}}_{U(s)}\phi_r)(s,V(s))\Big) ds.
\end{aligned} \tag{3.59}$$

where for $u \in \Lambda$, $\hat{\mathcal{L}}_u$ is the "controlled generator" defined as

$$\hat{\mathcal{L}}_u \phi_r(t,x) = \nabla_x \phi_r(t,x)(\eta(t,u) + \beta(t)x) + \frac{1}{2}\mathrm{Tr}(\sigma(t)D^2\phi_r(t,x)\sigma'(t)).$$

By the definition of π, and since $u \mapsto \hat{\mathcal{L}}_u \phi_r(t,x)$ is linear we see that

$$\int_\Lambda (\hat{\mathcal{L}}_u \phi_r)(s,x)\pi(s,x)(du) = (\mathcal{L}_{g^*}\phi)(s,x),\ (s,x) \in [0,T] \times \mathbb{V}_{d-1}.$$

From this it follows that

$$\begin{aligned}
&\mathbb{E}\int_0^t \left(\int_\Lambda \bar{k}_r(V(s),u) m_s(du) + (\hat{\mathcal{L}}_{U(s)}\phi_r)(s,V(s)) \right) ds \\
&\quad = \mathbb{E}\int_0^t \left(\int_\Lambda \bar{k}_r(V(s),u)\pi(s,V(s))(du) + (\mathcal{L}_{g^*}\phi_r)(s,V(s)) \right) ds.
\end{aligned}$$

Thus (3.58) implies that the right hand side of (3.59) is 0 and thus $\mathbb{E}[Y_r(t)] = \mathbb{E}[Y_r(0)] = \phi_r(0,v_0)$ for all $t \in [0,T]$. From the convexity of k_2 we see that

$$\begin{aligned}
\phi_r(0,v_0) &= \mathbb{E}^*\left[\int_0^T \int_\Lambda \bar{k}_r(V^*(s),u)\pi_s(du)ds \right] \\
&\geq \mathbb{E}^*\left[\int_0^T \bar{k}_r(V^*(s), g^*(s,V^*(s))ds \right] \\
&\doteq J_r(U_{g^*}, v_0).
\end{aligned}$$

Using the monotone convergence theorem it now follows that

$$\begin{aligned}
J(U,v_0) &= \lim_{r\to\infty} \mathbb{E}[Y_r(T)] = \lim_{r\to\infty} \mathbb{E}Y_r(0) \\
&= \lim_{r\to\infty} \phi_r(0,v_0) \geq \lim_{r\to\infty} J_r(U_{g^*},v_0) = J(U_{g^*},v_0).
\end{aligned}$$

The result follows. □

We will next show in Theorem 9 below that the above theorem can be strengthened in that the class $\mathcal{A}_{fb}$ can be replaced by the smaller class $\mathcal{A}^c_{fb}$ of all *continuous* feedback controls, i.e. feedback controls for which that corresponding map g is continuous. Recall the orthogonal matrix Q defined in Section 3.1.3. Fix $v_0 \in \mathbb{V}_{d-1}$ and let $g^* : [0,T]\times\mathbb{R}^d \to \Lambda$ be a measurable map.

Let U_{g^*} be the corresponding feedback control given on some system $\Xi = (\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\}, \{W_t\})$ and let V^* be the solution of (3.50) with g replaced by g^* on the right side. Define the $(d-1)$-dimensional process $\hat{V}^*$ such that $V^* = Q\left(\begin{smallmatrix}\hat{V}^* \\ 0\end{smallmatrix}\right)$ and the map $\hat{g}^* : [0,T] \times \mathbb{R}^{d-1} \to \Lambda$ as $\hat{g}^*(t,v) = g^*(t, Q\left(\begin{smallmatrix}v \\ 0\end{smallmatrix}\right))$ for $v \in \mathbb{R}^{d-1}$. Then,

$$\hat{V}^*(t) = \hat{v}_0 + \int_0^t \hat{\eta}(s, \hat{g}^*(s, \hat{V}^*(s)))ds + \int_0^t \hat{\beta}(s)V^*(s)ds + \int_0^t \alpha^{1/2}(s)d\hat{W}(s) \tag{3.60}$$

where $\hat{\eta}$, $\hat{\beta}$, and α are as in (3.41), (3.42), and (3.17), respectively. In addition, $v_0 = Q\left(\begin{smallmatrix}\hat{v}_0 \\ 0\end{smallmatrix}\right)$ and $Q'W = \left(\begin{smallmatrix}\hat{W} \\ B_d\end{smallmatrix}\right)$. Define $\varrho \in \mathcal{P}([0,T] \times \mathbb{R}^{d-1})$ as

$$\varrho(A) \doteq \bar{c} \int_A e^{-\frac{(\|x\|^2 + t^2)}{2}} dxdt \tag{3.61}$$

for $A \in \mathcal{B}([0,T] \times \mathbb{R}^{d-1})$ where $\bar{c}$ is a normalizing constant. We denote by $\bar{\mathcal{B}}$ the Lebesgue σ-field on $[0,T] \times \mathbb{R}^{d-1}$, namely the completion of $\mathcal{B}([0,T] \times \mathbb{R}^{d-1})$ with respect to the Lebesgue measure.

Lemma 3. *For each $n \in \mathbb{N}$ there exists a $\bar{\mathcal{B}}$-measurable function $\hat{g}_n : [0,T] \times \mathbb{R}^{d-1} \to \Lambda$ and compact sets $A_n \in \mathcal{B}([0,T] \times \mathbb{R}^{d-1})$ such that $\hat{g}_n$ is continuous and,*

$$\{(s,v) \in [0,T] \times \mathbb{R}^{d-1} : \hat{g}^*(s,v) \neq \hat{g}_n(s,v)\} \subset A_n^c \quad \text{and} \quad \varrho(A_n^c) \leq \frac{1}{2^{n+1}}. \tag{3.62}$$

Proof. From Lusin's theorem (cf. 2.24 of (Rudin, 1986)) for each $n \in \mathbb{N}$ there exists a continuous function $\hat{g}'_n : [0,T] \times \mathbb{R}^{d-1} \to \mathbb{R}^\ell$ such that (3.62) is satisfied. Since Λ is a closed convex set, there is a continuous map $\Pi_\Lambda : \mathbb{R}^\ell \to \Lambda$ such that $\Pi_\Lambda(u) = u$ for all $u \in \Lambda$. Define $\hat{g}_n : [0,T] \times \mathbb{R}^{d-1} \to \Lambda$ as $\hat{g}_n(s,v) = \Pi_\Lambda(\hat{g}'_n(s,v))$. The result now follows on noting that

$$\{(s,v) : \hat{g}_n(s,v) = \hat{g}^*(s,v)\} \supset \{(s,v) : \hat{g}'_n(s,v) = \hat{g}^*(s,v)\}.$$

□

Let $\{v_n\} \subset \mathbb{V}_{d-1}$ be such that $v_n \to v_0$ and let $\Xi^n = (\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}, \mathbb{P}^n, \{W^n\})$ be a system on which the process V^n is the unique (weak) solution to

$$V^n(t) = v_n + \int_0^t \eta(s, g_n(s, V^n(s)))ds + \int_0^t \beta(s)V^n(s)ds + \int_0^t \sigma(s)dW^n(s) \tag{3.63}$$

where $g_n : [0,T] \times \mathbb{V}_{d-1} \to \Lambda$ is the continuous function defined as $g_n(s, Q\binom{v}{0}) = \hat{g}_n(s,v),\ v \in \mathbb{R}^{d-1}$. We can extend g_n continuously to $[0,T] \times \mathbb{R}^d$ as before using the projection map $\Pi_{\mathbb{V}_{d-1}}$. Defining $\hat{V}^n$ as $Q'V^n = \binom{\hat{V}^n}{0}$, we can write

$$\hat{V}^n(t) = \hat{v}_n + \int_0^t \hat{\eta}(s, \hat{g}_n(s, \hat{V}^n(s)))ds + \int_0^t \hat{\beta}(s)\hat{V}^n(s)ds + \int_0^t \alpha^{1/2}(s)d\hat{W}^n(s)$$

where $Q'v_n = \binom{\hat{v}_n}{0}$ and $\hat{W}^n$ is a $(d-1)$-dimensional Brownian motion.

Theorem 9. *Given $v_0 \in \mathbb{V}_{d-1}$, let V^* be as introduced in (3.60). Let v_n, g_n and $\{V^n\}$ be as introduced above. Then $V^n \Rightarrow V^*$ as a sequence of $\mathbb{C}([0,T] : \mathbb{R}^d)$-valued random variables.*

Proof. It suffices to show that $\hat{V}^n \Rightarrow \hat{V}^*$. Let $G = \mathbb{R}^{d-1} \times \Lambda$ and define $m^n \in \mathcal{M}([0,T] \times G)$ as

$$m^n(A \times B \times C) \doteq \int_0^T 1_A(s)1_B(\hat{V}^n(s))1_C(\hat{g}_n(s, \hat{V}^n(s)))ds,$$

where $A \in \mathcal{B}([0,T]),\ B \in \mathcal{B}(\mathbb{R}^{d-1}),\ C \in \mathcal{B}(\Lambda)$. Since $u \mapsto \hat{\eta}(s,u)$ is a linear function and $\int_0^t \hat{g}_n(s, \hat{V}^n(s))ds = \int_0^t um^n(ds\ dv\ du)$, $\hat{V}^n(t)$ can be expressed as

$$\hat{V}^n(t) = \hat{v}_n + \int_0^t \hat{\eta}(s,u)m^n(ds\ dv\ du) + \int_0^t \hat{\beta}(s)\hat{V}^n(s)ds + \int_0^t \alpha^{1/2}(s)d\hat{W}^n(s).$$

We can disintegrate m^n as $m_t^n(dv\ du)dt$, where $m_t^n(dv\ du) = \delta_{\hat{V}^n(t)}(dv)\delta_{\hat{g}_n(t,\hat{V}^n(t))}(du)$ and δ_x is the Dirac measure at the point x. From the boundedness of $\hat{\eta},\ \hat{\beta}$, and $\alpha^{1/2}$, we get by a standard application of Gronwall's inequality that for some $C \in (0,\infty)$

$$\mathbb{E}[\hat{V}^n(t)] \le C(1 + \hat{v}_n)e^{Ct}, \text{ for all } n \in \mathbb{N},\ t \in [0,T]. \tag{3.64}$$

Using this moment bound and a similar bound on the increments of $\hat{V}^n$ we have that $\{\hat{V}^n\}$ is a tight sequence of $\mathbb{C}([0,T] : \mathbb{R}^{d-1})$-valued random variables. Now the tightness of $\{m^n\}$

as a sequence of $\mathcal{M}([0,T]\times G)$-valued random variables is immediate since the first marginal is the Lebesgue measure (i.e. $m^n([0,t)\times G) = t$ for all $t \in [0,T]$), $\{\hat{V}^n\}$ is tight, and Λ is compact. Also, the tightness of $\{\hat{W}^n\}$ as a sequence of $\mathbb{C}([0,T]:\mathbb{R}^{d-1})$-valued random variables is immediate since $\hat{W}^n$ is a standard Brownian motion for each n. Therefore $\{\hat{V}^n, \hat{W}^n, m^n\}$ is a tight collection of $\mathbb{C}([0,T]:\mathbb{R}^{2(d-1)})\times\mathcal{M}([0,T]\times G)$-valued random variables.

Suppose $\{\hat{V}^n, \hat{W}^n, m^n\}$ converges along a subsequence (also denoted $\{n\}$) to a process, $\{\hat{V}, \hat{W}, m\}$. Let $(\Omega', \mathcal{F}', \mathbb{P}')$ be the probability space on which the limit processes are defined. Then $\hat{W}$ is a $\mathbb{P}'$-Brownian motion and using the continuity of $\hat{\eta}$, $\hat{\beta}$ and $\alpha^{1/2}$ we see that $(\hat{V}, \hat{W}, m)$ satisfy

$$\hat{V}(t) = \hat{v}_0 + \int_0^t \hat{\eta}(s,u)dm(ds\ dv\ du) + \int_0^t \hat{\beta}(s)\hat{V}(s)ds + \int_0^t \alpha^{1/2}(s)d\hat{W}(s)$$

$\mathbb{P}'$-almost surely.

Define $\mathcal{F}'_t = \sigma\{\hat{V}_s, \hat{W}_s, m([0,s]\times A) : 0 \le s \le t, A \in \mathcal{B}(G)\}$. It is easy to check that $\{\hat{W}_t\}$ is a $\{\mathcal{F}'_t\}$-martingale. Indeed, let $k \in \mathbb{N}$ and $\mathcal{H} : (\mathbb{R}^{2(d-1)}\times\mathbb{R})^k \to \mathbb{R}$ be a bounded and continuous function. Define $\mathcal{Z}_t \doteq (\hat{V}_t, \hat{W}_t, m(t,f))$ and $\mathcal{Z}^n_t \doteq (\hat{V}^n_t, \hat{W}^n_t, m^n(t,f))$, where $f \in \mathbb{C}_b(G)$ and $\nu(t,f) = \int_0^t f(v,u)\nu(ds\ dv\ du)$ for $\nu = m, m^n$. Then for $s \le t \le T$ and $0 \le t_1 \le \ldots t_k \le s$,

$$\mathbb{E}'\mathcal{H}(\mathcal{Z}_{t_1}, \ldots, \mathcal{Z}_{t_k})[\hat{W}_t - \hat{W}_s] = \lim_{n\to\infty} \mathbb{E}^n\mathcal{H}(\mathcal{Z}^n_{t_1}, \ldots, \mathcal{Z}^n_{t_k})[\hat{W}^n_t - \hat{W}^n_s] = 0,$$

where the second equality uses the fact that $\hat{W}^n$ is a $\{\mathcal{F}^n_t\}$-Brownian motion and $\mathcal{Z}^n_t$ is $\{\mathcal{F}^n_t\}$-adapted. This proves that $(\hat{W}_t)$ is an $\{\mathcal{F}'_t\}$-martingale.

Note that m, m^n can be disintegrated as

$$m(ds\ dv\ du) = m_s(dv\ du)ds,\ m^n(ds\ dv\ du) = m^n_s(dv\ du)ds.$$

We will now argue that for all $t \in [0,T]$,

$$\int_0^t \int_G u m_s(dv\ du)ds = \int_0^t \hat{g}^*(s, \hat{V}(s))ds \quad \text{a.s. } \mathbb{P}'. \tag{3.65}$$

Note that (3.65), the linearity of $\hat{\eta}$ in u, together with the weak-uniqueness of solutions to (3.60) (which was established in Section 3.4) completes the proof of the theorem.

Note that for any $f \in \mathbb{C}_b(\mathbb{R}^{d-1})$ we have $\int_0^t \int_G f(v) m_s^n(dv\ du)ds = \int_0^t f(\hat{V}^n(s))ds$. Since $(m^n, \hat{V}^n) \Rightarrow (m, \hat{V})$, we have for any such f

$$\int_0^t \int_G f(v) m_s(dv\ du)ds = \int_0^t f(\hat{V}(s))ds \quad \text{for all } t \in [0,T],\ \text{a.s. } \mathbb{P}'.$$

Denote by $\hat{m}_t^i,\ i = 1,2$ the marginal of m_t on its i-th coordinate. Then the above display can be rewritten as

$$\int_0^t \int_{\mathbb{R}^{d-1}} f(v)\hat{m}_s^1(dv)ds = \int_0^t f(\hat{V}(s))ds,\ \text{for } t \in [0,T],\ \text{a.s. } \mathbb{P}',\ \text{for every } f \in \mathbb{C}_b(\mathbb{R}^{d-1} : \mathbb{R}).$$

This shows that

$$\hat{m}_t^1(dv) = \delta_{\hat{V}(t)}(dv),\ [\lambda \otimes \mathbb{P}']\ \text{a.e. } (t, w') \tag{3.66}$$

where λ is the Lebesgue measure on $[0,T]$.

Recall the definition of A_n from Lemma 3 and ϱ from (3.61). Define $B_n \doteq \cap_{m=n}^{\infty} A_n$. Then

$$\varrho(B_n) \geq 1 - \frac{1}{2^n} \quad \text{for all } n \geq 1$$

and $\hat{g}^*(s,v) = \hat{g}_n(s,v) = \hat{g}_{n+1}(s,v) = \ldots$ for all $(s,v) \in B_n$. Since $\{\hat{v}_n\}$ is bounded we have from the moment bound in (3.64) that for every $\varepsilon > 0$, there is a compact $F \subset \mathbb{R}^{d-1}$ such that

$$\sup_{n \in \mathbb{N}} \sup_{0 \leq t \leq T} \mathbb{P}^n[\hat{V}^n(t) \in F^c] \leq \frac{\varepsilon}{2}. \tag{3.67}$$

Note that this says in particular that $\{\hat{v}_n\} \subset F$. For $t \in [0,T]$ and $v \in \mathbb{R}^{d-1}$, let $p(t,v,z)$ be the transition probability density of the Gaussian random variable $\hat{V}_0^v(t)$ given as the solution of the SDE

$$\hat{V}_0^v(t) = v + \int_0^t \hat{\beta}(s)\hat{V}_0^v ds + \int_0^t \alpha^{1/2}(s) d\hat{W}(s).$$

It is easy to see that there exists a function $\Psi : [0,T] \to \mathbb{R}_+$ and $\kappa \in (0,\infty)$ such that

$$\sup_{v,z\in F} p(t,v,z) \le \Psi(t),\ t \in [0,T],\ \text{and}\ \int_0^T e^{-\kappa/t}\Psi(t)dt < \infty. \tag{3.68}$$

Using the boundedness of $\hat{\eta}$ and $\alpha^{-1/2}$, Girsanov's theorem, and the Cauchy-Schwarz inequality we see that there exists a $\theta \in (0,\infty)$ such that for any bounded measurable $f : [0,T] \times \mathbb{R}^{d-1} \to \mathbb{R}$ and $t \in [0,T]$

$$\mathbb{E}^n \left| \int_0^t f(s, \hat{V}^n(s))ds \right| \le \theta \left[\mathbb{E}' \left(\int_0^t f(s, \hat{V}_0^{v_n}(s))^2 ds \right) \right]^{1/2}. \tag{3.69}$$

Since $e^{-\kappa/s}\psi(s)1_F(v)dvds$ is a finite measure on $[0,T]\times\mathbb{R}^{d-1}$ that is absolutely continuous with respect to ϱ, we have for any $\varepsilon > 0$ a $n_0 \in \mathbb{N}$ such that

$$\int_0^T \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/s} 1_F(v)\Psi(s)dvds < \frac{\varepsilon^2}{4\theta^2}. \tag{3.70}$$

Together with (3.68), (3.70) implies

$$\mathbb{E}' \int_0^T e^{-\kappa/s} 1_{B_{n_0}^c}(s, \hat{V}_0^v(s)) 1_F(\hat{V}_0^v(s))ds < \frac{\varepsilon^2}{4\theta^2} \tag{3.71}$$

for all $v \in F$. From (3.67), (3.69), and (3.71) we have

$$\begin{aligned} \mathbb{E}^n \int_0^T e^{-\kappa/2s} 1_{B_{n_0}^c}(s, \hat{V}^n(s))ds &< \mathbb{E}^n \int_0^T 1_F(\hat{V}^n(s)) e^{-\kappa/2s} 1_{B_{n_0}^c}(s, \hat{V}^n(s))ds + \frac{\varepsilon}{2} \\ &\le \theta \left[\mathbb{E}' \left(\int_0^T 1_F(\hat{V}_0^{v_n}(s)) e^{-\kappa/s} 1_{B_{n_0}^c}(s, \hat{V}_0^{v_n}(s))ds \right) \right]^{1/2} + \frac{\varepsilon}{2} \\ &\le \varepsilon. \end{aligned} \tag{3.72}$$

Denote by $\hat{m}_t^{n,i}$ the marginal of m_t^n on the i-th coordinate for $i = 1, 2$. Then, for any $n \geq n_0,\ t \in [0,T],\ f \in \mathbb{C}(\Lambda)$, and $h \in \mathbb{C}([0,T])$

$$\begin{aligned}\int_0^t \int_G e^{-\kappa/2s} h(s) f(u) m_s^n(dv\ du) ds &= \int_0^t \int_{\mathbb{R}^{d-1}} e^{-\kappa/2s} h(s) f(\hat{g}_n(s,v)) \hat{m}_s^{n,1}(dv) ds \\ &= \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}}(s,v) e^{-\kappa/2s} h(s) f(\hat{g}_{n_0}(s,v)) \hat{m}_s^{n,1}(dv) ds \\ &\quad + \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/2s} h(s) f(\hat{g}_n(s,v)) \hat{m}_s^{n,1}(dv) ds,\end{aligned}$$

where the second equality follows on noting that for $(s,v) \in B_{n_0}$, $\hat{g}_n(s,v) = \hat{g}_{n_0}(s,v)$ when $n \geq n_0$. Thus

$$\begin{aligned}&\left| \int_0^t \int_G e^{-\kappa/2s} h(s) f(u) m_s^n(dv\ du) ds - \int_0^t \int_{\mathbb{R}^{d-1}} e^{-\kappa/2s} h(s) f(\hat{g}_{n_0}(s,v)) \hat{m}_s^{n,1}(dv) ds \right| \\ &\quad \leq 2\|f\|_\infty \|h\|_\infty \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/2s} \hat{m}_s^{n,1}(dv) ds.\end{aligned} \tag{3.73}$$

It follows from (3.72) that the expectation of (3.73) is bounded above by $2\|f\|_\infty \|h\|_\infty \varepsilon$ and thus, letting $n \to \infty$

$$\begin{aligned}&\mathbb{E}' \left| \int_0^t \int_G e^{-\kappa/2s} h(s) f(u) m_s(dv\ du) ds - \int_0^t \int_{\mathbb{R}^{d-1}} e^{-\kappa/2s} h(s) f(\hat{g}_{n_0}(s,v)) \hat{m}_s^1(dv) ds \right| \\ &\quad \leq 2\|f\|_\infty \|h\|_\infty \varepsilon.\end{aligned}$$

Therefore, since $\hat{g}_{n_0}(s,v) = \hat{g}^*(s,v)$ on B_{n_0}

$$\begin{aligned}&\mathbb{E}' \left| \int_0^t \int_G e^{-\kappa/2s} h(s) f(u) m_s(dv\ du) ds - \int_0^t \int_{\mathbb{R}^{d-1}} e^{-\kappa/2s} h(s) f(\hat{g}^*(s,v)) \hat{m}_s^1(dv) ds \right| \\ &\quad \leq 2\|f\|_\infty \|h\|_\infty \left[\varepsilon + \mathbb{E}' \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/2s} \hat{m}_s^1(dv) ds \right].\end{aligned}$$

Since $B_{n_0}^c$ is open, it then follows from (3.72)

$$\mathbb{E}' \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/2s} \hat{m}_s^1(dv) ds \leq \liminf_{n \to \infty} \mathbb{E}^n \int_0^t \int_{\mathbb{R}^{d-1}} 1_{B_{n_0}^c}(s,v) e^{-\kappa/2s} \hat{m}_s^{n,1}(dv) ds \leq \varepsilon.$$

Letting $\varepsilon \to 0$ we have for all $t \in [0,T],\ h \in \mathbb{C}([0,T]),\ f \in \mathbb{C}(\Lambda)$ that

$$\int_0^t \int_G e^{-\kappa/2s} h(s) f(u) m_s(dv\ du) ds = \int_0^t \int_{\mathbb{R}^{d-1}} e^{-\kappa/2s} f(\hat{g}^*(s,v)) \hat{m}_s^1(dv) ds \quad \text{a.e. } \mathbb{P}'.$$

Combined with (3.66) this implies that

$$m_s(dv\ du) = \delta_{\hat{V}(s)}(dv) \delta_{\hat{g}^*(s,\hat{V}(s))}(du), [\lambda \times \mathbb{P}'] \text{ a.e. } (s,w').$$

This proves (3.65) and, as argued previously, completes the proof of the theorem. □

3.6 Example

The following class of models is studied in (Antunes et al., 2008). Consider a system consisting of n identical servers (nodes) of capacity $C \in \mathbb{N}$ and K different classes of jobs each with its own capacity requirement $A_k \in \mathbb{N},\ k \in \{1,\dots,K\}$. External jobs of type k arrive at each server with rate λ_k. A job of type k remains at a given node for an exponential holding time with mean γ_k^{-1} before attempting to move to another randomly chosen node. If the server has available capacity it accepts the job, otherwise the job is rejected and exits the system. If not rejected first, a type k job remains in the system for an exponential amount of time with mean τ_k^{-1} before leaving the system. We make the usual assumptions of mutual independence, in particular a.s. at most one job may arrive, switch nodes, or exit the system at a given time, but note that such an event may correspond to the change in state of multiple servers.

For the discussion below, for simplicity, we consider the case where there are only two classes of jobs. In the notation of the current paper, the state process $X_n(t) = \{X_n^1(t),\dots,X_n^n(t)\}$ is the pure jump Markov process where $X_n^i(t)$ takes values in

$$\mathbb{X} = \{(j,i) \in \mathbb{N}_0 \times \mathbb{N}_0 : jA_1 + iA_2 \le C\}.$$

Let, as before, $d = |\mathbb{X}|$, $\mathcal{S} = \mathcal{P}(\mathbb{X})$, and $\mathcal{S}_n = \mathcal{P}(\mathbb{X}) \cap \frac{1}{n}\mathbb{N}^d$. The empirical measure process, $\mu_n(t) \in \mathcal{S}_n$, is a d-dimensional pure jump Markov process where $\mu_n^{j,i}(t) = \frac{1}{n}\sum_{k=1}^n 1_{\{X_n^k(t)\}}((j,i))$ represents the proportion of nodes with exactly j and i jobs of type 1 and 2, respectively. We suppose that $\mu_n(0) = x_n$ a.s. for some deterministic $x_n \in \mathcal{S}_n$ such that $x_n \to x_0$ as $n \to \infty$ and

$x_0^{j,i} > 0$ for all $(j,i) \in \mathbb{X}$. Also suppose that $v_n \doteq \sqrt{n}(x_n - x_0) \to v_0$ as $n \to \infty$. The rate function $\bar{\Gamma}_n^k$ associated with this system is described in (Antunes et al., 2008) but we present it below in our notation for completeness. Jobs can enter or leave the system or switch nodes which means that there are three transition types for each class of job. Thus the set $\mathbf{K}$ of different jump types can be represented as $\mathbf{K} = \{E^i, L^i, C^i : i = 1, 2\}$ where $n_{E^i} = n_{L^i} = 1$ and $n_{C^i} = 2$ for $i = 1, 2$. Let for $(j,i) \in \mathbb{X}$, $\hat{e}_{j,i} = (\delta_{(j,i),(k,\ell)})_{(k,\ell)\in\mathbb{X}}$ be the d-dimensional vector which is 1 for entry (j,i) and 0 for all other entries. The sets corresponding to the possible jumps of each type are

$$\begin{aligned}
\Delta^{E^1} &= \{(\hat{e}_{j,i}, \hat{e}_{j+1,i}) : (j,i) \in S^{E^1}\}, \quad \Delta^{E^2} = \{(\hat{e}_{j,i}, \hat{e}_{j,i+1}) : (j,i) \in S^{E^2}\} \\
\Delta^{L^1} &= \{(\hat{e}_{j,i}, \hat{e}_{j-1,i}) : (j,i) \in S^{L^1}\}, \quad \Delta^{L^2} = \{(\hat{e}_{j,i}, \hat{e}_{j,i-1}) : (j,i) \in S^{L^2}\} \\
\Delta^{C^1} &= \Delta^{L^1} \cup \{(\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j-1,i} + \hat{e}_{j'+1,i'}) : (j,i,j',i') \in S^{C^1}\} \\
\Delta^{C^2} &= \Delta^{L^2} \cup \{(\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j,i-1} + \hat{e}_{j',i'+1}) : (j,i,j',i') \in S^{C^2}\}.
\end{aligned}$$

where $S^{E^1} = \{(j,i) \in \mathbb{X} : (j+1,i) \in \mathbb{X}\}$ and S^{E^2}, S^{L^1}, S^{L^2}, S^{C^1}, S^{C^2} are defined similarly.

Let $r \in \mathcal{S}_n$. The rate of jumps corresponding to a job arriving at a node with j and i jobs of classes 1 and 2, respectively, is equal to the number of nodes in this configuration multiplied by the rate at which jobs enter the system. Namely, the rate $\bar{\Gamma}_n^k(r,\nu)$ when $\nu = (\hat{e}_{j,i}, \hat{e}_{j+1,i}) \in \Delta^k$ and $k = E^1$ is $nr^{j,i} \times \lambda_1$, and similarly $\bar{\Gamma}_n^k(r,\nu) = nr^{j,i} \times \lambda_2$, $\nu = (\hat{e}_{j,i}, \hat{e}_{j,i+1}) \in \Delta^k$, $k = E^2$. The rate of departures is given similarly but, since all jobs are processed simultaneously, we need to multiply the processing rate by the number of jobs at a given node. Specifically, $\bar{\Gamma}_n^k(r,\nu) = j \times nr^{j,i} \times \tau_1$ for $\nu = (\hat{e}_{j,i}, \hat{e}_{j-1,i}) \in \Delta^k$, $k = L^1$ and $\bar{\Gamma}_n^k(r,\nu) = i \times nr^{j,i} \times \tau_2$ for $\nu = (\hat{e}_{j,i}, \hat{e}_{j,i-1}) \in \Delta^k$, $k = L^2$. When jobs attempt to change nodes there are two possible outcomes (successful and unsuccessful switching) which we will consider separately. The case in which a job successfully switches nodes is analogous to a job leaving the system but rates are multiplied by the proportion of nodes in the configuration to which the job is switching. Thus for a job switching from a node with j and i jobs to a node with j' and i' jobs (of types 1 and 2, respectively) we have $\bar{\Gamma}_n^k(r,\nu) = j \times nr^{j,i} \times \gamma_1 \times \frac{nr^{j',i'}}{n-1}$ where $\nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j-1,i} + \hat{e}_{j'+1,i'}) \in \Delta^k$, $k = C^1$ and $\bar{\Gamma}_n^k(r,\nu) = i \times nr^{j,i} \times \gamma_2 \times \frac{nr^{j',i'}}{n-1}$ for $\nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j,i-1} + \hat{e}_{j',i'+1}) \in \Delta^k$, $k = C^2$. Next consider unsuccessful switches. Recall that if a job attempts to switch to a node at which there is not

enough room, then the job is rejected from the system. The rate at which such jumps occur is, again, analogous to the previous scenario except we instead multiply by the proportion of nodes without enough room for the job attempting to move. Let r_i^C be the proportion of nodes without enough room to accommodate a job of type i (i.e. nodes in states (i', j') with $(j'A_1 + i'A_2 + A_i > C)$). Then $\bar{\Gamma}_n^k(r,\nu) = j \times nr^{j,i} \times \gamma_1 \times \frac{nr_1^C}{n-1}$ for $\nu = (\hat{e}_{j,i}, \hat{e}_{j-1,i}) \in \Delta^k,\ k = C^1$ and $\bar{\Gamma}_n^k(r,\nu) = i \times nr^{j,i} \times \gamma_2 \times \frac{nr_2^C}{n-1}$ for $\nu = (\hat{e}_{j,i}, \hat{e}_{j,i-1}) \in \Delta^k,\ k = C^2$.

With the above definition of $\bar{\Gamma}_n^k$, the generator of $\{\mu_n(t)\}$ is as given by (3.2). Γ^k is defined to be the limit of $\bar{\Gamma}_n^k$ which is simply given as

$$\Gamma^k(r,\nu) = \begin{cases} j \times r^{j,i} \times \gamma_1 \times r^{j',i'} & \text{for } \nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j-1,i} + \hat{e}_{j'+1,i'}) \in \Delta^k,\ k = C^1 \\ i \times r^{j,i} \times \gamma_2 \times r^{j',i'} & \text{for } \nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j,i-1} + \hat{e}_{j',i'+1}) \in \Delta^k,\ k = C^2 \\ j \times r^{j,i} \times \gamma_1 \times r_1^C & \text{for } \nu = (\hat{e}_{j,i}, \hat{e}_{j-1,i}) \in \Delta^k,\ k = C^1 \\ i \times r^{j,i} \times \gamma_2 \times r_2^C & \text{for } \nu = (\hat{e}_{j,i}, \hat{e}_{j,i-1}) \in \Delta^k,\ k = C^2 \\ \bar{\Gamma}_1^k(r,\nu) & \text{otherwise} \end{cases} \tag{3.74}$$

for $r \in \mathcal{S}$. Clearly $\Gamma^k(\cdot,\nu)$ is Lipschitz for all $k \in \mathbf{K}, \nu \in \Delta^k$ and (3.3) is satisfied so Condition 3.1.1 holds in this example. From Proposition 1 we then have that $\mu_n(t) \to \mu(t)$ uniformly on $[0,T]$ where $\dot{\mu}(t) = F(\mu(t))$ and F is as in (3.4), with Γ^k as defined above.

Now suppose that the arrival rates $\lambda_i,\ i = 1,2$ can be modulated by exercising an additive control with values in $\frac{1}{\sqrt{n}}[-D, D],\ D < \infty,\ i = 1,2$. One can also consider control of any of the other parameters $\{\tau_i,\ \gamma_i : i = 1,2\}$ but for simplicity we will only consider the control of the arrival rates. Let

$$\begin{aligned} \Lambda = \{u \in \mathbb{R}^{\ell_1} \times \{0\}^{\ell-\ell_1} | u_j = u_1^* \in [-D, D], j = 1, \ldots, |\Delta^{E^1}|, \\ u_k = u_2^* \in [-D, D], k = |\Delta^{E^1}| + 1, \ldots, |\Delta^{E^2}|\} \end{aligned} \tag{3.75}$$

where $\ell = \sum_{i=1}^2 \left(|\Delta^{E^i}| + |\Delta^{L^i}| + |\Delta^{C^i}|\right)$ and $\ell_1 = \sum_{i=1}^2 |\Delta^{E^i}|$. The controls will take values in $\Lambda_n = \frac{1}{\sqrt{n}}\Lambda$. For a $u \in \Lambda$ or Λ_n let u_1^* refer to the value of the first $|\Delta^{E_1}|$ coordinates and u_2^*

refer to the value of the next $|\Delta^{E^2}|$ coordinates. Define the controlled rate function as

$$\Gamma_n^k(r,u,\nu) = \begin{cases} nr^{j,i} \times (\lambda_1 + u_1^*) & \text{for } k = E^1,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j+1,i}) \in \Delta^{E^1} \\ nr^{j,i} \times (\lambda_2 + u_2^*) & \text{for } k = E^2,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j,i+1}) \in \Delta^{E^2} \\ \bar{\Gamma}_n^k(r,\nu) & \text{otherwise,} \end{cases} \tag{3.76}$$

where $u \in \Lambda_n$. Since controls in Λ_n are $\mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, Condition 3.1.2 is easily seen to be satisfied for the example.

From our assumption that $x_0^{j,i} > 0$ for all $(j,i) \in \mathbb{X}$, it follows that $\mu_t^{j,i} > 0$ for all $(j,i) \in \mathbb{X}$ and $0 \le t \le T$. Using this and the form of Γ^k given in (3.74), it is then easy to check that Condition 3.1.5 is satisfied. Similarly our assumption on the initial conditions in Theorem 2 is satisfied as well. Recalling the definitions of Γ_n^k and Γ^k in (3.76) and (3.74), respectively, we see that there exists a $\kappa \in (0,\infty)$ such that for all $y \in B(2\sqrt{n}), u \in \Lambda_n, \xi \in \mathcal{S}_n(y)$

$$\sqrt{n}\left(\frac{1}{n}\Gamma_n^k\left(\frac{1}{\sqrt{n}}y + \xi, u, \nu\right) - \Gamma^k(\xi,\nu)\right) \le \kappa(1 + \|y\|)$$

and therefore Condition 3.1.3 is satisfied. For $k \in \mathbf{K}, \nu \in \Delta^k$ define $h_1^k(\nu,\cdot) : \mathcal{S} \to \mathbb{R}$ as

$$h_1^k(\nu,r) = \begin{cases} r^{j,i} & \text{for } k = E^1,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j+1,i}) \in \Delta^{E^1} \\ r^{j,i} & \text{for } k = E^2,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j,i+1}) \in \Delta^{E^2} \\ 0 & \text{otherwise} \end{cases}$$

and $h_2^k(\nu,\cdot)$ as

$$\begin{cases} \lambda_1 \times e_{j,i} & \text{for } k = E^1,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j+1,i}) \in \Delta^{E^1} \\ \lambda_2 \times e_{j,i} & \text{for } k = E^2,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j,i+1}) \in \Delta^{E^2} \\ j \times \mu_1 \times e_{j,i} & \text{for } k = L^1,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j-1,i}) \in \Delta^{L^1} \\ i \times \mu_2 \times e_{j,i} & \text{for } k = L^2,\ \nu = (\hat{e}_{j,i}, \hat{e}_{j,i-1}) \in \Delta^{L^2} \\ j \times \gamma_1 \times (r^{j,i} \times e_{j',i'} + r^{j',i'} \times e_{j,i}) & \text{for } \nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j-1,i} + \hat{e}_{j'+1,i'}) \in \Delta^k,\ k = C^1 \\ i \times \gamma_2 \times (r^{j,i} \times e_{j',i'} + r^{j',i'} \times e_{j,i}) & \text{for } \nu = (\hat{e}_{j,i} + \hat{e}_{j',i'}, \hat{e}_{j,i-1} + \hat{e}_{j',i'+1}) \in \Delta^k,\ k = C^2 \\ j \times \gamma_1 \times (r^{j,i} \times e_C^1 + r_1^C \times e_{j,i}) & \text{for } \nu = (\hat{e}_{j,i}, \hat{e}_{j-1,i}) \in \Delta^k,\ k = C^1 \\ i \times \gamma_2 \times (r^{j,i} \times e_C^2 + r_2^C \times e_{j,i}) & \text{for } \nu = (\hat{e}_{j,i}, \hat{e}_{j,i-1}) \in \Delta^k,\ k = C^2. \end{cases}$$

Defining H^k, β_k^n as in Condition 3.1.4 with h_1^k and h_2^k we see that (3.14) is satisfied and thus Condition 3.1.4 holds for the example.

We now introduce the following finite time horizon cost

$$J^n(U^n, v_n) = \mathbb{E}\int_0^T (\|V_n(t)\|^2 + \alpha\|\sqrt{n}U^n(t)\|^2)dt,\ U^n \in \mathcal{A}_n, \tag{3.77}$$

where $\alpha \in (0,\infty)$. The cost function penalizes both the deviation from the nominal behavior and exercising rate control. Note that this cost function satisfies the condition introduced below (3.10). We have thus verified all the conditions needed for Theorem 2 and from this result it follows that a near optimal continuous feedback control for the diffusion control problem can be used to construct an asymptotically optimal sequence of control policies for this system. The diffusion control problem here takes the same form as (3.20) with η and β as in (3.15) and σ as in (3.18) with cost given as

$$J(U, v_0) = \mathbb{E}\int_0^T (\|V(t)\|^2 + \alpha\|U(t)\|^2)dt,\ U \in \mathcal{A}(\Xi). \tag{3.78}$$

This is the classical stochastic linear-quadratic regulator problem which has been well studied (cf. (Fleming and Rishel, 1976)). Replacing $[-D, D]$ with $\mathbb{R}$ in the definition of the control set

in (3.75), the optimal control for the limit stochastic LQR is given in feedback form as follows

$$u^*(s,y) = -B'(s)K^*(s)V(s)$$

where B is defined in terms of $\{h_1^k, k \in \mathbf{K}\}$ via the relation $\eta(t,u) = B(t)u$ and K^* solves an appropriate Riccati equation (see (Fleming and Rishel, 1976)). For implementing this feedback control for the prelimit system we truncate u^* suitably; such a modification, in practice, has little to no effect for large n. We construct U_g^n as in Section 3.1.4, by taking $U_g^n(t) = \sqrt{n}u^*(t, V_n(t))$.

We now present our numerical results. The above control policy was implemented (for $\alpha = .01$ and $.001$) on $n_{\text{trials}} = 128$ different realizations of the stochastic process with the following parameters $n = 10,000, T = 10, C = 6, A_1 = 1, A_2 = 1, \lambda_1 = 1, \lambda_2 = 1, \tau_1 = 1, \tau_2 = 1, \gamma_1 = 1, \gamma_2 = 1$. We also simulate 128 realizations of the corresponding uncontrolled system. Table 3.1 shows the averaged cost over the 128 simulations for the controlled and uncontrolled systems. The control policy based on the optimal feedback control for the stochastic LQR leads to a reduction in cost of 12.7% for $\alpha = .01$ and 15.5% for $\alpha = .001$. The deviations from the nominal values

Table 3.1. Cost over 128 Simulations

	Uncontrolled	Controlled with $\alpha = .01$	Controlled with $\alpha = .001$
Deviation Cost	8.9556	8.1271	7.5649
Control Cost	0	$.01 \times 25.37$	$.001 \times 256.8$
Total Cost	8.9556	8.3809	7.8217

under the controlled and uncontrolled systems are computed by calculating the average,

$$\frac{1}{n_{\text{trials}}} \sum_{i=1}^{n_{\text{trials}}} \int_0^T \|V_n(s)\|^2 ds$$

for the two systems and the cost of exercising control is computed by the average,

$$\alpha \times \frac{1}{n_{\text{trials}}} \sum_{i=1}^{n_{\text{trials}}} \int_0^T \|\sqrt{n}U^n(t)\|^2 ds.$$

The deviations are smaller for the controlled system as expected. In general, one can achieve higher reduction in such deviations by decreasing the parameter α in the cost function. In

practice the tuning parameter α suitably balances the cost of deviating from the nominal values and the cost for exercising control.

CHAPTER 4

Load Balancing Mechanisms in Cloud Storage Systems

In this chapter we are interested in developing a rigorous limit theory for a class of models used for the analysis of load balancing schemes in large cloud storage networks. We consider a system of n-serves storing a set of $I(n) = c\binom{n}{L}$ files using an MDS coding scheme with parameters L and k. Namely, each file is broken down into L chunks with each chunk being $\frac{1}{k}$-th the size of the original file. In addition, any subset of k chunks is sufficient for reconstructing the original file. Each server maintains its own FIFO queue and processes jobs at rate k. A stream of file requests arrive in the system at rate $n\lambda$. Each file request chooses a file uniformly at random and the files are distributed such that this request corresponds to the selection of L randomly chosen servers. A centralized dispatcher then routes the file request into the k shortest queues out of the L which are chosen. The evolution of the collection of queue lengths can be modeled as a continuous time Markov chain. The transition rates in the system scale with n and so, for large n, the state process of interest is jumping extremely quickly making a direct analysis intractable. In order to provide model simplifications we consider the behavior of the system as the number of queues approaches infinity. Under a suitable scaling we establish asymptotic approximations in the form of ODE and SDE. These continuous processes provide a more tractable means of analyzing the original system. For example, simulating the original system can take quite a long time for large n since every event in the system much be accounted for and such events are occurring extremely quickly. The limiting ODE and SDE can be discretized on a much coarser scale via numerical ODE solvers and Euler discretizations leading to a massive improvement in simulation time.

The starting point of our analysis is to consider, as the state descriptor, the empirical measure of the n queue lengths rather than the individual values of the queue lengths. Thus the state space for our system will be the space $\mathcal{P}_n(\mathbb{N}_0)$ of probability measures on $\mathbb{N}_0$ that assign weights in $\frac{1}{n}\mathbb{N}_0$ to sets in $\mathbb{N}_0$ rather than the space $\mathbb{R}_+^n$. With this formulation the state

processes for all n-server systems can be regarded as taking values in a common space $\mathcal{S} \doteq \mathcal{P}(\mathbb{N}_0)$ (the space of probability measures on $\mathbb{N}_0$). It follows from our symmetry assumptions that the state-evolution of the n-server system describes a pure-jump Markov process with values in $\mathcal{P}(\mathbb{N}_0)$ and thus one can bring to bear the theory of weak convergence of Markov processes to study scaling limits as n becomes large. In particular, in Theorem 10 we prove a law of large numbers for the empirical measure process $(\pi^n(t))_{0\le t\le T}$ as $n \to \infty$. We show that π^n converges to a deterministic function π in $\mathbb{D}([0,T] : \mathcal{S})$, where $\mathbb{D}([0,T] : \mathcal{S})$ is the space of functions from $[0,T]$ to $\mathcal{S}$ that are right continuous and have left limits, equipped with the usual Skorohod topology. We then show that the limiting ODE system the characterizes the limit π has a fixed point $\bar{\pi}$ which is stable. Namely, starting from an arbitrary initial condition, the solution to the ODE converges to this fixed point as $t \to \infty$. Instead of working with $\bar{\pi}$, it will instead be convenient to work with $\bar{u} = (\bar{u}_i)_{i\in\mathbb{N}_0}$ where $\bar{u}_i = \sum_{j=i}^{\infty} \bar{\pi}_j$ for each $i \in \mathbb{N}_0$. Intuitively, $\bar{u}_i$ represents the proportion of queues with length at least i. We also show that the queue length distribution given by the fixed point has tails which decay super-exponentially extending this well known property of the supermarket model (i.e. $k = 1$) to a general $k < L$. We give explicit upper and lower bounds (cf. Theorem 11) on these tails which are sharp in the sense that they coincide when $k = 1$. An important interchange of limits property is then established. In (Li et al., 2016), it has been shown that queue length process Q^n for the n-server system is positive recurrent and, thus, has a unique invariant probability measure. This then implies that the occupancy measure process has a unique invariant distribution. We show that this invariant measure converges to $\delta_{\bar{u}}$ in probability, as $n \to \infty$. Roughly speaking, this result says that the limits $n \to \infty$ and $t \to \infty$ can be interchanged and, in particular, the fixed point of the ODE is a good approximation for the steady state behavior of the occupancy process for large n.

Next we consider the fluctuation process $X^n \doteq \sqrt{n}(\pi^n - \pi)$. This process can be regarded as taking values in the space of signed measures on $\mathbb{N}_0$, however for an asymptotic analysis it is convenient to view it as taking values in the Hilbert space of square summable real sequences, $\boldsymbol{\ell}_2$. The study of the asymptotics of these fluctuations is the subject of Theorem 14 which shows that $X^n \doteq \sqrt{n}(\pi^n - \pi)$ converges in $\mathbb{D}([0,T] : \boldsymbol{\ell}_2)$ to a $\boldsymbol{\ell}_2$-valued diffusion process.

A basic assumption in our analysis of the fluctuations around the law of large number limit (see statement of Theorem 14) is a uniform (in n) bound on the second moment of the

empirical measure at time 0. This condition is not very stringent as in practice one may consider systems starting from empty or with finitely many jobs (independent of n). We argue that these integrability properties at time 0 propagate through to any finite future time T. Tightness of the scaled fluctuation processes X^n which is shown by establishing, uniform in n, second moment bounds (on X^n) and by employing criteria for tightness of Hilbert space-valued semimartingales (cf. (Joffe and Métivier, 1986), (Métivier, 1982)), relies on these integrability properties. Another ingredient in the proof of tightness is a suitable Lipschitz property of the map F introduced in (4.4) that enables the use of a Gronwall argument. For this argument one needs a Lipschitz estimate in the $\boldsymbol{\ell}_2$ norm, however, it is not clear that F, as a map from $\boldsymbol{\ell}_2$ to $\boldsymbol{\ell}_2$, is Lipschitz. We instead restrict attention to a smaller space

$$\mathcal{V}_M \doteq \left\{ r \in \boldsymbol{\ell}_2 : r_i \geq 0, \sum_{i=0}^{\infty} r_i = 1, \ \sum_{i=0}^{\infty} i r_i \leq M \right\}$$

and argue that for each M, the map F is Lipschitz from $\mathcal{V}_M$ to $\boldsymbol{\ell}_2$. This 'local' Lipschitz property plays an important role in the proof of Proposition 6.

For characterization of limit points in the proof of the central limit theorem, one needs to argue that the associated stochastic differential equation (SDE) in $\boldsymbol{\ell}_2$ (see (4.14)) has a unique weak solution in an appropriate class of processes. It turns out that arguing this uniqueness among adapted processes with paths in $\mathbb{C}([0,T] : \boldsymbol{\ell}_2)$ (the space of continuous functions from $[0,T]$ to $\boldsymbol{\ell}_2$) is not straightforward due to a lack of suitable regularity of the function G introduced in (4.19). In particular, once more, the Lipschitz property of the map $x \mapsto G(x,\pi)$ (for a fixed $\pi \in \mathcal{P}(\mathbb{N}_0)$) from $\boldsymbol{\ell}_2$ to itself is not immediate. The key observation here is that this map is Lipschitz when restricted to the space

$$\tilde{\boldsymbol{\ell}}_2 \doteq \{ x \in \boldsymbol{\ell}_2 : \sum_{j=0}^{\infty} j^2 x_j^2 < \infty, \sum_{j=0}^{\infty} x_j = 0 \}.$$

This observation, together with the property that the limit points X of $X^n = \sqrt{n}(\pi^n - \pi)$. satisfy $X(t) \in \tilde{\boldsymbol{\ell}}_2$ for all $t \geq 0$ almost surely, is key to the characterization of the limit points as the unique solution of the SDE (4.14) in a suitable class (see Proposition 4).

The chapter is organized as follows. In Section 4.1 we give a precise mathematical formulation of our model and a statement of our main results. Specifically, Theorem 10 provides the convergence in probability of the empirical measure process in $\mathbb{D}([0,T]:\mathcal{S})$ to the unique solution of the ODE defined in (4.7). The fixed point of the limiting ODE system is given in (4.12). Theorem 11 gives explicit upper and lower bounds on the rate of decay for the tail of the queue length distribution determined by (4.12). In Theorem 12 we show that (4.12) is, in fact, a stable fixed point of the ODE (3.5) and Theorem 13 presents the interchange of limits property discussed earlier. In Theorem 14, we give the main diffusion approximation result. This result says that the sequence of centered and scaled processes X^n, defined in (4.13), converges to the unique solution (in a suitable class) of the $\boldsymbol{\ell}_2$-valued SDE, driven by a cylindrical Brownian motion, given in (4.14). In Section 4.1.1 we record the corollaries of these results for the special setting of power-of-d schemes. The remainder of the chapter is devoted to proofs of the above results. In Section 4.2 we give a convenient representation of the state processes through a countable number of time-changed unit rate Poisson processes. Such Poisson representations have been used extensively (cf. (Kurtz, 1980; Kang et al., 2014; Anderson and Kurtz, 2015)) in the study of diffusion approximations for pure jump processes. Using this we obtain a semimartingale decomposition (see (4.23)) that is central to our analysis. Section 4.3 is devoted to the study of asymptotic behavior under the LLN scaling. In Section 4.3.1 we prove tightness of the sequence of state processes $\{\pi^n\}_{n\in\mathbb{N}}$ (see Proposition 5) and the proof of Theorem 10 is completed in Section 4.3.2. In Section 4.3.3 we prove a lemma which will be needed in the proof of Theorems 12 and 13. Proofs of Theorems 11, 12, and 13 are then given in Sections 4.3.4, 4.3.5, and 4.3.6, respectively. Section 4.4 proves Theorem 14. In Section 4.4.1 we prove the propagation of integrability properties that was discussed earlier and in Section 4.4.2 (see Proposition 6) we prove the key tightness property for the sequence of processes $\{X^n\}_{n\in\mathbb{N}}$ which relies on the Lipschitz property of F, in the $\boldsymbol{\ell}_2$ norm, on $\mathcal{V}_M$ (Lemma 14). Theorem 14 is then proved in Section 4.4.3. Finally, in Section 4.5, we present some numerical results. In particular, we use our results to give numerical confidence intervals for several performance measures of interest and compare the results to those obtained from a direct simulation of the corresponding n-server systems.

4.1 Model Description and Main Result

We consider a system with n servers each with its own infinite capacity queue. In all, there are $I(n)$ equally sized files stored over the n servers. Each file is stored in equally sized pieces at L servers such that any k pieces can reconstruct the original file. The files are distributed such that each combination of L servers has exactly c files. This, in particular, implies $I(n) = c\binom{n}{L}$. Jobs arrive from outside according to a Poisson process with rate $n\lambda$ and request one of the $I(n)$ files uniformly at random. Such a request corresponds to selection of one of the $\binom{n}{L}$ sets of L servers, uniformly at random, which is the set of servers containing the pieces of the requested file. The job is then routed to the k shortest queues among this set of L servers. Each server processes queued jobs according to the first-in-first-out (FIFO) discipline. Processing times at each server are mutually independent and exponentially distribution with mean k^{-1}.

Let $Q^n(t) = \{Q_i^n(t)\}_{i=1}^n$ where $Q_i^n(t)$ represents the length of the i-th queue at time t and let $\pi^n(t) = \{\pi_i^n(t)\}_{i\in\mathbb{N}_0}$ where $\pi_i^n(t)$ represents the proportion of queues with length exactly i at time t. This can explicitly be written as

$$\pi_i^n(t) = \frac{1}{n}\sum_{j=1}^{n} 1_{\{Q_j^n(t)=i\}}. \tag{4.1}$$

It will be convenient to work with the process $u^n(t) = \{u_i^n\}_{i\in\mathbb{N}_0}$ where $u_i^n(t)$ represents the proportion of queues with length at least i. Namely, $u_i^n(t) = \sum_{j=i}^{\infty}\pi_j^n(t)$. We will assume for simplicity that $Q^n(0) = q^n$ is nonrandom and thus $\pi^n(0)$ and $u^n(0)$ are nonrandom as well. We identify $\mathcal{P}(\mathbb{N}_0)$ with the infinite dimensional simplex $\mathcal{S} = \{s \in \mathbb{R}_+^\infty | \sum_{i=0}^{\infty} s_i = 1\}$ and let $\mathcal{S}_n = \frac{1}{n}\mathbb{N}_0^\infty \cap \mathcal{S}$. The spaces $\mathcal{S}$ and $\mathcal{S}_n$ can be identified with subsets of $\bar{\mathcal{U}} = \{u \in \mathbb{R}_+^\infty | 1 = u_0 \geq u_1 \geq \ldots \geq 0\}$ and $\bar{\mathcal{U}}_n = \{u \in \bar{\mathcal{U}} | u_i = r_i/n, r_i \in \mathbb{Z}\}$, respectively, each endowed with the product metric,

$$\rho(x,y) \doteq \sum_{j=1}^{\infty}\frac{|x_j - y_j|}{2^j}.$$

The identification map $\iota : \mathcal{S} \to \bar{\mathcal{U}}$ is defined as $\iota(p)_j \doteq \sum_{k=j}^{\infty} p_k,\ j \in \mathbb{N}_0,\ p \in \mathcal{S}$. Note that for $p^n, p \in \mathcal{S}$, $d_0(p^n, p) \to 0$ if and only if $\rho(\iota(p^n), \iota(p)) \to 0$. Additionally, note that $\pi^n(t) \in \mathcal{S}_n$ and $u^n(t) \in \bar{\mathcal{U}}_n$ for all $t \in [0,T]$. Let $\Sigma = \{\ell = (\ell_i)_{i=1}^L \in \mathbb{N}_0^L | \ell_1 \leq \ell_2 \leq \cdots \leq \ell_L\}$ and for $\ell \in \Sigma$ define $\rho_i(\ell) \doteq \sum_{j=1}^{L} 1_{\{\ell_j = i\}},\ i \in \mathbb{N}_0$. Roughly speaking, Σ will represent the set of possible states

for L selected queues arranged by non-decreasing queue length. Note that each file will be stored at L servers and that at any given time t the queue lengths of these L servers (up to a reordering) will correspond to an element in Σ. We will refer to the elements of Σ as "queue length configurations". Given a configuration $\boldsymbol{\ell} \in \Sigma$, $\rho_i(\ell)$ gives the number of queues of length i (among the L selected). From the above description of the system it follows that the empirical measure process, $\pi^n(t)$, is a continuous time Markov chain with state space $\mathcal{S}_n$ and generator

$$\begin{aligned}\mathcal{L}^n f(r) = \frac{n\lambda}{\binom{n}{L}} \sum_{\ell \in \Sigma} \left(\prod_{i=0}^{\infty} \binom{nr_i}{\rho_i(\ell)} \right) \left[f\left(r + \frac{1}{n} \Delta_\ell \right) - f(r) \right] \\ + k \sum_{i=1}^{\infty} nr_i \left[f\left(r + \frac{1}{n}(e_{i-1} - e_i) \right) - f(r) \right],\end{aligned} \tag{4.2}$$

for $f : \mathcal{S}_n \to \mathbb{R}$ where

$$\Delta_\ell \doteq \sum_{i=1}^{k} e_{\ell_i + 1} - \sum_{i=1}^{k} e_{\ell_i} \tag{4.3}$$

and for $y \in \mathbb{N}_0$, $e_y \in \boldsymbol{\ell}_2$ is a vector with 1 at the y-th coordinate and 0 elsewhere. Here we use the standard conventions that $0^0 = \binom{0}{0} = 0! = 1$, and $\binom{a}{b} = 0$ when $a < b$. The above generator can be understood as follows. A typical term in the second expression corresponds to a jump as a result of a server, with exactly i jobs queued, completing a job. The term in the square brackets gives the change in value of f as a result of such a jump and the prefactor knr_i corresponds to the fact that servers process jobs at rate k and there are in all nr_i queues (prior to the jump) with exactly i jobs. The first expression in (4.2) corresponds to a jump resulting from an arrival of a job to the system. Typically, such an arrival makes a request for L servers with queue length configuration $\ell_1 \le \ell_2 \le \cdots \le \ell_L$ and results in the jump $\frac{1}{n}\Delta_\ell$. The sum in (4.3) only goes up to k (instead of L) since only the smallest k queues are affected by such a jump. Since prior to the jump, there are nr_i queues with exactly i jobs, the overall rate associated with the configuration $\ell = \{\ell_1 \le \ell_2 \le \cdots \le \ell_L\} \in \Sigma$ equals

$$\frac{n\lambda}{\binom{n}{L}} \left(\prod_{i=0}^{\infty} \binom{nr_i}{\rho_i(\ell)} \right).$$

In our setting the first entry in an element of $\boldsymbol{\ell}_2$ will typically correspond to the number of empty queues and thus we refer to it as the "0-th" coordinate and any $r \in \boldsymbol{\ell}_2$ will correspond to a vector of the form $(r_0, r_1, \ldots)$. For notational convenience, for $r \in \boldsymbol{\ell}_2$ we set $r_{-1} \doteq 0$.

The main results in this chapter provide scaling limits for π^n. We first present the law of large numbers which describes the nominal state of the system for large n. Define, for $r \in \boldsymbol{\ell}_1$,

$$F(r) \doteq \lambda L! \sum_{j=0}^{\infty} \bar{\zeta}^{\delta}(j,r) e_j + k \sum_{j=0}^{\infty} [r_{j+1} - r_j] e_j + r_0 e_0 \tag{4.4}$$

where

$$\bar{\zeta}^{\delta}(j,r) \doteq \bar{\zeta}(j-1,r) - \bar{\zeta}(j,r)$$

and, adopting the convention that $\sum_{i=b}^{a} x_i = 0$ for $a < b$,

$$\bar{\zeta}(j,r) \doteq \sum_{i_1=0}^{k-1} \frac{\left(\sum_{m=0}^{j-1} r_m\right)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k-i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{\left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2}}{(L-i_1-i_2)!}. \tag{4.5}$$

For $j \geq 0$, the quantities $k[r_{j+1} - r_j]$ in (4.4) roughly represent the rate at which the j-th coordinate of the state changes (in the limit) as a result of job-completions while the quantity $\lambda L!(\bar{\zeta}(j-1,r) - \bar{\zeta}(j,r))$ represents a similar quantity as a result of job-arrivals. The various terms in (4.5) can be interpreted as follows. An arrival to a queue with j jobs implies that a queue length configuration vector $\ell = \{\ell_1 \leq \ell_2 \leq \cdots \leq \ell_L\}$ was selected which has the property that at least one of the k smallest ℓ_i's equals j, or equivalently, exactly i_1 ($i_1 = 0, 1, \ldots, k-1$) of the smallest L selected are less than j, i_2 ($i_2 = 1, \ldots L - i_1$) of these are equal to j, and $L - i_1 - i_2$ are greater than j. The three ratios in (4.5) are contributions from these three types of queues. The term $[i_2 \wedge (k - i_1)]$ is from the fact that only the smallest k of the L queues are affected.

Also observe that for some $c_\zeta \in (0, \infty)$

$$\bar{\zeta}(j,r) \leq c_\zeta r_j \text{ for all } j \in \mathbb{N}_0 \text{ and } r = (r_j)_{j=0}^{\infty} \in \mathcal{S}. \tag{4.6}$$

Thus the infinite sum in (4.4) is well defined since $\sum_{j=0}^{\infty} r_j = 1$ and consequently F is a well defined map from $\mathcal{S}$ to $\boldsymbol{\ell}_1$. A similar estimate shows that F is a well defined map from $\boldsymbol{\ell}_1$ to $\boldsymbol{\ell}_1$ and $\sum_{j=0}^{\infty} F_j(r) = 0$ for all $r \in \boldsymbol{\ell}_1$.

Consider the system of ODEs

$$\dot{\pi}(t) = F(\pi(t)), \qquad \pi(0) = \pi_0 \tag{4.7}$$

where F is defined in (4.4) and $\pi_0 \in \mathcal{S}$. The solution of the equation is a continuous map $\pi : [0,T] \to \mathcal{S}$ such that

$$\pi(t) = \pi_0 + \int_0^t F(\pi(s))ds,\ t \in [0,T] \tag{4.8}$$

where the integral on the right side is the classical Bochner integral which is well defined since, from (4.4) and (4.6),

$$\sup_{0\le s\le T} \|F(\pi(s))\|_1 \le \sup_{r\in\mathcal{S}} \|F(r)\|_1 < \infty. \tag{4.9}$$

Equation (4.7) will characterize the law of large number limit of π^n.

The following result on the wellposedness of (4.7) will be shown in Section 4.3.2.

Proposition 3. *Let $\pi_0 \in \mathcal{S}$. Then there exists a $\pi \in \mathbb{C}([0,T] : \mathcal{S})$ that solves (4.7). Furthermore, if $\pi, \tilde{\pi}$ are two elements of $\mathbb{C}([0,T] : \mathcal{S})$ solving (4.7) with $\pi(0) = \tilde{\pi}(0) = \pi_0$, then $\pi = \tilde{\pi}$.*

The next theorem gives a law of large numbers for the sequence $\{\pi^n\}_{n\in\mathbb{N}}$. Recall we take $\pi^n(0)$ to be nonrandom.

Theorem 10. *Suppose that $\pi^n(0) \to \pi_0$, in $\mathcal{S}$, as $n \to \infty$. Then $\pi^n \to \pi$, in probability, in $\mathbb{D}([0,T] : \mathcal{S})$ where π is the unique solution of (4.7) in $\mathbb{C}([0,T] : \mathcal{S})$.*

Proof of Theorem 10 will be given in Section 4.3.2.

We now consider the long-time behavior of π^n. Following (Li et al., 2016), let $f \equiv f^{(L,k)} : [0,1] \to \mathbb{R}$ be defined as

$$f(x) \doteq \sum_{i=1}^{k} \binom{L}{L-k+i} \binom{L-k+i-2}{i-1} (-1)^{i-1} x^{L-k+i}.$$

The following lemma gives a representation for $\bar{\zeta}$ in terms of f.

Lemma 4. *Fix $r \in \mathcal{S}$ and let $u = \iota(r)$, i.e. $u_m = \sum_{i=m}^{\infty} r_i$, $m \in \mathbb{N}_0$. Then, for $j \in \mathbb{N}_0$*

$$L!\bar{\zeta}(j,r) = f(u_j) - f(u_{j+1}). \tag{4.10}$$

Proof of this lemma will be give in Section 4.3.3. It follows from Lemma 4 and Theorem 10 that the law of large number limit of u^n solves the following ODE,

$$\dot{u}_j(t) = \lambda[f(u_{j-1}(t)) - f(u_j(t))] - k[u_j(t) - u_{j+1}(t)], \qquad u(0) = g \in \bar{\mathcal{U}}. \tag{4.11}$$

Consider the queue length distribution $\bar{u} = (\bar{u}_m)_{m\in\mathbb{N}_0}$ defined recursively through,

$$\begin{cases} \bar{u}_{m+1} = \lambda \frac{f(\bar{u}_m)}{k} & \text{for } m \in \mathbb{N}_0 \\ \bar{u}_0 = 1 \end{cases} \tag{4.12}$$

We will see in Theorem 12 that $\bar{u}$ is the unique fixed point of (4.11). The following result shows that the vector $(\bar{u}_m)_{m\in\mathbb{N}_0}$ which, roughly speaking, represents the steady state distribution of the queue lengths for large n, decays super-exponentially in m with rate determined by L and k.

Theorem 11. *Suppose $\bar{u}$ satisfies* (4.12)*. Then the following upper and lower bounds hold:*

i) $\bar{u}_m \le \lambda^{\frac{(L/k)^m - 1}{L/k - 1}}$ *for all* $m \in \mathbb{N}_0$.

ii) $\bar{u}_m \ge \lambda^{\frac{(L-k+1)^m - 1}{L-k}}$ *for all* $m \in \mathbb{N}_0$.

We note that the bounds are tight in the sense that when $k = 1$ the upper and lower bounds agree. Proof of this theorem is given in Section 4.3.4. Since f is a polynomial it is easy to see that $f(x) = \mathcal{O}(x^{L-k+1})$ as $x \to 0$. Intuitively, it makes sense that the queue length distribution should have an upper bound of the form $\lambda^{\frac{(L-k+1)^m - 1}{L-k}}$. Indeed, we can establish an upper bound of this form for large m, however due to the higher order terms in f the bound will not hold for small m. In fact, the threshold for a large enough m will depend on L and k. Furthermore, the coefficient of x^{L-k+1} in f depends on L and k and, using its form, it can be shown that the upper bound (for large m) will be of the form $a^{\frac{(L-k+1)^m - 1}{L-k}}$ where a depends on L and k. Recall that the routing scheme considered here corresponds to the well-known "Power-of-d" or super market model when $L = d$ and $k = 1$. The above result reduces to results in (Graham, 2000) and (Vvedenskaya et al., 1996) in this case.

Following (Vvedenskaya et al., 1996), define

$$v_j(u) = \sum_{i=j}^{\infty} u_i, \quad u \in \bar{\mathcal{U}}.$$

Let $\mathcal{U} \doteq \{u \in \bar{\mathcal{U}} | v_1(u) < \infty\}$ and note that this can be identified with the space of probability measures on $\mathbb{N}_0$ with finite first moment. The space $\mathcal{U}$ is endowed with the topology inherited from $\bar{\mathcal{U}}$. We now characterize the long time behavior of the law of large number limit. Note that $\bar{u} \in \mathcal{U}$. The next theorem shows that $\bar{u}$ is the unique fixed point in $\mathcal{U}$ for the system defined by (4.11) and this fixed point is, in fact, stable.

Theorem 12. *Suppose $\lambda < 1$ and u is a solution to (4.11) with $g \in \mathcal{U}$. Then*

i) $u(t) \in \mathcal{U}$ for all t.

ii) For each $j \in \mathbb{N}_0$, $\lim_{t\to\infty}(u_j(t) - \bar{u}_j) = 0$ and thus $\lim_{t\to\infty} \rho(u(t), \bar{u}) = 0$. In particular, $\bar{u}$ is the unique fixed point of (4.11) in $\mathcal{U}$.

The proof of this theorem will be given in Section 4.3.5.

From Proposition 1 of (Li et al., 2016) the process Q^n is positive recurrent and, thus, has a unique invariant distribution $\tilde{\mathcal{L}}_n \in \mathcal{P}(\mathbb{N}_0^n)$. Note that $\tilde{\mathcal{L}}_n$ can be identified with a measure $\mathcal{L}_n \in \mathcal{P}(\bar{\mathcal{U}}_n)$ which is an invariant measure for u^n. Furthermore, for any $t \geq 0$, $u^n(t)$ can be mapped to $\tilde{Q}^n(t) \in \mathbb{N}_0$ which is equal (up to a relabeling) to $Q^n(t)$. Due to symmetry, Q^n and $\tilde{Q}^n$ must have the same invariant distribution. Therefore $\mathcal{L}_n$ is the unique invariant measure for u^n. The following result shows that this invariant measure converges, as $n \to \infty$, to the Dirac measure concentrated at $\bar{u}$.

Theorem 13. *Let $\mathcal{L}_n$ be the unique invariant distribution for the process u^n. Then $\mathcal{L}_n \Rightarrow \delta_{\bar{u}}$. Furthermore, we have*

$$\lim_{n\to\infty} \lim_{t\to\infty} \mathbb{E}u^n(t) = \bar{u}.$$

Proof of Theorem 13 is given in Section 4.3.6.

We now study the fluctuations of π^n from its law of large number limit. Consider

$$X^n(t) = \sqrt{n}[\pi^n(t) - \pi(t)],\ t \in [0,T]. \tag{4.13}$$

where π^n is the state process introduced in (4.1) and π is the unique solution of (4.7) in $\mathbb{C}([0,T] : \mathcal{S})$.

We will show that, under conditions, X^n converges in distribution in $\mathbb{D}([0,T] : \boldsymbol{\ell}_2)$ to a stochastic process that can be characterized as the solution of a stochastic differential equation (SDE) of the following form.

$$dX(t) = G(X(t), \pi(t))dt + a(t)dW(t), \qquad X(0) = x_0. \tag{4.14}$$

The equation is again interpreted in the integrated form,

$$X(t) = x_0 + \int_0^t G(X(s), \pi(s))ds + \int_0^t a(s)dW(s), \qquad t \in [0,T]. \tag{4.15}$$

In the above equations, a is a measurable map from $[0,T]$ to the space of Hilbert-Schmidt operators from ℓ_2 to ℓ_2 such that $\int_0^T \|a(t)\|_{\text{HS}}^2 dt < \infty$, where $\|\cdot\|_{\text{HS}}$ denotes the Hilbert-Schmidt norm (see Appendix B), and W is a $\boldsymbol{\ell}_2$-cylindrical Brownian motion. Precise definitions are given in Appendix C, but roughly speaking, W can be identified with an iid sequence $\{\beta_i\}_{i\in\mathbb{N}_0}$ of standard real Brownian motions over $[0,T]$ and the stochastic integral $\int_0^t a(s)dW(s)$ represents a $\boldsymbol{\ell}_2$-valued Gaussian martingale $M(t)$ given as

$$M_i(t) = \sum_{j=0}^{\infty} \int_0^t A_{ij}(s)d\beta_j(s),\ t \in [0,T],\ i \in \mathbb{N}_0, \tag{4.16}$$

where $A_{ij}(s) = \langle e_i, a(s)e_j\rangle_2$, $s \in [0,T]$, $i,j \in \mathbb{N}_0$. We refer the reader to Chapter 4 of (Da Prato and Zabczyk, 2014) for construction and properties of the stochastic integral in (4.15). The Hilbert-Schmidt and integrability property of a ensure that the infinite sum in (4.16) converges. The operator $a(t)$ is determined from the system parameters and the law of large number limit

π in Theorem 10 as the symmetric square root of the following non-negative trace class operator

$$\Phi(t) \doteq \lambda L! \sum_{\ell \in \Sigma} \Delta_\ell \Delta_\ell^T \prod_{i=0}^{\infty} \frac{\pi_i(t)^{\rho_i(\ell)}}{\rho_i(\ell)!} + k \sum_{i=1}^{\infty} (e_{i-1} - e_i)(e_{i-1} - e_i)^T \pi_i(t). \tag{4.17}$$

The trace class property of $\Phi(t)$ and the integrability of the squared Hilbert-Schmidt norm of $a(t)$ are shown in Lemma 17. Define the space $\tilde{\boldsymbol{\ell}}_2 \subset \boldsymbol{\ell}_2$ as

$$\tilde{\boldsymbol{\ell}}_2 \doteq \{x \in \boldsymbol{\ell}_2 : \sum_{j=0}^{\infty} j^2 x_j^2 < \infty, \sum_{j=0}^{\infty} x_j = 0\}. \tag{4.18}$$

In (4.14) G is a map from $\tilde{\boldsymbol{\ell}}_2 \times \mathcal{S}$ to $\boldsymbol{\ell}_2$ defined as

$$G_i(x, r) \doteq \frac{\partial}{\partial u} F_i(r + ux)\Big|_{u=0} \qquad i \in \mathbb{N}_0,\ u \in \mathbb{R}. \tag{4.19}$$

One of the difficulties in the analysis is that G as a map from $\boldsymbol{\ell}_2 \times \mathcal{S}$ to $\boldsymbol{\ell}_2$ is not well behaved and we need to restrict attention to the smaller space $\tilde{\boldsymbol{\ell}}_2 \times \mathcal{S}$ in order to get unique solvability of (4.14). Note that under the condition $\sum_{j=0}^{\infty} j^2 x_j^2 < \infty$, the series $\sum_{j=0}^{\infty} |x_j| < \infty$ and thus the series $\sum_{j=0}^{\infty} x_j$ is convergent. Additionally, the right side of (4.19) is well defined for every $x \in \tilde{\boldsymbol{\ell}}_2$ and $r \in \mathcal{S}$, since for each $j \in \mathbb{N}_0$ and $r \in \boldsymbol{\ell}_1$ with $\sum_{i=0}^{\infty} r_i = 1$, $r \mapsto F_j(r)$ is a polynomial in $(r_0, r_1, \ldots, r_{j+1})$ given as

$$F_j(r) = \lambda L! [\bar{\zeta}(j-1, r) - \bar{\zeta}(j, r)] + k(r_{j+1} - r_j)$$

where

$$\bar{\zeta}(j, r) = \sum_{i_1=0}^{k-1} \frac{\left(\sum_{m=0}^{j-1} r_m\right)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k - i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{\left(1 - \sum_{m=0}^{j} r_m\right)^{L-i_1-i_2}}{(L - i_1 - i_2)!}.$$

Also, from (4.4) and (4.5) it is easily checked that there is a $c \in (0, \infty)$ such that for all $x \in \tilde{\boldsymbol{\ell}}_2$ and $r \in \mathcal{S}$

$$|G_i(x, r)| \le c \left[|x_{i-1}| + |x_i| + |x_{i+1}| + (r_{i-1} + r_i) \sum_{m=0}^{\infty} |x_m| \right].$$

This in particular implies that $G(x, r) \doteq (G_i(x, r))_{i \in \mathbb{N}_0} \in \boldsymbol{\ell}_1 \subset \boldsymbol{\ell}_2$ for all $(x, r) \in \tilde{\boldsymbol{\ell}}_2 \times \mathcal{S}$.

The following result shows the well-posedness of (4.15). The definition of an $\boldsymbol{\ell}_2$-cylindrical Brownian motion is given in Section C.

Proposition 4. *There exists a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ on which is given a $\boldsymbol{\ell}_2$-cylindrical Brownian motion W and a continuous $\{\mathcal{F}_t\}$-adapted process $(X(t))_{0\le t\le T}$ with sample paths in $\mathbb{C}([0,T]:\boldsymbol{\ell}_2)$ that satisfies the integral equation (4.15) and is such that $X(t) \in \tilde{\boldsymbol{\ell}}_2 \subset \boldsymbol{\ell}_2$ for all $t \in [0,T]$ almost surely. Furthermore if $\{\tilde{X}_t\}_{0\le t\le T}$ is another such process then $\tilde{X}_t = X_t$ for all $t \in [0,T]$, almost surely.*

The above result establishes weak existence and pathwise uniqueness of (4.15). By a standard argument (cf. (Ikeda and Watanabe, 1989, Section IV.1)) it follows that (4.15) has a unique weak solution. We can now present our main result on fluctuations of π^n. Recall that $X^n(0) = \sqrt{n}(\pi^n(0) - \pi_0)$ is deterministic.

Theorem 14. *Suppose $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^2 \pi_j^n(0) < \infty$ and $\pi^n(0) \to \pi_0$ in $\mathcal{S}$ as $n \to \infty$. Let π be the unique solution of (4.7) and, with X^n defined as in (4.13), $X^n(0) \to x_0$ in $\boldsymbol{\ell}_2$. In addition, suppose that*

$$\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^2 (X_j^n(0))^2 < \infty. \tag{4.20}$$

Then $X^n \Rightarrow X$ in $\mathbb{D}([0,T]:\boldsymbol{\ell}_2)$ where X is the unique weak solution to (4.14) given by Proposition 4.

Proposition 4 and Theorem 14 will be proved in Section 4.4. In Section 4.5 we will describe how Theorems 10 and 14 can be used for numerical computation of various performance measures using simulation of diffusion processes.

4.1.1 Supermarket Model

Consider a system of n servers, each with its own queue. Jobs arrive in the system according to a Poisson process with rate $n\lambda$. When a job enters the system, d servers are chosen uniformly at random and the job is routed to the shortest of the d selected queues. All servers process jobs according to the FIFO discipline. Service times are mutually independent and exponentially distributed with mean 1. This model has been well studied and is known as Power-of-d routing or the "Supermarket Model" (see (Vvedenskaya et al., 1996; Mitzenmacher, 2001; Graham, 2000)). The model is a special case of the system considered in the current chapter, corresponding to

$L = d$ and $k = 1$. Theorems 10 and 14 then provide, as corollaries, the following law of large numbers and central limit theorem for the Power-of-d routing scheme.

Define by π_d^n the empirical measure process of queue lengths in the Power-of-d system. For $r \in \boldsymbol{\ell}_1$, define

$$F_d(r) \doteq \lambda \left[\sum_{i=1}^{d} \binom{d}{i} r_{j-1}^i \left(\sum_{m=j}^{\infty} r_m \right)^{d-i} - \sum_{i=1}^{d} \binom{d}{i} r_j^i \left(\sum_{m=j+1}^{\infty} r_m \right)^{d-i} \right] e_j + \sum_{j=0}^{\infty} [r_{j+1} - r_j] e_j.$$

The following is a direct corollary of Theorem 10.

Corollary 1. *Suppose that $\pi_d^n(0) \to \pi_d(0)$, in $\mathcal{S}$, as $n \to \infty$. Then $\pi_d^n \to \pi_d$, in probability, in $\mathbb{D}([0,T] : \mathcal{S})$ where π_d is the unique solution in $\mathbb{C}([0,T] : \mathcal{S})$ to the following ODE*

$$\dot{\pi}_d(t) = F_d(\pi_d(t)), \qquad \pi_d(0) = \pi_0.$$

Remark 4.1.1. This result has been established in (Graham, 2000) (see Theorem 3.4 therein). In particular, it is easy to verify that $v_m(t) \doteq \sum_{j=m}^{\infty} (\pi_d(t))_j$ is the same function as in (3.9) of (Graham, 2000) (see also (Vvedenskaya et al., 1996)).

Our second corollary studies the fluctuations of π_d^n from its law of large number limit. Consider

$$X_d^n(t) = \sqrt{n}[\pi_d^n(t) - \pi_d(t)], \ t \in [0,T].$$

Analogous to $a(t)$ introduced in (4.14), let $a_d(t)$ be the symmetric square root of the following non-negative operator

$$\begin{aligned} \Phi_d(t) \doteq \lambda \sum_{j=0}^{\infty} (e_{j+1} - e_j)(e_{j+1} - e_j)^T \left(\sum_{i=1}^{d} \binom{d}{i} [(\pi_d)_j(t)]^i \left(\sum_{m=j+1}^{\infty} (\pi_d)_m(t) \right)^{d-i} \right) \\ + \sum_{j=1}^{\infty} (e_{j-1} - e_j)(e_{j-1} - e_j)^T (\pi_d)_j(t). \end{aligned} \tag{4.21}$$

Analogous to G in (4.19), let G_d be a map from $\tilde{\boldsymbol{\ell}}_2 \times \mathcal{S}$ to $\boldsymbol{\ell}_2$, where $\tilde{\boldsymbol{\ell}}_2$ is as in (4.18), defined as

$$(G_d)_i(x,r) \doteq \frac{\partial}{\partial u} (F_d)_i(r + ux) \Big|_{u=0} \qquad i \in \mathbb{N}_0, \ u \in \mathbb{R}. \tag{4.22}$$

In the special case that $d = 2$, this function simply reduces to

$$(G_2)_i(x,r) = 2\lambda \sum_{m=i}^{\infty} [x_{i-1}r_m + r_{i-1}x_m - x_i r_{m+1} - r_i x_{m+1}] + (x_{i+1} - x_i).$$

The following result is immediate from Theorem 14.

Corollary 2. *Suppose* $\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^2(\pi_d^n)_j(0) < \infty$ *and* $\pi_d^n(0) \to \pi_0$ *in* $\mathcal{S}$ *as* $n \to \infty$. *Also, suppose* $X_d^n(0) = \sqrt{n}[\pi_d^n(0) - \pi_0] \to x_0$ *in probability in* $\boldsymbol{\ell}_2$ *and that*

$$\sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^2((X_d^n)_j(0))^2 < \infty.$$

Then $X_d^n \Rightarrow X_d$ *in* $\mathbb{D}([0,T] : \boldsymbol{\ell}_2)$ *where* X_d *is the unique weak solution to* (4.14) *with values in* $\tilde{\boldsymbol{\ell}}_2$, *with* G *replaced by* G_d *defined by* (4.22) *and* $a(t)$ *replaced by* $a_d(t)$ *which is given as the symmetric square root of the operator* $\Phi_d(t)$ *in* (4.21).

4.2 Semimartingale Representation

In this section we write the state processes using compensated time-changed Poisson processes to give a semimartingale representation for the system. Let $\{N_\ell,\ \ell \in \Sigma\}$ and $\{D_i,\ i \in \mathbb{N}_0\}$ be collections of mutually independent unit rate Poisson processes. The process N_ℓ will be used to represent the stream of jobs requesting files which are stored at servers with queue length configuration (immediately before the time of arrival of the request) $\ell = (\ell_1, \ldots, \ell_L)$. Similarly D_i will represent the stream of jobs completed by servers whose queue length (immediately before the time of completion) is equal to i. From the form of the generator in (4.2) we see that the state process π^n can be expressed as,

$$\pi^n(t) = \pi^n(0) + \frac{1}{n}\sum_{\ell\in\Sigma} \Delta_\ell N_\ell\left(\int_0^t \frac{n\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds\right) + \frac{1}{n}\sum_{i=1}^{\infty}(e_{i-1} - e_i) D_i\left(k\int_0^t n\pi_i^n(s)ds\right).$$

By adding and subtracting the compensators of the Poisson processes one can write the state process as a semimartingale. Namely,

$$\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t) \tag{4.23}$$

where

$$A^n(t) \doteq \sum_{\ell\in\Sigma} \Delta_\ell \int_0^t \frac{\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} ds + k\sum_{i=1}^{\infty}(e_{i-1}-e_i)\int_0^t \pi_i^n(s)ds \tag{4.24}$$

and

$$\begin{aligned} M^n(t) \doteq \sum_{\ell\in\Sigma} \frac{1}{n}\Delta_\ell N_\ell\left(\frac{n\lambda}{\binom{n}{L}}\int_0^t \prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)}ds\right) - \sum_{\ell\in\Sigma}\Delta_\ell \frac{\lambda}{\binom{n}{L}}\int_0^t\prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)}ds \\ + \sum_{i=1}^{\infty}\frac{1}{n}(e_{i-1}-e_i)D_i\left(k\int_0^t n\pi_i^n(s)ds\right) - k\sum_{i=1}^{\infty}(e_{i-1}-e_i)\int_0^t \pi_i^n(s)ds. \end{aligned} \tag{4.25}$$

From (4.46) and (4.73), it follows that for some $c_\zeta \in (0,\infty)$

$$A_j^n(t) \le \int_0^t \left(\frac{\lambda}{\binom{n}{L}} c_\zeta n^L[\pi_{j-1}^n(s)+\pi_j^n(s)] + k[\pi_{j+1}^n(s)+\pi_j^n(s)]\right)ds$$

for all $t\in[0,T]$, $n\in\mathbb{N}$, and $j\in\mathbb{N}_0$. Thus, there exists a $\kappa\in(0,\infty)$ such that

$$\sum_{j=0}^{\infty} A_j^n(t)^2 \le \kappa\sum_{j=0}^{\infty}\int_0^t [\pi_{j-1}^n(s)^2+\pi_{j+1}^n(s)^2+\pi_j^n(s)^2]ds \le 3\kappa t$$

for all $t\in[0,T]$. Consequently both $M^n(t)$ and $A^n(t)$ take values in $\boldsymbol{\ell}_2$. A similar argument shows that $A^n(t)$ in fact takes values in $\boldsymbol{\ell}_1$.

Similarly, using (4.23) and (4.8) for $\pi(t)$, we can express X^n as a semimartingale through the equation

$$X^n(t) = X^n(0) + \bar{A}^n(t) + \bar{M}^n(t) \tag{4.26}$$

where

$$\bar{A}^n(t) = \sqrt{n}\left[A^n(t) - \int_0^t F(\pi(s))ds\right] \tag{4.27}$$

and $\bar{M}^n(t) = \sqrt{n}M^n(t)$. We note that there is a natural filtration $\{\mathcal{F}_t^n\}_{0\le t\le T}$ on the probability space where the processes N_ℓ, D_i, and π^n are defined such that A^n, M^n, π^n, X^n, $\bar{M}^n$, $\bar{A}^n$ are RCLL processes adapted to the filtration and M^n, $\bar{M}^n$ are $\{\mathcal{F}_t^n\}$-local martingales.

4.3 Fluid Limit

In this section we present the proof of Theorem 10. First, in Section 4.3.1, we use the semimartingale representation from Section 4.2 to prove a key tightness property (see Proposition

5). Then, in Section 4.3.2, we prove the unique solvability of (4.7) and complete the proof of Theorem 10 by proving convergence of π^n to the unique solution of (4.7) in $\mathbb{C}([0,T]:\mathcal{S})$.

4.3.1 Tightness

In this section we prove tightness of $\{(\pi^n, M^n)\}_{n\in\mathbb{N}}$. We first recall the notion of $\mathbb{C}$-tightness.

Definition 4.1. Let $(\mathcal{Z}, d_{\mathcal{Z}})$ be a Polish space. For $z \in \mathbb{D}([0,T]:\mathcal{Z})$ let

$$j_T(z) \doteq \sup_{0\le t\le T} d_{\mathcal{Z}}(z(t), z(t-)).$$

We say a tight sequence of $\mathbb{D}([0,T]:\mathcal{Z})$-valued random variables $\{Z_n\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight if $j_T(Z_n) \Rightarrow 0$.

If Z_n, Z are $\mathbb{D}([0,T]:\mathcal{Z})$-valued random variables and $Z_n \Rightarrow Z$ then $\mathbb{P}(Z \in \mathbb{C}([0,T]:\mathcal{Z})) = 1$ if and only if $\{Z_n\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight (Ethier and Kurtz, 2009). The following proposition proves the $\mathbb{C}$-tightness of $\{\pi^n\}_{n\in\mathbb{N}}$ and convergence of M^n to the zero process.

Proposition 5. *Suppose that $\pi^n(0) \to \pi_0$, in $\mathcal{S}$, as $n \to \infty$. Then $\{(\pi^n, M^n)\}_{n\in\mathbb{N}}$ is a $\mathbb{C}$-tight sequence of $\mathbb{D}([0,T]:\mathcal{S}\times\boldsymbol{\ell}_2)$-valued random variables. Furthermore, $M^n \Rightarrow 0$ in $\mathbb{D}([0,T]:\boldsymbol{\ell}_2)$.*

Proof. We first prove the second statement by arguing that $\mathbb{E}\sup_{0\le s\le T}\|M^n(s)\|_2^2 \to 0$ as $n\to\infty$. For this, from Doob's inequality, it suffices to show $\mathbb{E}|\langle M^n\rangle(T)| \to 0$ as $n\to\infty$ where

$$\langle M^n\rangle(s) \doteq \sum_{j=0}^{\infty}\langle M_j^n\rangle(s), \qquad s\in[0,T].$$

From (4.25) and observing

$$\sum_{i=1}^{\infty}\langle e_j, (e_{i-1}-e_i)(e_{i-1}-e_i)^T e_j\rangle_2 \pi_i^n(s) = \pi_{j+1}^n(s) + \pi_j^n(s)$$

it follows that

$$\langle M_j^n\rangle(t) = \frac{\lambda}{n\binom{n}{L}}\int_0^t Z(j, n\pi^n(s))ds + \frac{k}{n}\int_0^t [\pi_{j+1}^n(s) + \pi_j^n(s)]ds. \tag{4.28}$$

where

$$Z(j, n\pi^n(s)) = \sum_{\ell \in \Sigma} \langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)}. \tag{4.29}$$

The ℓ-th term in the sum on the right side of (4.29) is the contribution from jobs that request servers with queue length configuration ℓ. A fixed $\ell \in \Sigma$ will make non-zero contribution to $\langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2$ if j or $j-1$ is one of the k-smallest coordinates in ℓ. Thus, for a fixed $\ell \in \Sigma$, the ℓ-th term in (4.29) is nonzero only if j or $j-1$ is a member of the set $(\ell_1, \ldots, \ell_k)$. The contribution from all such ℓ's in the sum (4.29) can be counted as follows. Suppose $0 \le i_1 \le k-1$ servers are selected among those with queue length less than $j-1$. This corresponds to $\binom{n \sum_{m=0}^{j-2} \pi_m^n(s)}{i_1}$ different choices of servers. In addition suppose $i_2 \le L - i_1$ and $i_3 \le L - i_1 - i_2$ servers are selected among those with queue length equal to $j-1$ and j, respectively. This corresponds to $\binom{n\pi_{j-1}^n(s)}{i_2}$ and $\binom{n\pi_j^n(s)}{i_3}$ choices, respectively. It follows that $L - i_1 - i_2 - i_3$ servers must be selected which have queue length larger than j which corresponds to $\binom{n \sum_{m=j+1}^{\infty} \pi_m^n(s)}{L - i_1 - i_2 - i_3}$ possible choices. Since jobs are only routed to the k shortest servers,

$$\langle e_j, \Delta_\ell \Delta_\ell^T e_j \rangle_2 = [i_2 \wedge (k - i_1) - i_3 \wedge (k - i_1 - i_2)_+]^2. \tag{4.30}$$

It follows that for $x \in n\mathcal{S}_n$

$$Z(j, x) = \sum_{i_1=0}^{k-1} \binom{\sum_{m=0}^{j-2} x_m}{i_1} \sum_{i_2=0}^{L-i_1} \binom{x_{j-1}}{i_2} \sum_{i_3=0}^{L-i_1-i_2} [i_2 \wedge (k - i_1) - i_3 \wedge (k - i_1 - i_2)_+]^2 \binom{x_j}{i_3} \binom{\sum_{m=j+1}^{\infty} x_m}{L - i_1 - i_2 - i_3}, \tag{4.31}$$

where, recall that we adopt the convention that for $a < b$, $\sum_{i=b}^{a} x_i = 0$.

Note that for non-negative integers $a, b,\ a \ge b$

$$\binom{a}{b} \le \frac{a^b}{b!}. \tag{4.32}$$

This fact, combined with (4.31) and recalling the fact that $\pi^n(s) \in \mathcal{S}$ for $s \in [0,T]$, gives the following bound on $Z(j, n\pi^n(s))$:

$$\begin{aligned} Z(j,n\pi^n(s)) &\leq \sum_{i_1=0}^{k-1} \frac{(n\sum_{m=0}^{j-2}\pi_m^n(s))^{i_1}}{i_1!} \sum_{i_2=0}^{L-i_1} \frac{(n\pi_{j-1}^n(s))^{i_2}}{i_2!} \\ &\qquad \times \sum_{i_3=0}^{L-i_1-i_2} k^2 1_{\{i_2 \vee i_3 > 0\}} \frac{(n\pi_j^n(s))^{i_3}}{i_3!} \frac{(n\sum_{m=j+1}^{\infty}\pi_m^n(s))^{L-i_1-i_2-i_3}}{(L-i_1-i_2-i_3)!} \\ &\leq n^L \sum_{i_1=0}^{k-1}\sum_{i_2=0}^{L-i_1}\sum_{i_3=0}^{L-i_1-i_2} k^2 1_{\{i_2\vee i_3>0\}} (\pi_{j-1}^n(s))^{i_2} (\pi_j^n(s))^{i_3} \\ &\leq c_Z n^L (\pi_{j-1}^n(s) + \pi_j^n(s)). \end{aligned} \tag{4.33}$$

for some $c_Z \in (0,\infty)$. Using (4.33) in (4.28) gives

$$\begin{aligned} \mathbb{E}|\langle M^n\rangle(t)| &\leq \mathbb{E}\left|\frac{2\lambda(n-L)!L!c_Z n^L}{n\times n!}\int_0^t \sum_{j=0}^{\infty}\pi_j^n(s)ds\right| + \mathbb{E}\left|\frac{2k}{n}\int_0^t\sum_{j=0}^{\infty}\pi_j^n(s)ds\right| \\ &\leq \left|\frac{2\lambda(n-L)!L!c_Z n^L}{n\times n!}t\right| + \left|\frac{2k}{n}t\right|. \end{aligned} \tag{4.34}$$

Thus $\mathbb{E}|\langle M^n\rangle_T| \to 0$ and consequently $\mathbb{E}\sup_{0\leq s\leq T}\|M^n(s)\|_2^2 \to 0$ as $n\to\infty$. It follows that $M^n \Rightarrow 0$ in $\mathbb{D}([0,T]:\ell_2)$ which completes the proof of (ii).

The tightness of $\{\pi^n\}_{n\in\mathbb{N}}$ in $\mathbb{D}([0,T]:\mathcal{S})$ follows as in the proof of Theorem 3.4 of (Graham, 2000). Namely, it suffices to show tightness of $\{Q_1^n\}_{n\in\mathbb{N}}$ in $\mathbb{D}([0,T]:\mathbb{N})$ (cf. (Sznitman, 1991)). However, this tightness is an immediate consequence of the fact that the jumps of Q_1^n can be embedded in a Poisson process with rate $\lambda L + k$.

Finally in order to show that $\{\pi^n\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight it suffices to show that

$$j_T(\pi^n) \doteq \sup_{0\leq t\leq T} d_0(\pi^n(t), \pi^n(t-)) \to 0 \text{ as } n\to\infty.$$

There are two types of jumps, those corresponding to incoming jobs and those corresponding to jobs being processed. When a job arrives in the system, the dispatcher assigns it to k different servers causing the queue length of each of the k chosen servers to increase by one. It follows that the jump size of such an event can be bounded by $\frac{2k}{n}$. When a job is processed, the corresponding queue length will drop by 1 and so the jump size of such an event can be bounded by $\frac{2}{n}$. Therefore $j_T(\pi^n) \leq \frac{2+2k}{n} \to 0$ which completes the proof. □

4.3.2 Convergence

In this section we provide the proof of Theorem 10. Since we have already proved tightness of $\{\pi^n\}_{n\in\mathbb{N}}$ in Section 4.3.1, all that remains is to prove uniqueness of solutions of (4.7) in an appropriate class and to characterize the limit of any weakly convergent subsequence as the unique solution to (4.7). We first present the following Lipschitz property for the map $F : \mathcal{S} \to \boldsymbol{\ell}_1$, defined in (4.4), that will give uniqueness of the solutions to (4.7). We remark that in the proof of Theorem 14 we will need a stronger Lipschitz property of F in the $\boldsymbol{\ell}_2$ norm. This Lipschitz property is not immediate on the space $\mathcal{S}$ but, as shown in Lemma 14, is satisfied on a smaller class $\mathcal{V}_M$.

Lemma 5. *The map F is a Lipschitz function from $\mathcal{S}$ to $\boldsymbol{\ell}_1$. Namely, there exists an $C_1 \in (0,\infty)$ such that for any $r, \tilde{r} \in \mathcal{S}$,*

$$\|F(r) - F(\tilde{r})\|_1 \le C_1 \|r - \tilde{r}\|_1. \tag{4.35}$$

Proof. Let $r, \tilde{r} \in \mathcal{S}$ and, for $i_1 \in \mathbb{N}_0$ and $j, i_2 \in \mathbb{N}$, define $R_{j,i_1,i_2}(r,\tilde{r})$ as

$$R_{j,i_1,i_2}(r,\tilde{r}) \doteq \left(\sum_{m=0}^{j-1} r_m\right)^{i_1} \left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2} r_j^{i_2} - \left(\sum_{m=0}^{j-1} \tilde{r}_m\right)^{i_1} \left(\sum_{m=j+1}^{\infty} \tilde{r}_m\right)^{L-i_1-i_2} \tilde{r}_j^{i_2}. \tag{4.36}$$

Note that for any $a, b, c, \tilde{a}, \tilde{b}, \tilde{c} \in \mathbb{R}_+$,

$$abc - \tilde{a}\tilde{b}\tilde{c} = ab(c-\tilde{c}) + a(b-\tilde{b})\tilde{c} + (a-\tilde{a})\tilde{b}\tilde{c}. \tag{4.37}$$

Combining (4.36), (4.37), and the fact that $r, \tilde{r} \in \mathcal{S}$, we have

$$\begin{aligned} |R_{j,i_1,i_2}(r,\tilde{r})| \le |r_j^{i_2} - \tilde{r}_j^{i_2}| + \tilde{r}_j^{i_2} \left| \left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2} - \left(\sum_{m=j+1}^{\infty} \tilde{r}_m\right)^{L-i_1-i_2} \right| \\ + \tilde{r}_j^{i_2} \left| \left(\sum_{m=0}^{j-1} r_m\right)^{i_1} - \left(\sum_{m=0}^{j-1} \tilde{r}_m\right)^{i_1} \right|. \end{aligned} \tag{4.38}$$

For any $a, b \in \mathbb{R}$ and $i \in \mathbb{N}$, $(a^i - b^i) = (a-b)\sum_{j=1}^{i} a^{i-j} b^{j-1}$. Thus, if $a, b \in [0,1]$ and $i \le L$, $|a^i - b^i| \le |a-b|L$. This inequality along with (4.38) implies there exist $\kappa_1, \kappa_1' > 0$ such that for

all $i_1, i_2 \le L,\ i_2 > 0$,

$$\begin{aligned} |R_{j,i_1,i_2}(r,\tilde{r})| &\le \kappa_1' \left(|r_j - \tilde{r}_j| + \tilde{r}_j^{i_2} \sum_{m=j+1}^{\infty} |r_m - \tilde{r}_m| + \tilde{r}_j^{i_2} \sum_{m=0}^{j-1} |r_m - \tilde{r}_m| \right) \\ &\le \kappa_1 (|r_j - \tilde{r}_j| + \tilde{r}_j \|r - \tilde{r}\|_1). \end{aligned} \tag{4.39}$$

The definition of F (see (4.4)) and the triangle inequality imply,

$$\|F(r) - F(\tilde{r})\|_1 \le \lambda L! \sum_{j=0}^{\infty} |\bar{\zeta}^{\delta}(j,r) - \bar{\zeta}^{\delta}(j,\tilde{r})| + k \sum_{j=0}^{\infty} |(r-\tilde{r})_{j+1} - (r-\tilde{r})_j|. \tag{4.40}$$

Noting that

$$\bar{\zeta}^{\delta}(j,r) - \bar{\zeta}^{\delta}(j,\tilde{r}) = [\bar{\zeta}(j-1,r) - \bar{\zeta}(j-1,\tilde{r})] - [\bar{\zeta}(j,r) - \bar{\zeta}(j,\tilde{r})],$$

it follows that

$$\sum_{j=0}^{\infty} |\bar{\zeta}^{\delta}(j,r) - \bar{\zeta}^{\delta}(j,\tilde{r})| \le 2 \sum_{j=0}^{\infty} |\bar{\zeta}(j,r) - \bar{\zeta}(j,\tilde{r})| \le \kappa_2 \sum_{j=0}^{\infty} \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} |R_{j,i_1,i_2}(r,\tilde{r})| \tag{4.41}$$

where the second inequality follows from the the definitions of $\bar{\zeta}$ and R. Combining (4.41) with (4.40) and applying (4.39) yields, for some $\kappa_3 > 0$,

$$\begin{aligned} \|F(r) - F(\tilde{r})\|_1 &\le \kappa_2 \lambda L! \sum_{j=0}^{\infty} \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} |R_{j,i_1,i_2}(r,\tilde{r})| + 2k \sum_{j=0}^{\infty} |r_j - \tilde{r}_j| \\ &\le \kappa_3 \sum_{j=0}^{\infty} [|r_j - \tilde{r}_j| + \tilde{r}_j \|r - \tilde{r}\|_1] + 2k \|r - \tilde{r}\|_1 \end{aligned}$$

and thus with $C_1 \doteq 2(\kappa_3 + k)$, (4.35) is satisfied for all $r, \tilde{r} \in \mathcal{S}$ which proves the result. □

Using the above Lipschitz property of F we can now complete the proof of Proposition 3.

Proof of Proposition 3. Existence of a $\pi \in \mathbb{C}([0,T] : \mathcal{S})$ that solves (4.7) will be shown below in the proof of Theorem 10. We now argue uniqueness. Suppose π and $\tilde{\pi}$ are two elements of $\mathbb{C}([0,T] : \mathcal{S})$ satisfying (4.7) with $\pi(0) = \tilde{\pi}(0) = \pi_0$. The Lipschitz property of F proved in

Lemma 5 implies, for all $t \in [0,T]$

$$\|\pi(t) - \tilde{\pi}(t)\|_1 = \left\| \int_0^t [F(\pi(s)) - F(\tilde{\pi}(s))] ds \right\|_1 \leq \int_0^t \|F(\pi(s)) - F(\tilde{\pi}(s))\|_1 ds$$
$$\leq C_1 \int_0^t \|\pi(s) - \tilde{\pi}(s)\|_1 ds.$$

The result follows. □

We now proceed to the proof of Theorem 10.

Proof of Theorem 10. From Proposition 5 we have that $\{\pi^n\}_{n\in\mathbb{N}}$ is a $\mathbb{C}$-tight sequence of $\mathbb{D}([0,T] : \mathcal{S})$-valued random variables.

Note from (4.23) that for all $j \in \mathbb{N}_0$,

$$\pi^n(t) = \pi^n(0) + V^n(t) + M^n(t) + \int_0^t F(\pi^n(s)) ds \tag{4.42}$$

where

$$V^n(t) \doteq A^n(t) - \int_0^t F(\pi^n(s)) ds.$$

From the definition of A^n in (4.24) we see that

$$A_j^n(t) = \int_0^t \left(\sum_{\ell\in\Sigma} \langle \Delta_\ell, e_j \rangle_2 \frac{\lambda}{\binom{n}{L}} \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} + k[\pi_{j+1}^n(s) - \pi_j^n(s)] \right) ds. \tag{4.43}$$

By a similar argument (see comments given below (4.45)) used to obtain the representation in (4.31),

$$\sum_{\ell\in\Sigma} \langle \Delta_\ell, e_j \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)} = [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))] \tag{4.44}$$

where for $x \in n\mathcal{S}_n$

$$\zeta(j,x) \doteq \sum_{i_1=0}^{k-1} \binom{\sum_{m=0}^{j-1} x_m}{i_1} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k-i_1)] \binom{x_j}{i_2} \binom{\sum_{m=j+1}^{\infty} x_m}{L-i_1-i_2}. \tag{4.45}$$

One can interpret $\zeta(j,x)$ as the rate at which jobs are being routed into queues of length j when the system is in state x. Recall that any incoming job corresponds to the selection of L queues. The term on the right side of (4.45) then sums over all possible queue length configurations

of this selection. In particular, i_1 represents the number of queues with lengths less than j, i_2 corresponds to the queues of length equal to j, and $L - i_1 - i_2$ are the queues of length greater than j. Since we are routing jobs to the k shortest queues the rate must be multiplied by the factor $[i_2 \wedge (k - i_1)]$ rather than i_2. From our convention that $x_{-1} = 0$, we see that $\zeta(-1, x) = 0$. In addition, recalling the conventions that for $a < b$, $\sum_{i=b}^{a} x_i = 0$ and that $\binom{0}{0} = 1$ we see $\zeta(0, x)$ is well defined. Combining (4.43), (4.44), and (4.45) gives the following representation for A_j^n

$$A_j^n(t) = \frac{\lambda}{\binom{n}{L}} \int_0^t [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))]ds + k \int_0^t [\pi_{j+1}^n(s) - \pi_j^n(s)]ds. \tag{4.46}$$

For each fixed $j, i_1 \in \mathbb{N}_0$ and $i_2 \in \mathbb{N}$ with $i_1, i_2 \leq L$ we have

$$\begin{aligned}
&\binom{n\sum_{m=0}^{j-1} \pi_m^n(s)}{i_1} [i_2 \wedge (k - i_1)] \binom{n\pi_j^n(s)}{i_2} \binom{n\sum_{m=j+1}^{\infty} \pi_m^n(s)}{L - i_1 - i_2} \\
&\quad = n^L \frac{\left(\sum_{m=0}^{j-1} \pi_m^n(s)\right)^{i_1}}{i_1!} [i_2 \wedge (k - i_1)] \frac{(\pi_j^n(s))^{i_2}}{i_2!} \frac{\left(\sum_{m=j+1}^{\infty} \pi_m^n(s)\right)^{L - i_1 - i_2}}{(L - i_1 - i_2)!} + \hat{R}_n(j, i_1, i_2, s)
\end{aligned} \tag{4.47}$$

where

$$\sup_{i_1, i_2 \leq L} |\hat{R}_n(j, i_1, i_2, s)| \leq \kappa_1 n^{L-1} \pi_j^n(s)$$

and thus, from the definition of ζ and $\bar{\zeta}$ in (4.45) and (4.5),

$$\left| \zeta(j, n\pi^n(s)) - \frac{n!}{(n-L)!} \bar{\zeta}(j, \pi^n(s)) \right| \leq \kappa_2 n^{L-1} \pi_j^n(s) \ \forall \ s \in [0, T]. \tag{4.48}$$

Furthermore, using the definition of A^n in (4.46) and F in (4.4), (4.48) implies

$$\sup_{0 \leq t \leq T} \|V^n(t)\|_2 = \sup_{0 \leq t \leq T} \left\| A^n(t) - \int_0^t F(\pi^n(s))ds \right\|_2 \leq \frac{\kappa_3}{n}. \tag{4.49}$$

Also from Proposition 5, $M^n \Rightarrow 0$ in $\mathbb{D}([0, T] : \boldsymbol{\ell}_2)$. Combining these observations with the tightness of π^n, we have subsequential convergence of (π^n, M^n, V^n) to $(\pi, 0, 0)$, in distribution, in $\mathbb{D}([0, T] : \mathcal{S} \times \boldsymbol{\ell}_2 \times \boldsymbol{\ell}_2)$ for some $\mathbb{C}([0, T] : \mathcal{S})$-valued π. By appealing to the Skorohod representation theorem we can assume that this convergence holds a.s. Noting that $r \mapsto F_j(r)$ is a continuous map from $\mathcal{S}$ to $\mathbb{R}$ for each $j \in \mathbb{N}_0$ we have that $F_j(\pi^n(s)) \to F_j(\pi(s))$ as $n \to \infty$ for all $j \in \mathbb{N}_0$ and $s \in [0, T]$. Thus, upon sending $n \to \infty$ in (4.42), (4.9) and the dominated

convergence theorem imply that almost surely,

$$\pi_j(t) = (\pi_0)_j + \int_0^t F_j(\pi(s))ds, \text{ for all } t \in [0,T],\ j \in \mathbb{N}_0.$$

This shows that π satisfies (4.7). The result now follows from the uniqueness property shown in Proposition 3. □

4.3.3 Proof of Lemma 4

The result will follow upon verifying,

$$L!\bar{\zeta}(m,r) = \sum_{\ell=1}^{k} \sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} \sum_{i_2=\ell-i_1}^{L-i_1} \binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \tag{4.50}$$

and, for $\ell = \{1,\dots,k\}$,

$$\sum_{j=m}^{\infty} \sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_j)^{i_1} \sum_{i_2=\ell-i_1}^{L-i_1} \binom{L-i_1}{i_2} (r_j)^{i_2} u_{j+1}^{L-i_1-i_2} = \sum_{j=L-\ell+1}^{L} \binom{L}{j} u_m^j (1-u_m)^{L-j}. \tag{4.51}$$

These equations can be interpreted as follows. Suppose the occupancy measure is in state r. Roughly speaking, a typical term in the outside summation on the RHS of (4.50), denoted as $p_m(\ell)$, corresponds to the probability that the ℓ-th largest out of L randomly selected queues is of length m. Then (4.50) states that the rate of jobs being routed into queues of length m is equal to the sum $\sum_{\ell=1}^{k} p_m(\ell)$. Recall that a file request will correspond to a queue length configuration $a_1 \le a_2 \le \cdots \le a_L$, where a_j corresponds to the length of the j-th largest queue. A typical term in the outside summation on the LHS of (4.51), denoted $\tilde{p}_\ell(j)$, corresponds to the probability that $a_\ell = j$. Terms in the summation on the RHS of (4.51), denoted $q_m(j)$, correspond to the probability that $a_{j-1} < m \le a_j$. The expression (4.51) then states that $\sum_{j=m}^{\infty} \tilde{p}_\ell(j) = \sum_{j=L-\ell+1}^{L} q_m(j)$. Once these equalities are established the remainder of the argument follows as in Appendix B of (Li et al., 2016) which argues that

$\sum_{\ell=1}^{k}\sum_{j=L-\ell+1}^{L} q_m(j) = f(u_m)$. Combining this fact with (4.50) and (4.51) then gives,

$$\begin{aligned} L!\bar{\zeta}(m,r) &= \sum_{\ell=1}^{k} p_m(\ell) = \sum_{\ell=1}^{k}\left[\sum_{j=m}^{\infty}\tilde{p}_\ell(j) - \sum_{j=m+1}^{\infty}\tilde{p}_\ell(j)\right] = \sum_{\ell=1}^{k}\left[\sum_{j=L-\ell+1}^{L} q_m(j) - \sum_{j=L-\ell+1}^{L} q_{m+1}(j)\right] \\ &= f(u_m) - f(u_{m+1}) \end{aligned}$$

which proves the result.

We now prove the two equalities. First consider (4.50). By rearranging and collecting combinatorial terms we can write

$$L!\bar{\zeta}(m,r) = \sum_{i_1=0}^{k-1}\sum_{i_2=1}^{L-i_1}[i_2 \wedge (k-i_1)]\binom{L}{i_1}\binom{L-i_1}{i_2}(1-u_m)^{i_1} r_m^{i_2} u_{m+1}^{L-i_1-i_2}. \tag{4.52}$$

Note that the RHS in (4.52) can be written as

$$\begin{aligned} &\sum_{i_1=0}^{k-1}\sum_{i_2=1}^{L-i_1}\sum_{\ell=1}^{[i_2\wedge(k-i_1)]}\binom{L}{i_1}\binom{L-i_1}{i_2}(1-u_m)^{i_1} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \\ &\quad = \sum_{i_1=0}^{k-1}\binom{L}{i_1}(1-u_m)^{i_1}\sum_{i_2=1}^{L-i_1}\sum_{\ell=i_1+1}^{(i_1+i_2)\wedge k}\binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2}. \end{aligned} \tag{4.53}$$

We then exchange the order of summations as follows

$$\begin{aligned} &\sum_{i_1=0}^{k-1}\binom{L}{i_1}(1-u_m)^{i_1}\sum_{i_2=1}^{L-i_1}\sum_{\ell=i_1+1}^{(i_1+i_2)\wedge k}\binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \\ &\quad = \sum_{i_1=0}^{k-1}\sum_{\ell=i_1+1}^{k}\binom{L}{i_1}(1-u_m)^{i_1}\sum_{i_2=\ell-i_1}^{L-i_1}\binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \\ &\quad = \sum_{\ell=1}^{k}\sum_{i_1=0}^{\ell-1}\binom{L}{i_1}(1-u_m)^{i_1}\sum_{i_2=\ell-i_1}^{L-i_1}\binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2}. \end{aligned} \tag{4.54}$$

Combining (4.52), (4.53), and (4.54) gives (4.50).

We now prove (4.51). Fix $\ell \in \{1,\ldots,k\}$ and note that

$$\sum_{j=L-\ell+1}^{L}\binom{L}{j} u_m^j (1-u_m)^{L-j} = \sum_{i_1=0}^{\ell-1}\binom{L}{i_1}(1-u_m)^{i_1} u_m^{L-i_1}. \tag{4.55}$$

Then, applying the binomial theorem to $u_m^{L-i_1} = (r_m + u_{m+1})^{L-i_1}$, (4.55) becomes

$$\sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} u_m^{L-i_1} = \sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} \sum_{i_2=0}^{L-i_1} \binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \tag{4.56}$$

which, by breaking up the summation indexed by i_2, can be rewritten as

$$\begin{aligned} &\sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} \sum_{i_2=\ell-i_1}^{L-i_1} \binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \\ &\quad + \sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} \sum_{i_2=0}^{\ell-i_1-1} \binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2}. \end{aligned} \tag{4.57}$$

Now consider the second term in (4.57). By relabeling the indices we get

$$\begin{aligned} &\sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_m)^{i_1} \sum_{i_2=0}^{\ell-i_1-1} \binom{L-i_1}{i_2} r_m^{i_2} u_{m+1}^{L-i_1-i_2} \\ &\quad = \sum_{i_1=L-\ell+1}^{L} \binom{L}{i_1} u_{m+1}^{i_1} \sum_{i_2=0}^{L-i_1} \binom{L-i_1}{i_2} (1-u_m)^{i_2} r_m^{L-i_1-i_2} \\ &\quad = \sum_{i_1=L-\ell+1}^{L} \binom{L}{i_1} u_{m+1}^{i_1} (1-u_{m+1})^{L-i_1} \end{aligned} \tag{4.58}$$

where the second equality follows from the binomial theorem. It then follows from (4.55)-(4.58) that for any $m' > m$

$$\begin{aligned} &\sum_{j=L-\ell+1}^{L} \binom{L}{j} u_m^j (1-u_m)^{L-j} \\ &\quad = \sum_{j=m}^{m'} \sum_{i_1=0}^{\ell-1} \binom{L}{i_1} (1-u_j)^{i_1} \sum_{i_2=\ell-i_1}^{L-i_1} \binom{L-i_1}{i_2} r_j^{i_2} u_{j+1}^{L-i_1-i_2} + \sum_{j=L-\ell+1}^{L} \binom{L}{j} u_{m'}^j (1-u_{m'})^{L-j}. \end{aligned}$$

The result then follows upon sending $m' \to \infty$. □

4.3.4 Proof of Theorem 11

It is proved in Lemma 5 of (Li et al., 2016) that $f(x) \le k x^{L/k}$ and thus (i) is immediate from (4.12).

We now verify (ii). From (4.12) it suffices to show that $f(x) \geq kx^{L-k+1}$ for $x \in [0,1]$. Since both sides of the inequality evaluate to zero at $x = 0$, it is equivalent to show

$$h(x) \doteq \frac{1}{k} \frac{f(x)}{x^{L-k+1}} \geq 1 \text{ for } x \in (0,1]. \tag{4.59}$$

Note that $f(1) = k$ (cf. Lemma 2 of (Li et al., 2016)), and thus $h(1) = 1$. It follows that $h'(x) \leq 0,\ x \in (0,1]$ is sufficient for verifying (4.59). Taking the derivative of h gives

$$h'(x) = \frac{1}{k} \frac{xf'(x) - (L-k+1)f(x)}{x^{L-k+2}}. \tag{4.60}$$

Denoting $xf'(x) - (L-k+1)f(x)$ as $w(x)$, we note that in order to show $h'(x) \leq 0$ one must only verify $w(x) \leq 0$. One can verify (cf. (68) and (69) of (Li et al., 2016)) that w can be expressed as follows

$$w(x) \doteq \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} (-1)^\ell \frac{1}{L-k+\ell} \frac{\ell}{L-k+\ell+1} x^{L-k+\ell+1}.$$

Therefore

$$w''(x) = x^{L-k-1} \sum_{\ell=0}^{k-1} \binom{k-1}{\ell} \ell (-x)^\ell = -(k-1)x^{L-k}(1-x)^{k-2}$$

and so $w''(x) \leq 0$ for all $x \in [0,1]$. Noting that $w'(0) = 0$, it follows that $w'(x) \leq 0$ for all $x \in (0,1]$ and since $w(0) = 0$, $w(x) \leq 0$ for $x \in (0,1]$. This verifies (4.59). □

4.3.5 Proof of Theorem 12

In this section we present the proof of Theorem 12. Namely, for every $g \in \mathcal{U}$, the solution u of (4.11) satisfies $u(t) \in \mathcal{U}$ for all $t \geq 0$ (Lemma 10), and there is a unique fixed point to (4.11) in $\mathcal{U}$ defined by (4.12) which is asymptotically stable. The argument follows along the lines of the proof of Theorem 1 of (Vvedenskaya et al., 1996) (cf. Lemmas 1-7 therein). The key difference is that the the term $\lambda[f(u_{i-1}) - f(u_i)]$ appears in the differential equation instead of $\lambda[u_{i-1}^2 - u_i^2]$. As we will see, this difference can be handled using the properties of f shown

in Lemma 2 of (Li et al., 2016). Specifically, we will use the facts that $f(0) = 0$, $f(1) = k$, f is strictly increasing, convex, and differentiable with derivative bounded by L.

We first consider a truncated version of (4.11). Fix $K \in \mathbb{N}$, $c \geq 0$, and consider the following boundary value problem

$$\begin{cases} \dot{s}_j(t) &= \lambda[f(s_{j-1}(t)) - f(s_j(t))] - k[s_j(t) - s_{j+1}(t)], \quad j = 1, \dots, K \\ s_0(t) &= 1 \\ s_j(0) &= g_j, \quad j = 1, \dots, K \end{cases} \tag{4.61}$$

with

$$s_{K+1}(t) = c. \tag{4.62}$$

The following two lemmas giving monotonicity and uniqueness properties for the truncated system will be used to extend the same properties to the full system in Lemma 11.

Lemma 6. *Suppose s is a solution to* (4.61)-(4.62) *with initial conditions satisfying*

$$1 = g_0 \geq g_1 \geq \ldots \geq g_K \geq g_{K+1} = c. \tag{4.63}$$

Then

$$1 = s_0(t) \geq s_1(t) \geq \ldots \geq s_K(t) \geq s_{K+1}(t) = c \tag{4.64}$$

for all $t \geq 0$.

Proof. Since solutions to (4.61)-(4.62) depend continuously on the initial conditions we can take the inequalities in (4.63) to be strict, without loss of generality. Let t_0 be the first time that an equality appears in (4.64). Since $s_0 > s_{K+1}$, then there exists an $i \in \{1, \dots, K\}$ such that either $s_{i-1}(t_0) > s_i(t_0) = s_{i+1}(t_0)$ or $s_{i-1}(t_0) = s_i(t_0) > s_{i+1}(t_0)$. In the former case, since f is strictly increasing, $\dot{s}_i(t_0) = \lambda[f(s_{i-1}(t_0)) - f(s_i(t_0))] > 0$ and $\dot{s}_{i+1}(t_0) = k[s_{i+2}(t_0) - s_{i+1}(t_0)] \leq 0$ if $i < K$ and $s_{i+1}(t_0) = 0$ if $i = K$, both of which contradict the assumption that $s_i(t) > s_{i+1}(t)$ for $t < t_0$. The latter case follows from a similar argument. □

Lemma 7. *Let* $\{s_i^{(1)}\}_{i=0}^K$ *and* $\{s_i^{(2)}\}_{i=0}^K$ *solve* (4.61) *and be such that* $s_i^{(1)}(0) \geq s_i^{(2)}(0)$ *for all* $i = 1, 2, \ldots, K$. *If, in addition,* $s_{K+1}^{(1)}(t) \geq s_{K+1}^{(2)}(t)$ *for all* $t \geq 0$ *then* $s_i^{(1)}(t) \geq s_k^{(2)}(t)$ *for all* $i = 1, 2, \ldots, K, K+1$ *and all* $t \geq 0$.

Proof. Again, assume without loss of generality that the inequalities are strict. I.e. $s_i^{(1)}(0) > s_i^{(2)}(0)$ for all $i = 1, 2, \ldots, K$ and $s_{K+1}^{(1)}(t) > s_{K+1}^{(2)}(t)$, for all $t \geq 0$. Suppose the first time equality appears is at time t_0. If $j \in \{1, \ldots, K\}$ is the largest index such that $s_j^{(1)}(t_0) = s_j^{(2)}(t_0)$ then, since f is strictly increasing,

$$\begin{aligned}\dot{s}_j^{(1)}(t_0) - \dot{s}_j^{(2)}(t_0) &= \lambda[f(s_{j-1}^{(1)}(t_0)) - f(s_{j-1}^{(2)}(t_0))] + k[s_{j+1}^{(1)}(t_0) - s_{j+1}^{(2)}(t_0)] \\ &\geq k[s_{j+1}^{(1)}(t_0) - s_{j+1}^{(2)}(t_0)] > 0\end{aligned}$$

which contradicts the assumption $s_j^{(1)}(t) > s_j^{(2)}(t)$ for $t < t_0$. □

Note that Lemma 7, in particular, shows that there is a unique solution to (4.61)-(4.62). We now consider the full system (4.11). In the following lemma we show that the full system can be constructed as the limit of the sequence of truncated systems defined through (4.61).

Lemma 8. *Let* $g \in \bar{\mathcal{U}}$.

i) There exists a unique solution to (4.11) *in* $\bar{\mathcal{U}}$.

ii) This solution can be obtained as the limit as $K \to \infty$ *of solutions to the truncated systems* (4.61)-(4.62) *associated with* $c = 0$.

Proof. Part (i) follows immediately from Lemma 4 and Proposition 2.1 of (Budhiraja and Friedlander, 2017). Let $s^K(t)$, $K = 1, 2, \ldots$ denote solutions to (4.61) with $s_{K+1}^K(t) = 0$. It follows from Lemma 6 that $s_{K+1}^{K+1}(t) \geq s_{K+1}^K(t) = 0$ and from Lemma 7 that for fixed t and $i \leq K$, $s_i^{K+1}(t) \geq s_i^K(t)$. It follows that $\lim_{K\to\infty} s_i^K(t) = s_i(t)$ exists, $s(t) \in \bar{\mathcal{U}}$, and s_i satisfies (4.11) which proves (ii). □

Lemma 9. *Let* u *be a solution to* (4.11) *taking values in* $\bar{\mathcal{U}}$. *Then the following estimate holds for all* t,

$$u_j(t) \leq \sum_{i=0}^{j} \frac{u_i(0)(\lambda k t)^{j-i}}{(j-i)!}, \qquad j \in \mathbb{N}_0. \tag{4.65}$$

Proof. The lemma follows from using an inductive argument. Note that the inequality is immediate for $j = 0$. Suppose now that (4.65) holds for $j-1$, for some $j \geq 1$. Then, since $f(0) = 0$, $f(1) = k$, and f is convex on $[0,1]$, it follows from (4.11) that

$$\dot{u}_j(t) \leq \lambda f(u_{j-1}(t)) \leq \lambda k u_{j-1}(t).$$

Since (4.65) holds for $j-1$ by our inductive hypothesis we have, by integrating over t on both sides of the above inequality, that (4.65) also holds for j. The result follows. □

Lemma 10. *Let u be a solution to (4.11) taking values in $\bar{\mathcal{U}}$. If $u(0) \in \mathcal{U}$, then $u(t) \in \mathcal{U}$ for all $t \geq 0$. Furthermore, $v_1(u(t)) \leq \exp(\lambda k t)[1 + v_1(u(0))]$.*

Proof. This follows immediately from the estimate (4.65). □

The following monotonicity property of the full system (4.11) is an immediate consequence of Lemma 7 and part (ii) of Lemma 8.

Lemma 11. *Let $u^{(1)}$ and $u^{(2)}$ be a solutions to (4.11) in $\bar{\mathcal{U}}$ with $u_j^{(1)}(0) \geq u_j^{(2)}(0)$ for all $j \in \mathbb{N}_0$. Then $u_j^{(1)}(t) \geq u_j^{(2)}(t)$ for all $j = \mathbb{N}_0$ and all $t \geq 0$.*

With the above lemmas we can now complete the proof of Theorem 12.

Proof of Theorem 12. Part (i) of the theorem was shown in Lemma 10.

Now consider part (ii). Suppose $g_i \leq \bar{u}_i,\ i \in \mathbb{N}_0$. Then from Lemma 11, it follows that $v_1(u(t)) \leq \sum_{i=1}^{\infty} \bar{u}_i < \infty$. If instead, $g_i \geq \bar{u}_i,\ i \in \mathbb{N}_0$ then from (4.12) and noting that $\bar{u}_0 = 1$ and $f(1) = k$, we have that $\bar{u}_1 = \lambda f^{(L,k)}(1)/k = \lambda$. Thus, from Lemma 11 once more, $u_1(t) \geq \lambda$ for all $t \geq 0$, from which it follows that

$$\dot{v}_1(u(t)) = \lambda f(1) - k u_1(t) = k(\lambda - u_1(t)) \leq 0.$$

Therefore, in both cases $v_1(u(t))$ is uniformly bounded in t. Assume for now that we are in one of these two cases.

We now prove that

$$\int_0^{\infty} |u_k(t) - \bar{u}_k| dt < \infty \tag{4.66}$$

for each k. Noting that f has derivative bounded by L (cf. Lemma 2 of (Li et al., 2016)) it will then follow that, for each of these two cases we have the desired convergence

$$\lim_{t\to\infty} |u_k(t) - \bar{u}_k| = 0, \text{ for all } k \in \mathbb{N}_0. \tag{4.67}$$

From this, convergence for an arbitrary initial condition will follow on noting that from Lemma 11, $u^-(t) \le u(t) \le u^+(t)$ where u^- and u^+ are the solutions to (4.11) with $u_k^-(0) = g_k \wedge \bar{u}_k$ and $u_k^+(0) = g_k \vee \bar{u}_k$. Finally, we prove (4.66) using an inductive argument. It is clear that (4.66) holds for $k = 0$. Now suppose (4.66) is true for $k-1$, for some $k \ge 1$. Then

$$\dot{v}_k(u(t)) = \lambda f(u_{k-1}(t)) - k u_k(t) = \lambda[f(u_{k-1}(t)) - f(\bar{u}_{k-1})] - k[u_k(t) - \bar{u}_k]$$

and thus

$$v_k(u(t)) - v_k(g) = \int_0^t (\lambda[f(u_{k-1}(s)) - f(\bar{u}_{k-1})] - k[u_k(s) - \bar{u}_k])\, ds.$$

Note that since $v_k(u(t)) \le v_1(u(t))$ we must have that $v_k(u(t)) - v_k(g)$ is uniformly bounded in t. From the inductive assumption and appealing again to the boundedness of the first derivative of f it follows that $\sup_{t\in[0,\infty)} \int_0^t \lambda[f(u_{k-1}(s)) - f(\bar{u}_{k-1})]ds < \infty$. Therefore (4.66) is satisfied for k which completes the proof. □

4.3.6 Proof of Theorem 13

Note that $\mathcal{L}_n$ is a probability measure on the set $\bar{\mathcal{U}}$ which is a compact set in the product topology. Thus, $\{\mathcal{L}_n\}_{n\in\mathbb{N}}$ is a tight sequence in $\mathcal{P}(\bar{\mathcal{U}})$. Let $\{\mathcal{L}_{n_k}\}_{k\in\mathbb{N}}$ be a weakly convergent subsequence with limit point $\mathcal{L}$. Suppose $u^{n_k}(0)$ is distributed according to $\mathcal{L}_{n_k}$ (we write $u^{n_k}(0) \sim \mathcal{L}_{n_k}$). Then $u^{n_k}(t) \sim \mathcal{L}_{n_k}$ for all $t \ge 0$. By a minor modification of the proof of Theorem 2.2 of (Budhiraja and Friedlander, 2017) it follows now that $u_{n_k} \Rightarrow u$ in $\mathbb{D}([0,T],\mathcal{U})$ where u solves the ODE (4.11) a.s. Theorem 2.2 of (Budhiraja and Friedlander, 2017) proves such a result for the case where the initial occupancy measure $u(0)$ is deterministic. However, the extension to the case where the initial conditions are stochastic is straight forward. Since at

any time t, $u_{n_k}(t) \sim \mathcal{L}_{n_k}$, it follows that $u(t) \sim \mathcal{L}$. From the fact that $\bar{u}$ is the unique fixed point of (4.11) it follows now that $\mathcal{L} = \delta_{\bar{u}}$ and thus $\delta_{\bar{u}}$ must be the limit point of every convergent subsequence. This completes the proof of the first statement in Theorem 13. The second statement is immediate on noting that for all $k \in \mathbb{N}_0$, $\mathbb{E}u_k^n(t) \to \int_{\mathcal{U}} u_k d\mathcal{L}^n(u)$ as $t \to \infty$. □

4.4 Diffusion Approximation

In this section we prove Theorem 14. Section 4.4.1 presents some moment estimates on π^n which will be used in the proof of Theorem 14. Section 4.4.2 then proves tightness of the sequence of centered and scaled state processes $\{X^n\}_{n\in\mathbb{N}}$. Section 4.4.3 completes the proof of Theorem 14 by proving unique solvability of the SDE (4.14) (Theorem 4) and characterizing limit points of X^n as this unique solution.

4.4.1 Moment Bounds

The following elementary lemma will be useful in the proof of Lemma 13.

Lemma 12. *For all $t \geq 0$, $k \in \mathbb{N}$, and $n \in \mathbb{N}$, $\lim_{m\to\infty} \mathbb{E}m^k \sup_{0\leq s\leq t} \pi_m^n(s) = 0$.*

Proof. Fix $n \in \mathbb{N}$. Note that file requests arrive at rate $n\lambda$. Let N be a Poisson process representing the total flow of such file requests. Also let $m^* = \sup\{m : \pi_m^n(0) > 0\}$ be the length of the largest queue at time 0. Note that since the system consists of n queues, m^* must be finite for any fixed n. Then for $m > m^*$,

$$\begin{aligned}\mathbb{E}m^k \sup_{0\leq s\leq t} \pi_m^n(s) &= \mathbb{E} \sup_{0\leq s\leq t} 1_{\{N(t)\geq m-m^*\}} m^k \pi_m^n(s) + \mathbb{E} \sup_{0\leq s\leq t} 1_{\{N(t)<m-m^*\}} m^k \pi_m^n(s) \\ &\leq m^k \mathbb{P}(N(t) \geq m - m^*).\end{aligned}$$

Thus, from Markov's inequality, for $m > m^*$

$$\mathbb{E}m^k \sup_{0\leq s\leq t} \pi_m^n(s) \leq m^k e^{-(m-m^*)} e^{n\lambda t(e-1)}.$$

The result follows. □

In the next lemma we will we establish two key moment bounds that will be needed in the tightness proof (see proof of Proposition 6).

Lemma 13. *Suppose* $\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty} j^2\pi_j^n(0) \doteq c_{\pi(0)} < \infty$. *Then*

$$\sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\le t\le T}\left(\sum_{j=0}^{\infty} j\pi_j^n(t)\right)^2 < \infty \tag{4.68}$$

and

$$\sup_{n\in\mathbb{N}} \mathbb{E}\int_0^T \sum_{j=0}^{\infty} j^2\pi_j^n(t)dt < \infty. \tag{4.69}$$

Proof. Since $\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t)$, we can write for fixed $K \in \mathbb{N}$

$$\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^{K} j\pi_j^n(t)\right|^2 \le 3\left|\sum_{j=0}^{K} j\pi_j^n(0)\right|^2 + 3\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^{K} jA_j^n(t)\right|^2 + 3\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^{K} jM_j^n(t)\right|^2. \tag{4.70}$$

Using (4.44), for $K \in \mathbb{N}$, we can write

$$\begin{aligned}\sum_{j=0}^{K} j\sum_{\ell\in\Sigma}\langle \Delta_\ell, e_j\rangle_2 \prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)} &= \sum_{j=1}^{K} j\left[\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))\right] \\ &= \sum_{j=0}^{K-1}\zeta(j, n\pi^n(s)) - K\zeta(K, n\pi^n(s))\end{aligned} \tag{4.71}$$

and

$$k\sum_{j=0}^{K} j[\pi_{j+1}^n(s) - \pi_j^n(s)] = -k\left(\sum_{j=1}^{K}\pi_j^n(s) - K\pi_{K+1}^n(s)\right). \tag{4.72}$$

Using similar bounds as in (4.33), for some $c_\zeta \in (0,\infty)$

$$\zeta(j, n\pi^n(s)) \le c_\zeta n^L \pi_j^n(s). \tag{4.73}$$

The above bound implies that for some $\kappa_1 \in (0,\infty)$, for all $n, K \in \mathbb{N}$

$$\mathbb{E}\sup_{0\le t\le T}\left[\frac{\lambda}{\binom{n}{L}}\int_0^t\sum_{j=1}^{K}\zeta(j-1, n\pi^n(s)) + k\int_0^t\sum_{j=0}^{K}\pi_j^n(s)ds\right]^2 \le \mathbb{E}\left[\left(c_\zeta n^L\frac{\lambda}{\binom{n}{L}} + k\right)T\right]^2 \le \kappa_1.$$

Combined with (4.46), (4.71), and (4.72), the above estimate gives, for all $n, K \in \mathbb{N}$,

$$\mathbb{E} \sup_{0\le t\le T} \left| \sum_{j=0}^{K} j A_j^n(t) \right|^2 \le \kappa_2 \left(1 + K\mathbb{E}\left[\sup_{0\le t\le T} (\pi_K^n(t) + \pi_{K+1}^n(t)) \right] \right). \tag{4.74}$$

We now consider $\mathbb{E}\sup_{0\le t\le T} |\sum_{j=0}^{K} j M_j^n(t)|^2$. Since $\sum_{j=0}^{K} j M_j^n(t)$ is a martingale, Doob's inequality implies that

$$\mathbb{E} \sup_{0\le t\le T} \left| \sum_{j=0}^{K} j M_j^n(t) \right|^2 \le 4\mathbb{E} \left\langle \sum_{j=0}^{K} j M_j^n \right\rangle (T) = 4\mathbb{E} \sum_{j_1=0}^{K} \sum_{j_2=0}^{K} j_1 j_2 \langle M_{j_1}^n, M_{j_2}^n \rangle (T). \tag{4.75}$$

The diagonal terms ($j_1 = j_2$) in the above sum are given by (4.28). We now consider the off-diagonal terms. Fix $0 \le j_1 < j_2 \le K$ and note that in order to compute $\langle M_{j_1}^n, M_{j_2}^n \rangle (T)$ we must expand

$$Z(j_1, j_2, n\pi^n(s)) \doteq \sum_{\ell \in \Sigma} \langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2} \rangle_2 \prod_{i=0}^{\infty} \binom{n\pi_i^n(s)}{\rho_i(\ell)}. \tag{4.76}$$

Similar to (4.29), the ℓ-th term in (4.76) is the contribution from jobs that request servers with queue length configuration ℓ. A fixed $\ell \in \Sigma$ will make non-zero contribution to $\langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2} \rangle_2$ if (j_1 or $j_1 - 1$) *and* (j_2 or $j_2 - 1$) are among the k-smallest coordinates in ℓ. That is, for a fixed $\ell \in \Sigma$, the ℓ-th term is nonzero only if (j_1 or $j_1 - 1$) is a member of the set $(\ell_1, \ldots, \ell_k)$ and (j_2 or $j_2 - 1$) is also a member. The contribution from all such ℓ's in the sum (4.76) can be counted in a method analogous to the one used to obtain (4.31). Namely, we count the number of choices of servers with queue length less than $j_1 - 1$, equal to $j_1 - 1$, equal to j_1, between j_1 and $j_2 - 1$, equal to $j_2 - 1$, equal to j_2, and larger than j_2. One must be careful in the cases $j_2 - 1 = j_1$ and $j_2 - 1 = j_1 + 1$. In both cases there are no servers with length between j_1 and $j_2 - 1$. In the first case above ($j_2 - 1 = j_1$), we must also be careful not to double count. To ensure this we include an indicator function $1_{\{j_2 > j_1 + 1\}}$ in the upper index of the binomial coefficient corresponding to the selection of servers with queue length equal to $j_2 - 1$. Combining these observations we see

that for $x \in n\mathcal{S}_n$,

$$\begin{aligned} Z(j_1,j_2,x) &= \sum_{\ell\in\Sigma}\langle e_{j_1}, \Delta_\ell \Delta_\ell^T e_{j_2}\rangle_2 \prod_{i=0}^{\infty}\binom{x_i}{\rho_i(\ell)} \\ &= \sum_{i_1=0}^{k-2}\binom{\sum_{m=0}^{j_1-2}x_m}{i_1}\sum_{i_2=0}^{k-i_1-1}\binom{x_{j_1-1}}{i_2}\sum_{i_3=0}^{k-i_1-i_2-1}[i_2-i_3]\binom{x_{j_1}}{i_3} \\ &\quad\times\sum_{i_4=0}^{k-i_1-i_2-i_3-1}\binom{\sum_{m=j_1+1}^{j_2-2}x_m}{i_4}\sum_{i_5=0}^{L-\sum_{n=1}^{4}i_n}\binom{x_{j_2-1}1_{\{j_2>j_1+1\}}}{i_5} \\ &\quad\times\sum_{i_6=0}^{L-\sum_{n=1}^{5}i_n}\left[(1_{\{j_2=j_1+1\}}(i_3-i_5)+i_5)\wedge\left(k-\sum_{n=1}^{4}i_n\right)_+ - i_6\wedge\left(k-\sum_{n=1}^{5}i_n\right)_+\right] \\ &\quad\times\binom{x_{j_2}}{i_6}\binom{\sum_{m=j_2+1}^{\infty}x_m}{L-\sum_{n=1}^{6}i_n}. \end{aligned} \tag{4.77}$$

For $j_1 > j_2$ we define $Z(j_1,j_2,x) \doteq Z(j_2,j_1,x)$. The contribution to $\langle M_{j_1}^n, M_{j_2}^n\rangle(T)$, for $j_1 \neq j_2$, from completed jobs is given by the following term:

$$\sum_{i=1}^{\infty}\langle e_{j_1}, (e_{i-1}-e_i)(e_{i-1}-e_i)^T e_{j_2}\rangle_2 \pi_i^n(s) = -1_{\{j_1=j_2-1\}}\pi_{j_2}^n(s) - 1_{\{j_1-1=j_2\}}\pi_{j_1}^n(s). \tag{4.78}$$

This follows on noting that if a job is completed from a queue of length j then its queue length become $j-1$. This implies that the contribution is zero unless $j_1 = j_2 - 1$ or $j_1 - 1 = j_2$ which results in the above expression. Combining (4.77) and (4.78) gives, for $j_1, j_2 \in \mathbb{N}_0$,

$$\begin{aligned} \langle M_{j_1}^n, M_{j_2}^n\rangle(T) &= \frac{\lambda}{n\binom{n}{L}}\int_0^T Z(j_1,j_2,n\pi^n(s))ds \\ &\quad + \frac{k}{n}\int_0^T [1_{\{j_1=j_2\}}[\pi_{j_1}^n(s)+\pi_{j_1+1}^n] - 1_{\{j_1=j_2-1\}}\pi_{j_2}^n(s) - 1_{\{j_1-1=j_2\}}\pi_{j_1}^n(s)]ds, \end{aligned} \tag{4.79}$$

where, by convention, $Z(j,j,x) \doteq Z(j,x)$. Referring to the definition of Z in (4.77), note that for $j_2 > j_1 + 1$, $Z(j_1,j_2,x) = 0$ unless (i_2 or i_3) are greater than zero and (i_5 or i_6) are greater than zero. In the case that $j_2 = j_1 + 1$, $Z(j_1,j_2,x) = 0$ unless (i_2 or i_3) are greater than zero and (i_3 or i_6) are greater than zero. Therefore (4.32) implies there exists a $\tilde{c}_Z \in (0,\infty)$ such that for $r \in \mathcal{S}_n$ and $j_1 < j_2$,

$$Z(j_1,j_2,nr) \le \tilde{c}_Z n^L [r_{j_1}r_{j_2} + r_{j_1-1}r_{j_2} + r_{j_1}r_{j_2-1} + r_{j_1-1}r_{j_2-1} + 1_{\{j_2=j_1+1\}}r_{j_1}]. \tag{4.80}$$

Combining this with (4.33) and (4.79), we have for some $\kappa_3', \kappa_3 \in (0,\infty)$ and all $n, K \in \mathbb{N}$

$$\begin{aligned}
&\sum_{j_1=0}^{K}\sum_{j_2=0}^{K} j_1 j_2 \langle M_{j_1}^n, M_{j_2}^n\rangle(T) \\
&\quad \le \frac{\kappa_3'}{n}\left[\int_0^T \sum_{j_1=0}^{K}\sum_{j_2=0}^{K}(j_1+1)(j_2+1)\pi_{j_1}^n(t)\pi_{j_2}^n(t)dt + \int_0^T \sum_{j=1}^{K+1} j(j+1)\pi_j^n(t)dt\right] \\
&\quad \le \frac{\kappa_3}{n}\left[\int_0^T \left(\sum_{j=0}^{K} j^2\pi_j^n(t) + (K+1)^2\pi_{K+1}^n(t)\right)dt + 1\right].
\end{aligned} \tag{4.81}$$

Recalling $\pi^n(t) = \pi^n(0) + A^n(t) + M^n(t)$ we have that for all $K, n \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}\int_0^T \sum_{j=0}^{K} j^2\pi_j^n(t)dt &= \int_0^T \sum_{j=0}^{K} j^2\pi_j^n(0)dt + \mathbb{E}\int_0^T \sum_{j=0}^{K} j^2 A_j^n(t)dt + \int_0^T \mathbb{E}\sum_{j=0}^{K} j^2 M_j^n(t)dt \\
&\le \mathbb{E}\int_0^T \sum_{j=0}^{K} j^2 A_j^n(t)dt + \kappa_4,
\end{aligned}$$

where $\kappa_4 = c_\pi(0)T$ and the last inequality follows on using the fact that $M_j^n(t)$ is a martingale. Thus, from (4.46), for some $\kappa_5 \in (0,\infty)$ and all $K, n \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}\int_0^T \sum_{j=0}^{K} j^2\pi_j^n(t)dt \le \frac{\kappa_5}{n^L}\mathbb{E}\int_0^T \sum_{j=1}^{K} j^2 \int_0^t [\zeta(j-1, n\pi^n(s)) - \zeta(j, n\pi^n(s))]ds\,dt \\
+ \kappa_5\mathbb{E}\int_0^T \sum_{j=1}^{K} j^2 \int_0^t [\pi_{j+1}^n(s) - \pi_j^n(s)]ds\,dt + \kappa_5.
\end{aligned} \tag{4.82}$$

Using the fact that for any $a_0, \ldots, a_K \in \mathbb{R}$

$$\sum_{j=1}^{K} j^2[a_{j-1} - a_j] = \sum_{j=1}^{K}[(j-1)^2 a_{j-1} - j^2 a_j + (2j-1)a_{j-1}] = -K^2 a_K + \sum_{j=1}^{K}(2j-1)a_{j-1}$$

and

$$\sum_{j=0}^{K} j^2[a_{j+1} - a_j] = \sum_{j=0}^{K}[(j+1)^2 a_{j+1} - j^2 a_j - (2j+1)a_{j+1}] = (K+1)^2 a_{K+1} - \sum_{j=0}^{K}(2j+1)a_{j+1}$$

in (4.82) we have that, for some $\kappa_6 \in (0,\infty)$ and all $K, n \in \mathbb{N}$

$$\begin{aligned}\mathbb{E}\int_0^T \sum_{j=0}^K j^2 \pi_j^n(t)dt &\le \frac{\kappa_5}{n^L}\mathbb{E}\int_0^T \int_0^t \sum_{j=0}^K (2j-1)\zeta(j-1, n\pi^n(s))ds\,dt \\ &\quad + \kappa_5 \mathbb{E}\int_0^T \int_0^t (K+1)^2 \pi_{K+1}^n(s)ds\,dt + \kappa_5 \\ &\le \kappa_6 \mathbb{E}\int_0^T [K^2 \sup_{0\le s\le t} \pi_{K+1}^n(s) + \sup_{0\le s\le t}\sum_{j=0}^K j\pi_j^n(s)]dt + \kappa_6\end{aligned} \tag{4.83}$$

where the second inequality follows from (4.73). Thus it follows from (4.75) and (4.81) that for some $\kappa_7 \in (0,\infty)$

$$\begin{aligned}\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^K jM^n(t)\right|^2 &\le \frac{\kappa_3}{n}\left[\int_0^T \mathbb{E}\sum_{j=0}^K j^2\pi_j^n(t)dt + \gamma_K^n T + 1\right] \\ &\le \frac{\kappa_7}{n}\left[1 + \gamma_K^n + \int_0^T \mathbb{E}\sup_{0\le u\le s}\left|\sum_{j=0}^K j\pi_j^n(u)\right|^2 ds\right],\end{aligned} \tag{4.84}$$

where $\gamma_K^n = \mathbb{E}(K^2 \sup_{0\le s\le T}\pi_{K+1}^n(s))$. Combining (4.70), (4.74), (4.84), and using the fact that $\left|\sum_{j=0}^\infty j\pi_j^n(0)\right| \le c_{\pi(0)}$,

$$\begin{aligned}\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^K j\pi_j^n(t)\right|^2 &\le \kappa_8\left(1 + \mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^K jA_j^n(t)\right|^2 + \mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^K jM_j^n(t)\right|^2\right) \\ &\le \kappa_9\left(1 + \gamma_K^n + \frac{1}{n}\int_0^T \mathbb{E}\sup_{0\le s\le t}\left|\sum_{j=0}^K j\pi_j^n(s)\right|^2 ds\right).\end{aligned}$$

By Gronwall's lemma (since the above inequality also holds for all $T_1 \le T$), there is a $\kappa_{10} \in (0,\infty)$ such that for all $n, K \in \mathbb{N}$

$$\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^K j\pi_j^n(t)\right|^2 \le \kappa_{10}\left(1 + \gamma_K^n\right).$$

Sending $K \to \infty$ and recalling from Lemma 12 that for each fixed n, as $K \to \infty$, $\gamma_K^n \to 0$ we have for all n

$$\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^\infty j\pi_j^n(t)\right|^2 \le \kappa_{10}$$

where κ_{10} is independent of n. This proves (4.68). Finally, (4.69) follows from (4.68) upon sending $K \to \infty$ in (4.83). □

4.4.2 Tightness

We now proceed with the proof of tightness of $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$. Let for $M \in \mathbb{R}_+$,

$$\mathcal{V}_M \doteq \left\{ r \in \mathcal{S} \middle| \sum_{i=0}^{\infty} i r_i \leq M \right\},$$

where $\mathcal{V}_M$ is equipped with the topology inherited from $\boldsymbol{\ell}_2$. We begin by establishing the following Lipschitz property for F on $\mathcal{V}_M$.

Lemma 14. *The map F is a Lipschitz function from $\mathcal{V}_M$ to $\boldsymbol{\ell}_2$ for each $M \in \mathbb{R}_+$. Namely, there exists an $C(M) \in (0, \infty)$ such that for any $r, \tilde{r} \in \mathcal{V}_M$,*

$$\|F(r) - F(\tilde{r})\|_2 \leq C(M)\|r - \tilde{r}\|_2. \tag{4.85}$$

Proof. Fix $M \in \mathbb{R}_+$. Let $r, \tilde{r} \in \mathcal{V}_M$ and, for $i_1 \in \mathbb{N}_0$ and $j, i_2 \in \mathbb{N}$, recall $R_{j,i_1,i_2}(r, \tilde{r})$ from (4.36). Using (4.37) and the fact that $r, \tilde{r} \in \mathcal{S}$, we have

$$\begin{aligned}
(R_{j,i_1,i_2}(r,\tilde{r}))^2 \leq\ & 3[r_j^{i_2} - \tilde{r}_j^{i_2}]^2 \\
& + 3\tilde{r}_j^{2i_2}\left[\left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2} - \left(\sum_{m=j+1}^{\infty} \tilde{r}_m\right)^{L-i_1-i_2}\right]^2 \\
& + 3\tilde{r}_j^{2i_2}\left[\left(\sum_{m=0}^{j-1} r_m\right)^{i_1} - \left(\sum_{m=0}^{j-1} \tilde{r}_m\right)^{i_1}\right]^2.
\end{aligned}$$

By an argument similar to the one used to derive (4.39) and an application of the Cauchy Schwartz inequality we have the following inequality for all $i_1, i_2 \leq L,\ i_2 > 0$,

$$(R_{j,i_1,i_2}(r,\tilde{r}))^2 \leq \kappa_1 \left([r_j - \tilde{r}_j]^2 + (j+1)\tilde{r}_j \|r - \tilde{r}\|_2^2 \right). \tag{4.86}$$

Using arguments analogous to those in the proof of Lemma 5 we have

$$\begin{aligned}\|F(r)-F(\tilde{r})\|_2 &\leq \kappa_2 \lambda L! \left(\sum_{j=0}^{\infty}\sum_{i_1=0}^{k-1}\sum_{i_2=1}^{L-i_1}\left[R_{j,i_1,i_2}(r,\tilde{r})\right]^2\right)^{1/2} + 2k\left(\sum_{j=0}^{\infty}(r-\tilde{r})_j^2\right)^{1/2} \\ &\leq \kappa_3\left(\sum_{j=0}^{\infty}\left[[r_j-\tilde{r}_j]^2+(j+1)\tilde{r}_j\|r-\tilde{r}\|_2^2\right]\right)^{1/2} + 2k\|r-\tilde{r}\|_2 \qquad (4.87)\\ &\leq \kappa_4\|r-\tilde{r}\|_2\left(1+\sum_{j=0}^{\infty}j\tilde{r}_j\right)^{1/2} + 2k\|r-\tilde{r}\|_2.\end{aligned}$$

Since $r,\tilde{r}\in\mathcal{V}_M$, (4.87) gives

$$\|F(r)-F(\tilde{r})\|_2 \leq \kappa_4(M+1)^{1/2}\|r-\tilde{r}\|_2 + 2k\|r-\tilde{r}\|_2$$

and thus with $C(M)\doteq\kappa_4(M+1)^{1/2}+2k$, (4.85) is satisfied for all $r,\tilde{r}\in\mathcal{V}_M$ which proves the result. □

Recall the process X^n introduced in (4.13) and $\bar{M}^n$ defined below (4.27). The following proposition gives tightness of $\{(X^n,\bar{M}^n)\}_{n\in\mathbb{N}}$.

Proposition 6. *Suppose that $\{\pi^n\}_{n\in\mathbb{N}}$ is as in the statement of Theorem 10 with $\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty}j^2\pi_j^n(0)<\infty$. Let $X^n(0)=\sqrt{n}(\pi^n(0)-\pi_0)$ and suppose that (4.20) is satisfied. Then $\{(X^n,\bar{M}^n)\}_{n\in\mathbb{N}}$ is a $\mathbb{C}$-tight sequence of $\mathbb{D}([0,T]:(\ell_2)^2)$-valued random variables.*

Proof. We will make use of Theorem 17 in the Appendix. We first prove that $\{\bar{M}^n\}_{n\in\mathbb{N}}$ is tight. In order to show that condition (A) in Theorem 17 is satisfied for $\{\bar{M}^n\}_{n\in\mathbb{N}}$ it suffices (cf. Theorem 2.3.2 in (Joffe and Métivier, 1986)) to show that the condition is satisfied for the real-valued process $\langle\bar{M}^n\rangle(t)\doteq\sum_{j=0}^{\infty}\langle\bar{M}_j^n\rangle(t)$. Fix $\varepsilon\in(0,T]$ and $N\in[0,T-\varepsilon]$. Let $\tau_n\leq N$ be a sequence of $\{\mathcal{F}_t^n\}$-stopping times. Then, (4.44) and (4.73) imply that for $\theta\in[0,\varepsilon]$

$$\begin{aligned}&|\langle\bar{M}^n(\tau_n+\theta)\rangle-\langle\bar{M}^n(\tau_n)\rangle| \\ &\quad=\left|\sum_{j=0}^{\infty}\left[\int_{\tau_n}^{\tau_n+\theta}\sum_{\ell\in\Sigma}\langle\Delta_\ell,e_j\rangle_2\frac{\lambda}{I(n)}\prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)}+k\int_{\tau_n}^{\tau_n+\theta}[\pi_{j+1}^n(s)-\pi_j^n(s)]ds\right]\right| \\ &\quad\leq\kappa_1\sum_{j=0}^{\infty}\int_{\tau_n}^{\tau_n+\theta}[\pi_j^n(s)+\pi_{j-1}^n(s)+\pi_{j+1}^n(s)]ds \\ &\quad\leq\kappa_1\varepsilon.\end{aligned}$$

Proof of (A) is now immediate.

We next show that $\{\bar{M}^n\}_{n\in\mathbb{N}}$ satisfies condition (T_1) in Theorem 17. For this we will apply Theorem 16. We first verify $\{\bar{M}^n(t)\}_{n\in\mathbb{N}}$ satisfies (a) of Theorem 16 for all $t \in [0,T]$. It follows from (4.34) that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\langle \bar{M}^n\rangle(T) = \sup_{n\in\mathbb{N}} n\mathbb{E}\langle M^n\rangle(T) \le \kappa_2. \tag{4.88}$$

This, combined with Doob's inequality, implies for each $N \in \mathbb{N}$

$$\sup_{n\in\mathbb{N}} \sum_{i=0}^{N} \mathbb{E} \sup_{0\le t\le T} \left|\bar{M}_i^n(t)\right| \le N + \sup_{n\in\mathbb{N}} \sum_{i=0}^{N} \mathbb{E}(\sup_{0\le t\le T} \bar{M}_i^n(t))^2 \le N + \kappa_3.$$

Using Markov's inequality, part (a) of Theorem 16 follows.

We now verify condition (b) in Theorem 16 for $\{\bar{M}^n(t)\}_{n\in\mathbb{N}}$ for each fixed $t \in [0,T]$. Note that $\langle \bar{M}_j^n\rangle(t) = n\langle M_j^n\rangle(t)$ and thus, from (4.28) and (4.33),

$$\langle \bar{M}_j^n\rangle(t) \le \kappa_4 \int_0^t (\pi_{j-1}^n(s) + \pi_j^n(s) + \pi_{j+1}^n(s))ds. \tag{4.89}$$

It follows from (4.89) and the Cauchy-Schwartz inequality that

$$\begin{aligned} \sum_{j=N}^{\infty} \mathbb{E}(\bar{M}_j^n(t))^2 = \sum_{j=N}^{\infty} \mathbb{E}\langle \bar{M}_j^n(t)\rangle &\le \kappa_5 \mathbb{E} \int_0^t \sum_{j=N-1}^{\infty} \pi_j^n(s)ds \\ &\le \kappa_5 \left(\sum_{j=N-1}^{\infty} \frac{1}{j^2}\right)^{1/2} \int_0^t \mathbb{E}\left(\sum_{j=N-1}^{\infty} j^2(\pi_j^n(s))^2\right)^{1/2} ds. \end{aligned} \tag{4.90}$$

From (4.69),

$$\sup_{n\in\mathbb{N}} \mathbb{E} \int_0^T \sum_{j=0}^{\infty} j^2(\pi_j^n(s))^2 ds \le \sup_{n\in\mathbb{N}} \mathbb{E} \int_0^T \sum_{j=0}^{\infty} j^2\pi_j^n(s)ds \doteq \kappa_6 < \infty. \tag{4.91}$$

Using this observation in (4.90) we have

$$\sum_{j=N}^{\infty} \mathbb{E}(\bar{M}_j^n(t))^2 \le \kappa_7 \left(\sum_{j=N-1}^{\infty} \frac{1}{j^2}\right)^{1/2} \int_0^t \mathbb{E}\left(\sum_{j=N-1}^{\infty} j^2\pi_j^n(s)\right)^{1/2} ds \le \kappa_8 \left(\sum_{j=N-1}^{\infty} \frac{1}{j^2}\right)^{1/2}.$$

From Markov's inequality we now see that for any $\delta > 0$

$$\lim_{N\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(\sum_{j=N}^{\infty} (\bar{M}_j^n(t))^2 > \delta \right) = 0$$

which verifies part (b) of Theorem 16. Thus we have shown that $\{\bar{M}^n(t)\}_{n\in\mathbb{N}}$ is a tight sequence of $\boldsymbol{\ell}_2$-valued random variables for all $t \in [0,T]$. From Theorem 17 it now follows that $\{\bar{M}^n\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{D}([0,T] : \boldsymbol{\ell}_2)$-valued random variables.

We will now argue that $\{X^n\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{D}([0,T] : \boldsymbol{\ell}_2)$-valued random variables. Again, via Theorem 17, it suffices to show that $\{X^n(t)\}_{n\in\mathbb{N}}$ is tight for every $t \in [0,T]$ (which will follow from verifying conditions (a) and (b) in Theorem 16) and that $\{X^n\}_{n\in\mathbb{N}}$ satisfies condition (A) of Theorem 17. We first show that, for all $t \in [0,T]$, condition (a) in Theorem 16 holds for $\{X^n(t)\}_{n\in\mathbb{N}}$. Namely we show that for each $N \in \mathbb{N}$ and $t \in [0,T]$

$$\lim_{A\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(\sum_{j=0}^{N} |X_j^n(t)| > A \right) = 0. \tag{4.92}$$

Fix $\varepsilon > 0$. From Lemma 13, there is a $M \in (0,\infty)$ such that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\left(\sup_{0\le t\le T} \sum_{j=0}^{\infty} j\pi_j^n(t) \right) \le \frac{M\varepsilon}{2}. \tag{4.93}$$

Let $B_M^n \doteq \{\sup_{0\le t\le T} \sum_{j=0}^{\infty} j\pi_j^n(t) \le M\}$. Then for $t \in [0,T]$ and $N \in \mathbb{N}$

$$\begin{aligned} \mathbb{P}\left(\sum_{j=0}^{N} |X_j^n(t)| > A \right) &\le \mathbb{P}\left(\sup_{0\le t\le T} \sum_{j=0}^{\infty} j\pi_j^n(t) > M \right) + \mathbb{P}\left(\sum_{j=0}^{N} |X_j^n(t)| > A, B_M^n \right) \\ &\le \frac{\varepsilon}{2} + \mathbb{P}\left(\sum_{j=0}^{N} |X_j^n(t)| > A, B_M^n \right). \end{aligned} \tag{4.94}$$

The Cauchy-Schwartz inequality yields

$$\sum_{i=0}^{N} |X_j^n(t)| \le \sqrt{N} \left(\sum_{j=0}^{N} |X_j^n(t)|^2 \right)^{1/2} \le \sqrt{N} \|X^n(t)\|_2. \tag{4.95}$$

Furthermore, from (4.23) and the triangle inequality,

$$\|X^n(t)\|_2 \le \|X^n(0)\|_2 + \|\bar{A}^n(t)\|_2 + \|\bar{M}^n(t)\|_2. \tag{4.96}$$

The definition of $\bar{A}^n$ in (4.27) gives

$$\|\bar{A}^n(t)\|_2 = \sqrt{n}\left\|A^n(t) - \int_0^t F(\pi(s))ds\right\|_2.$$

The moment bound (4.68) proved in Lemma 13 implies

$$\sup_{n\in\mathbb{N}}\mathbb{E}\sup_{0\le t\le T}\left|\sum_{j=0}^{\infty} j\pi_j^n(t)\right|^2 \doteq \kappa_7 < \infty. \tag{4.97}$$

and thus, for some $\kappa_8 \in (0,\infty)$,

$$\sup_{0\le t\le T}\left|\sum_{j=0}^{\infty} j\pi_j(t)\right|^2 \le \kappa_8 \tag{4.98}$$

as well. From (4.49) and the Lipschitz property proved in Lemma 14, with $M \ge \kappa_7 \vee \kappa_8$ on the set B_M^n,

$$\|\bar{A}^n(t)\|_2 \le \sqrt{n}\int_0^t \|F(\pi^n(s)) - F(\pi(s))\|_2 ds + \frac{\kappa_9}{\sqrt{n}} \le C(M)\int_0^t \|X^n(s)\|_2 ds + \frac{\kappa_9}{\sqrt{n}}.$$

Thus, from (4.96) and Gronwall's lemma, on the set B_M^n, for all $n \ge 1$

$$\sup_{0\le t\le T}\|X^n(t)\|_2 \le \kappa_{10}\left(\frac{1}{\sqrt{n}} + \|X^n(0)\|_2 + \sup_{0\le t\le T}\|\bar{M}^n(t)\|_2\right)e^{C(M)T}. \tag{4.99}$$

From (4.88) and Doob's inequality

$$\sup_{n\in\mathbb{N}}\mathbb{E}\sup_{0\le t\le T}\|\bar{M}^n(t)\|_2^2 < \infty. \tag{4.100}$$

Also by assumption, $X^n(0) \to x_0$ in $\boldsymbol{\ell}_2$. Thus for the given $\varepsilon > 0$, we can find α_0 such that for all $\alpha \ge \alpha_0$

$$\mathbb{P}\left(\sup_{0\le t\le T}\|X^n(t)\|_2 \ge \frac{\alpha}{\sqrt{N}}, B_M^n\right) \le \frac{\varepsilon}{2}.$$

Therefore from (4.94) and (4.95) we have that for all $A \ge \frac{\alpha_0}{\sqrt{N}}$,

$$\sup_{n\in\mathbb{N}}\mathbb{P}\left(\sum_{j=0}^{N}|X_j^n(t)| > A\right) \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary we get (4.92). Thus, we have verified part (a) of Theorem 16 for $\{X^n(t)\}_{n\in\mathbb{N}}$, for each $t \in [0,T]$.

We next consider part (b) of Theorem 16. Namely, we show that for every $\delta > 0$ and $t \in [0,T]$,

$$\lim_{N\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(\sum_{j=N}^{\infty} (X_j^n(t))^2 > \delta \right) = 0.$$

For this it suffices to show that

$$\sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\le t\le T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 < \infty. \tag{4.101}$$

Recalling that $X_j^n(t) = X_j^n(0) + \bar{A}_j^n(t) + \bar{M}_j^n(t)$ for each $j \in \mathbb{N}$, it follows that

$$\begin{aligned} \mathbb{E} \sup_{0\le t\le T} \sum_{j=0}^{\infty} j^2 (X_j^n(t))^2 \le 3 \sup_{n\in\mathbb{N}} \sum_{j=0}^{\infty} j^2 (X_j^n(0))^2 + 3 \sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\le t\le T} \sum_{j=0}^{\infty} j^2 (\bar{A}_j^n(t))^2 \\ + 3 \sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\le t\le T} \sum_{j=0}^{\infty} j^2 (\bar{M}_j^n(t))^2. \end{aligned} \tag{4.102}$$

Using the definitions of $\bar{A}^n$, A^n, and F in (4.27), (4.46), and (4.4), respectively, we can write

$$\begin{aligned} (\bar{A}_j^n(t))^2 \le \kappa_{11} \Bigg\{ & \int_0^t n \left[\frac{(n-L)!}{n!} \zeta(j, n\pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 ds \\ & + \int_0^t n \left[\frac{(n-L)!}{n!} \zeta(j-1, n\pi^n(s)) - \bar{\zeta}(j-1, \pi(s)) \right]^2 ds \\ & + \int_0^t n \left[\pi_j^n(s) - \pi_j(s) \right]^2 ds + \int_0^t n \left[\pi_{j+1}^n(s) - \pi_{j+1}(s) \right]^2 ds \Bigg\}. \end{aligned} \tag{4.103}$$

From (4.48) and in a similar manner as in (4.41) we have

$$\begin{aligned} n \left[\frac{(n-L)!}{n!} \zeta(j, n\pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 &\le \kappa_{12} \left\{ (\pi_j^n(s))^2 + n \left[\bar{\zeta}(j, \pi^n(s)) - \bar{\zeta}(j, \pi(s)) \right]^2 \right\} \\ &\le \kappa_{13} \left\{ (\pi_j^n(s))^2 + n \sum_{i_1=0}^{k-1} \sum_{i_2=1}^{L-i_1} R_{j,i_1,i_2}(\pi^n(s), \pi(s))^2 \right\}, \end{aligned}$$

where R_{j,i_1,i_2} is as in (4.36). By (4.86) and the Cauchy Schwartz inequality we now have,

$$nR_{j,i_1,i_2}(\pi^n(s),\pi(s))^2 \le \kappa_{14}\left[(X_j^n(s))^2+\pi_j(s)\left(\sum_{m=0}^{\infty}|X_m^n(s)|\right)^2\right]$$
$$\le \kappa_{14}\left[(X_j^n(s))^2+\pi_j(s)\left(\sum_{m=0}^{\infty}\frac{1}{m^2}\right)\sum_{m=0}^{\infty}m^2(X_m^n(s))^2\right].$$

Therefore

$$n\left[\frac{(n-L)!}{n!}\zeta(j,n\pi^n(s))-\bar{\zeta}(j,\pi(s))\right]^2$$
$$\le \kappa_{15}\left\{(\pi_j^n(s))^2+(X_j^n(s))^2+\pi_j(s)\sum_{m=0}^{\infty}m^2(X_m^n(s))^2\right\}.$$

Combining this estimate with (4.103) and (4.91) yields

$$\mathbb{E}\sum_{j=0}^{\infty}j^2(\bar{A}_j^n(t))^2 \le \kappa_{16}\mathbb{E}\left\{\int_0^t\sum_{j=1}^{\infty}j^2\left[(X_{j-1}^n(s))^2+(X_j^n(s))^2+(X_{j+1}^n(s))^2\right.\right.$$
$$\left.\left.+(\pi_j(s)+\pi_{j-1}(s))\sum_{m=0}^{\infty}m^2(X_m^n(s))^2\right]ds\right\}+\kappa_{16} \tag{4.104}$$
$$\le \kappa_{17}\mathbb{E}\int_0^t\left(1+\sum_{j=1}^{\infty}j^2\pi_j(s)\right)\left(\sum_{j=1}^{\infty}j^2(X_j^n(s))^2\right)ds+\kappa_{17}.$$

Additionally, it follows from (4.89) and (4.69) that,

$$\mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty}j^2\bar{M}_j^n(t)^2 \le \kappa_{18}'\mathbb{E}\sum_{j=0}^{\infty}j^2\langle\bar{M}_j^n\rangle_T \le \kappa_{18}\int_0^T\left[1+\mathbb{E}\sum_{j=0}^{\infty}j^2\pi_j^n(s)\right]ds \le \kappa_{19}.$$

Therefore, from (4.20), (4.102), and (4.104), for all $t\in[0,T]$,

$$\mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty}j^2(X_j^n(t))^2 \le \kappa_{20}+\kappa_{20}\int_0^T\left(1+\sum_{j=1}^{\infty}j^2\pi_j(t)\right)\mathbb{E}\sup_{0\le s\le t}\left(\sum_{j=1}^{\infty}j^2(X_j^n(s))^2\right)dt.$$

From (4.69) and Fatou's lemma, $\int_0^T\sum_{j=1}^{\infty}j^2\pi_j(s)ds<\infty$ and thus by Gronwall's lemma

$$\sup_{n\in\mathbb{N}}\mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty}j^2(X_j^n(t))^2 \le \kappa_{19}e^{\kappa_{20}\int_0^T\left(1+\sum_{j=1}^{\infty}j^2\pi_j(s)\right)}<\infty.$$

This proves (4.101) and verifies part (b) of Theorem 16 for $\{X^n(t)\}_{n\in\mathbb{N}}$ for each $t\in[0,T]$. Thus $\{X^n(t)\}_{n\in\mathbb{N}}$ is a tight sequence of $\boldsymbol{\ell}_2$-valued random variables for every $t\in[0,T]$.

We now show that condition (A) of Theorem 17 holds for $\{X^n\}_{n\in\mathbb{N}}$. Since $X^n(t) = X^n(0) + \bar{A}^n(t) + \bar{M}^n(t)$ and we have shown the condition is satisfied by $\{\bar{M}^n\}_{n\in\mathbb{N}}$, it suffices to show that the condition holds for $\{\bar{A}^n\}_{n\in\mathbb{N}}$. Let $N, \eta, \varepsilon, \theta > 0,\ N \le T - \theta$, and suppose $\{\tau_n\}_{n\in\mathbb{N}}$ is a family of stopping times such that $\tau_n \le N$. From the definition of $\bar{A}^n$ (cf. (4.27)) and (4.49) we have that

$$\left\|\bar{A}^n(\tau_n+\theta) - \bar{A}^n(\tau_n)\right\|_2 \le \int_{\tau_n}^{\tau_n+\theta} \sqrt{n}\left\|F(\pi^n(t)) - F(\pi(t))\right\|_2 dt + \frac{\kappa_{21}}{\sqrt{n}} \tag{4.105}$$

where κ_{21} is independent of the choice of τ_n and N. Fix $n_0 \in \mathbb{N}$ such that $\eta - \frac{\kappa_{21}}{\sqrt{n_0}} > 0$ and let $\eta' = \eta - \frac{\kappa_{21}}{\sqrt{n_0}}$. Recall κ_7, κ_8 introduced in (4.97) and (4.98), and B_M^n introduced below (4.93). Select $M \in (0,\infty)$ large enough that $M > \kappa_7 \vee \kappa_8$ and (4.93) holds. Then for all $n \ge n_0$,

$$\begin{aligned}
&\mathbb{P}\left\{\left\|\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))]dt\right\|_2 > \eta'\right\} \\
&\quad\le \mathbb{P}\left\{\left\|\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))]dt\right\|_2 > \eta', B_M^n\right\} + \mathbb{P}\left\{\sup_{0\le t\le T}\sum_{j=0}^{\infty} j\pi_j^n(t) > M\right\} \\
&\quad\le \mathbb{P}\left\{\left\|\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))]dt\right\|_2 > \eta', B_M^n\right\} + \frac{\varepsilon}{2}.
\end{aligned} \tag{4.106}$$

It follows from the Lipschitz property of F proved in Lemma 14 that

$$\mathbb{P}\left\{\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}\left\|F(\pi^n(t)) - F(\pi(t))\right\|_2 dt > \eta', B_M^n\right\} \le \mathbb{P}\left\{C(M)\int_{\tau_n}^{\tau_n+\theta}\left\|X^n(t)\right\|_2 dt > \eta', B_M^n\right\}. \tag{4.107}$$

Recall from (4.99) that for some $\tilde{C}(M) \in (0,\infty)$ on the set B_M^n

$$C(M)\sup_{0\le t\le T}\|X^n(t)\|_2 \le \tilde{C}(M)(1 + \sup_{0\le t\le T}\|\bar{M}^n(t)\|_2).$$

Thus from (4.107), Markov's inequality, and (4.100) we have

$$\begin{aligned}
\mathbb{P}\left\{\int_{\tau_n}^{\tau_n+\theta} \sqrt{n}\left\|F(\pi^n(t)) - F(\pi(t))\right\|_2 dt > \eta', B_M^n\right\} &\le \mathbb{P}\{\theta\tilde{C}(M)(1 + \sup_{0\le t\le T}\|\bar{M}^n(t)\|_2) > \eta'\} \\
&\le \frac{\theta\tilde{C}(M)(1 + \mathbb{E}\sup_{0\le t\le T}\|\bar{M}^n(t)\|_2)}{\eta'} \\
&\le \theta\tilde{C}(M)\kappa_{22}
\end{aligned} \tag{4.108}$$

Combining (4.106) and (4.108) gives, whenever $\theta \le \delta$,

$$\sup_{0\le\theta\le\delta} \mathbb{P}\left\{\left\| \int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))]dt \right\|_2 > \eta' \right\} \le C(M)\kappa_{22}\delta + \frac{\varepsilon}{2}.$$

Selecting δ small enough that the first term on the RHS is less than $\varepsilon/2$ we have,

$$\sup_{0\le\theta\le\delta} \mathbb{P}\left\{\left\| \int_{\tau_n}^{\tau_n+\theta} \sqrt{n}[F(\pi^n(t)) - F(\pi(t))]dt \right\|_2 > \eta' \right\} \le \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \tag{4.109}$$

Therefore, combining (4.105) and (4.109), gives

$$\sup_{n\ge n_0} \sup_{0\le\theta\le\delta} \mathbb{P}\left\{\left\| \bar{A}^n(\tau_n+\theta) - \bar{A}^n(\tau_n) \right\|_2 > \eta \right\} \le \varepsilon$$

which shows that condition (A) of Theorem 17 is satisfied for $\{\bar{A}^n\}_{n\in\mathbb{N}}$. Therefore, as discussed earlier, $\{X^n\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{D}([0,T] : \ell_2)$-valued random variables and thus $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{D}([0,t] : (\ell_2)^2)$-valued random variables.

Finally, the $\mathbb{C}$-tightness of $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$ is immediate from the estimate

$$j_T(X^n) = j_T(\bar{M}^n) \le \frac{2+2k}{\sqrt{n}}, \qquad n \in \mathbb{N}$$

which follows as in the proof of Proposition 5. □

4.4.3 Convergence

In this section we give the proofs of Proposition 4 and Theorem 14. Since we have shown tightness of $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$ in Section 4.4.2, all that remains in order to complete the proof of Theorem 14 is to characterize the weak limit points of this sequence of processes. This will be argued by showing that the limit point of any weakly convergent subsequence of $\{X^n\}_{n\in\mathbb{N}}$ will be a solution to the SDE (4.14) and that uniqueness holds for (4.14) in an appropriate class, which will also prove Proposition 4. We begin by establishing a uniform integrability property for the sequence $\{\bar{M}^n\}_{n\in\mathbb{N}}$.

Lemma 15. *Suppose $\{\pi^n\}_{n\in\mathbb{N}}$ satisfies conditions in Proposition 6. Then the sequence $\{\sup_{0\le t\le T} \sum_{j=0}^{\infty} |\bar{M}_j^n(t)|^2\}_{n\in\mathbb{N}}$ is uniformly integrable.*

Proof. It follows from the Cauchy-Schwartz and Burkholder-Davis-Gundy inequalities that

$$\sup_{n\in\mathbb{N}}\mathbb{E}\sup_{0\le t\le T}\left(\sum_{j=0}^{\infty}|\bar{M}_j^n(t)|^2\right)^2 \le \sup_{n\in\mathbb{N}}\left(\sum_{m=0}^{\infty}\frac{1}{m^2}\right)\sum_{j=0}^{\infty}\mathbb{E}\sup_{0\le t\le T} j^2|\bar{M}_j^n(t)|^4 \le \kappa_1\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty} j^2\mathbb{E}[\bar{M}_j^n](T)^2. \tag{4.110}$$

Recalling the definition of M^n from (4.25), for each j, $\mathbb{E}[\bar{M}_j^n](T)^2$ can be written as

$$\mathbb{E}[\bar{M}_j^n](T)^2 = \mathbb{E}\left\{\sum_{\ell\in\Sigma}\frac{1}{n}\langle e_j, \Delta_\ell\Delta_\ell^T e_j\rangle_2 N_\ell\left(\frac{n\lambda}{\binom{n}{L}}\int_0^T\prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)}ds\right)\right.$$
$$\left.+\frac{1}{n}\left[D_j\left(k\int_0^T n\pi_j^n(s)ds\right)+D_{j+1}\left(k\int_0^T n\pi_{j+1}^n(s)ds\right)\right]\right\}^2$$

The first term in the expectation on the RHS of the above equation corresponds to the stream of incoming jobs assigned to queues of length j. From (4.29), (4.30), (4.33), and the independence of $N_\ell, N_{\ell'}$ for $\ell \neq \ell'$ we have

$$\mathbb{E}\left(\sum_{\ell\in\Sigma}\frac{1}{n}\langle e_j, \Delta_\ell\Delta_\ell^T e_j\rangle_2 N_\ell\left(\frac{n\lambda}{\binom{n}{L}}\int_0^T\prod_{i=0}^{\infty}\binom{n\pi_i^n(s)}{\rho_i(\ell)}ds\right)\right)^2 \le \kappa_2\mathbb{E}\int_0^T[\pi_j^n(s)+\pi_{j-1}^n(s)]ds.$$

Similarly,

$$\mathbb{E}\left(\frac{1}{n}\left[D_j\left(k\int_0^T n\pi_j^n(s)ds\right)+D_{j+1}\left(k\int_0^T n\pi_{j+1}^n(s)ds\right)\right]\right)^2 \le \kappa_3\mathbb{E}\int_0^T[\pi_j^n(s)+\pi_{j+1}^n(s)]ds.$$

Combining these estimates and using (4.69)

$$\sup_{n\in\mathbb{N}}\sum_{j=0}^{\infty} j^2 E[\bar{M}_j^n](T)^2 \le \kappa_4\sup_{n\in\mathbb{N}}\mathbb{E}\int_0^T\sum_{j=1}^{\infty} j^2(\pi_{j-1}^n(s)+\pi_j^n(s)+\pi_{j+1}^n(s)) < \infty$$

which, in view of (4.110), gives the desired uniform integrability. □

The following lemma together with (4.101) shows that any weak limit point X of $\{X^n\}_{n\in\mathbb{N}}$ satisfies $X(t) \in \tilde{\boldsymbol{\ell}}_2$ for all $t \in [0,T]$, a.s.

Lemma 16. *Let z^n, z be $\mathbb{D}([0,T]:\boldsymbol{\ell}_2)$-valued random variables such that*

$$\sup_{0\le t\le T}\|z^n(t)-z(t)\|_2 \to 0 \text{ in probability as } n\to\infty.$$

Suppose that $\sup_{n\in\mathbb{N}} \mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty} j^2(z_j^n(t))^2 < \infty$. *Then* $\sup_{0\le t\le T}\sum_{j=0}^{\infty} j^2(z_j(t))^2 < \infty$ *almost surely and* $\sup_{0\le t\le T}\left|\sum_{j=0}^{\infty} z_j^n(t) - \sum_{j=0}^{\infty} z_j(t)\right| \to 0$ *in probability.*

Proof. Let $\kappa = \sup_{n\in\mathbb{N}} \mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty} j^2[z_j^n(t)]^2$. Note that

$$\sup_{n\in\mathbb{N}} \mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty}|z_j^n(t)| \le \left(\sum_{j=1}^{\infty}\frac{1}{j^2}\right)^{1/2}\sqrt{\kappa} < \infty.$$

Also, by Fatou's lemma $\mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty} j^2(z_j(t))^2 \le \kappa$ and so we have $\sup_{0\le t\le T}\sum_{j=0}^{\infty}|z_j(t)| < \infty$ almost surely as well. Now

$$\begin{aligned}
&\mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{j=0}^{\infty} z_j^n(t) - \sum_{j=0}^{\infty} z_j(t)\right| \wedge 1\right] \\
&\quad\le \mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{j=0}^{m} z_j^n(t) - \sum_{j=0}^{m} z_j(t)\right| \wedge 1\right] + \mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{j=m+1}^{\infty} z_j^n(t)\right| \wedge 1\right] + \mathbb{E}\left[\sup_{0\le t\le T}\left|\sum_{j=m+1}^{\infty} z_j(t)\right| \wedge 1\right] \\
&\quad\equiv T_1^m(n) + T_2^m(n) + T_3^m(n).
\end{aligned}$$

Then for $\kappa_1 \in (0,\infty)$

$$(T_2^m(n))^2 \le \left(\sum_{j=m+1}^{\infty}\frac{1}{j^2}\right)\kappa_1 \quad \text{and} \quad (T_3^m(n))^2 \le \left(\sum_{j=m+1}^{\infty}\frac{1}{j^2}\right)\kappa_1.$$

The result now follows on first sending $n\to\infty$ and then $m\to\infty$. □

The following result that shows that $\Phi(t)$ is a trace class operator will be useful in characterizing the martingale term in the limiting diffusion. Note that, from the definition (4.17), $\Phi(t)$ is a non-negative operator.

Lemma 17. *For each* $t \in [0,T]$, $\Phi(t)$ *is a non-negative trace class operator. Denote by* $a(t)$ *the non-negative square root of* $\Phi(t)$. *Then* $\int_0^T \|a(s)\|_{HS}^2 ds < \infty$.

Proof. We first show that $\Phi(t)$ is a trace class operator. Since $\Phi(t)$ is non-negative (and hence self-adjoint) it suffices to show

$$\sum_{j=0}^{\infty}\langle e_j, \Phi(s)e_j\rangle_2 < \infty$$

Using an argument similar to that used in the derivation of (4.31) one can write $\langle e_j, \Phi(s)e_j\rangle_2$, as

$$\langle e_j, \Phi(s)e_j\rangle_2 = \lambda L! \bar{Z}(j, \pi(s)) + k(\pi_j(s) + \pi_{j+1}(s)) \tag{4.111}$$

where the definition of $\bar{Z}$ is analogous to Z, given as,

$$\begin{aligned}\bar{Z}(j, \pi(s)) \doteq \sum_{i_1=0}^{k-2} \frac{\left(\sum_{m=0}^{j-1} \pi_m(s)\right)^{i_1}}{i_1!} \sum_{i_2=0}^{L-i_1} \frac{\pi_{j-1}(s)^{i_2}}{i_2!} \sum_{i_3=0}^{L-i_1-i_2} [i_2 \wedge (k-i_1)_+ - i_3 \wedge (k-i_1-i_2)_+]^2 \\ \times \frac{\pi_j(s)^{i_3}}{i_3!} \frac{\left(\sum_{m=j+1}^{\infty} \pi_m(s)\right)^{L-i_1-i_2-i_3}}{(L-i_1-i_2-i_3)!}.\end{aligned} \tag{4.112}$$

Using arguments as in (4.33) and (4.80) it is easy to see that there exists $c_{\bar{Z}} \in (0, \infty)$ such that for all $j \in \mathbb{N}_0$,

$$\bar{Z}(j, \pi(s)) \leq c_{\bar{Z}}(\pi_{j-1}(s) + \pi_j(s)). \tag{4.113}$$

From (4.111) and (4.113) it follows that there exists a $\kappa_1 \in (0, M)$ such that,

$$\sum_{j=0}^{\infty} \langle e_j, \Phi(t)e_j\rangle_2 \leq \kappa_2 \sum_{j=0}^{\infty} [\pi_{j-1}(t) + \pi_j(t) + \pi_{j+1}(t)] \leq 3\kappa_1.$$

Therefore, $\Phi(t)$ is a trace class operator. Finally, note that

$$\int_0^T \|a(s)\|_{\mathrm{HS}}^2 ds = \int_0^T \sum_{j=0}^{\infty} \langle a(s)e_j, a(s)e_j\rangle_2 ds = \int_0^T \sum_{j=0}^{\infty} \langle e_j, \Phi(s)e_j\rangle_2 ds \leq 3\kappa_1 T$$

which completes the proof. □

We now proceed with the proofs of Proposition 4 and Theorem 14.

Proof of Proposition 4. The existence of a $(X(t))_{0\leq t\leq T}$ as in the statement of Proposition 4 will be proved as part of Theorem 14. We now consider the second statement in Proposition 4 and let $(X(t))_{0\leq t\leq T}$, $(\tilde{X}(t))_{0\leq t\leq T}$ be two $\{\mathcal{F}_t\}$-adapted processes solving (4.15) with sample paths in $\mathbb{C}([0,T] : \boldsymbol{\ell}_2)$ such that $X(t) \in \tilde{\boldsymbol{\ell}}_2$ and $\tilde{X}(t) \in \tilde{\boldsymbol{\ell}}_2$ for all t, almost surely. In order to show that $X(t) = \tilde{X}(t)$ for all $t \in [0,T]$ almost surely it suffices to show the following Lipschitz property

on G: There exists a $C \in (0,\infty)$ such that for all $x, \tilde{x} \in \tilde{\ell}_2$,

$$\sup_{0\le t\le T} \|G(x,\pi(t)) - G(\tilde{x},\pi(t))\|_2 \le C\|x-\tilde{x}\|_2. \tag{4.114}$$

Note from (4.4), (4.5), and (4.19) that for $j \in \mathbb{N}_0$ and $(x,r) \in \tilde{\ell}_2 \times \mathcal{S}$,

$$G_j(x,r) = \lambda L![\xi^1_{j-1}(x,r) - \xi^1_j(x,r) + \xi^2_{j-1}(x,r) - \xi^2_j(x,r) + \xi^3_{j-1}(x,r) - \xi^3_j(x,r)] + k\xi^4_j(x) \tag{4.115}$$

where

$$\xi^1_j(x,r) \doteq \sum_{i_1=0}^{k-1} i_1 \frac{\left(\sum_{m=0}^{j-1} r_m\right)^{i_1-1}}{i_1!} \sum_{i_2=1}^{L-i_1} [i_2 \wedge (k-i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{\left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2}}{(L-i_1-i_2)!} \sum_{m=0}^{j-1} x_m,$$

$$\xi^2_j(x,r) \doteq \sum_{i_1=0}^{k-1} \frac{\left(\sum_{m=0}^{j-1} r_m\right)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} i_2[i_2 \wedge (k-i_1)] \frac{(r_j)^{i_2-1}}{i_2!} \frac{\left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2}}{(L-i_1-i_2)!} x_j,$$

$$\xi^3_j(x,r) \doteq \sum_{i_1=0}^{k-1} \frac{\left(\sum_{m=0}^{j-1} r_m\right)^{i_1}}{i_1!} \sum_{i_2=1}^{L-i_1} (L-i_1-i_2)[i_2 \wedge (k-i_1)] \frac{(r_j)^{i_2}}{i_2!} \frac{\left(\sum_{m=j+1}^{\infty} r_m\right)^{L-i_1-i_2-1}}{(L-i_1-i_2)!} \sum_{m=j+1}^{\infty} x_m,$$

and

$$\xi^4_j(x) = [x_{j+1} - x_j].$$

Also, let $\xi^i \doteq (\xi^i_j)_{j=0}^{\infty}$ for $i = 1,2,3,4$. Using the triangle inequality, it suffices to show that (4.114) holds with G replaced with $\xi^i, i = 1,2,3,4$. Since $\pi(t) \in \mathcal{S}$ for all $t \in [0,T]$

$$\begin{aligned}
\sup_{0\le t\le T} \|\xi^1(x,\pi(t)) - \xi^1(\tilde{x},\pi(t))\|_2^2 &\le \kappa_1' \sup_{0\le t\le T} \sum_{j=0}^{\infty} \pi_j(t)^2 \left[\sum_{m=0}^{j-1} x_m - \sum_{m=0}^{j-1} \tilde{x}_m\right]^2 \\
&\le \kappa_1' \sup_{0\le t\le T} \sum_{j=0}^{\infty} j\pi_j(t)\|x-\tilde{x}\|_2^2 \\
&\le \kappa_1 \|x-\tilde{x}\|_2^2,
\end{aligned} \tag{4.116}$$

where the last inequality is from (4.98). Also,

$$\sup_{0\le t\le T} \|\xi^2(x,\pi(t)) - \xi^2(\tilde{x},\pi(t))\|_2^2 \le \kappa_2 \sum_{j=0}^{\infty} [x_j - \tilde{x}_j]^2 = \kappa_2 \|x-\tilde{x}\|_2^2.$$

Using the fact that $\sum_{m=0}^{\infty} x_m = \sum_{m=0}^{\infty} \tilde{x}_m = 0$ and the calculation in (4.116)

$$\begin{aligned}\sup_{0\le t\le T} \|\xi^3(x,\pi(t)) - \xi^3(\tilde{x},\pi(t))\|_2^2 &\le \kappa_3' \sup_{0\le t\le T} \sum_{j=0}^{\infty} \pi_j(t)^2 \left[\sum_{m=j+1}^{\infty} x_m - \sum_{m=j+1}^{\infty} \tilde{x}_m\right]^2 \\ &= \kappa_3' \sup_{0\le t\le T} \sum_{j=0}^{\infty} \pi_j(t)^2 \left[\sum_{m=0}^{j} \tilde{x}_m - \sum_{m=0}^{j} x_m\right]^2 \\ &\le \kappa_3 \|x-\tilde{x}\|_2^2.\end{aligned}$$

Finally,

$$\|\xi^4(x) - \xi^4(\tilde{x})\|_2^2 \le \sum_{j=0}^{\infty} [x_j - \tilde{x}_j]^2 + \sum_{j=0}^{\infty} [x_{j+1} - \tilde{x}_{j+1}]^2 \le 2\|x-\tilde{x}\|_2^2.$$

Combining the above Lipschitz estimates for $\xi^i,\ i = 1,2,3,4$, we have (4.114) and the result follows. □

We now proceed to the proof of Theorem 14.

Proof of Theorem 14. From Proposition 6 $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$ is $\mathbb{C}$-tight in $\mathbb{D}([0,T] : (\ell_2)^2)$. Suppose $(X, \bar{M})$ is a weak limit of a subsequence of $\{(X^n, \bar{M}^n)\}_{n\in\mathbb{N}}$ (also indexed by $\{n\}$) given on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Let $m \in \mathbb{N}$ and $\mathcal{H} : (\ell_2 \times \ell_2)^m \to \mathbb{R}$ be a bounded and continuous function. For $s \le t \le T$ and $0 \le t_1 \le \ldots \le t_m \le s$ we let $\xi_i^n = (X^n(t_i), \bar{M}^n(t_i))$ and $\xi_i = (X(t_i), \bar{M}(t_i))$. Then, for all $j \in \mathbb{N}_0$,

$$\mathbb{E}\mathcal{H}(\xi_1,\ldots,\xi_m)[\bar{M}_j(t) - \bar{M}_j(s)] = \lim_{n\to\infty} \mathbb{E}\mathcal{H}(\xi_1^n,\ldots,\xi_m^n)[\bar{M}_j^n(t) - \bar{M}_j^n(s)] = 0$$

where the first equality comes from the uniform integrability property proved in Lemma 15 and the second comes from the fact that $\bar{M}^n$ is a martingale for each $n \in \mathbb{N}$. It follows that $\bar{M}$ is a $\{\mathcal{F}_t\}$-martingale where $\mathcal{F}_t = \sigma\{X(s), \bar{M}(s), s \le t\}$.

As was shown in (4.79),

$$\begin{aligned}\langle \bar{M}_i^n, \bar{M}_j^n\rangle(t) &= n\langle M_i^n, M_j^n\rangle(t) \\ &= \frac{\lambda}{\binom{n}{L}} \int_0^t Z(i,j,n\pi^n(s))ds - k\int_0^t 1_{\{i=j+1\}}\pi_i^n(s) \\ &\quad - k\int_0^t 1_{\{i+1=j\}}\pi_j^n(s)ds + k\int_0^t 1_{\{i=j\}}(\pi_j^n(s) + \pi_{j+1}^n(s))ds\end{aligned} \tag{4.117}$$

(see (4.31) and (4.77) for definition of Z). Using similar arguments as in (4.77), we have the estimate

$$\langle e_i, \Phi(s)e_j\rangle_2 = \lambda L!\bar{Z}(i,j,\pi(s)) - k1_{\{i+1=j\}}\pi_j(s) - k1_{\{i=j+1\}}\pi_i(s) + k1_{\{i=j\}}(\pi_j(s)+\pi_{j+1}(s)),$$

where for $i<j$,

$$\begin{aligned}
\bar{Z}(i,j,\pi(s)) \doteq & \sum_{i_1=0}^{k-2} \frac{\left(\sum_{m=0}^{i-2}\pi_m(s)\right)^{i_1}}{i_1!} \sum_{i_2=0}^{k-i_1-1} \frac{\pi_{i-1}(s)^{i_2}}{i_2!} \sum_{i_3=0}^{k-i_1-i_2-1} [i_2-i_3]\frac{\pi_i(s)^{i_3}}{i_3!} \\
& \times \sum_{i_4=0}^{k-i_1-i_2-i_3-1} \frac{\left(\sum_{m=i+1}^{j-2}\pi_m(s)\right)^{i_4}}{i_4!} \sum_{i_5=0}^{L-\sum_{n=1}^4 i_n} \frac{\pi_{j-1}(s)^{i_5}1_{\{j>i+1\}}}{i_5!} \\
& \times \sum_{i_6=0}^{L-\sum_{n=1}^5 i_n} \left[(1_{\{j=i+1\}}(i_3-i_5)+i_5) \wedge \left(k-\sum_{n=1}^4 i_n\right)_+ - i_6 \wedge \left(k-\sum_{n=1}^5 i_n\right)_+\right] \\
& \times \frac{\pi_j(s)^{i_6}}{i_6!} \frac{\left(\sum_{m=j+1}^{\infty}\pi_m(s)\right)^{L-\sum_{n=1}^6 i_n}}{\left(L-\sum_{n=1}^6 i_n\right)!},
\end{aligned}$$

for $i>j$, $\bar{Z}(i,j,\pi(s)) \doteq \bar{Z}(j,i,\pi(s))$, and for $i=j$, $\bar{Z}(j,j,\pi(s)) \doteq \bar{Z}(j,\pi(s))$, where $\bar{Z}(j,r)$ is defined in (4.112). Using arguments similar to those used in (4.47) and (4.48) one can write

$$\left|Z(i,j,n\pi^n(s)) - \frac{n!}{(n-L)!}\bar{Z}(i,j,\pi^n(s))\right| \le \kappa_1 n^{L-1}.$$

It follows from this, (4.117), (4.111), and the fact that $\pi^n \to \pi$ in probability that

$$\sup_{0\le t\le T}\left|\langle \bar{M}_i^n(t), \bar{M}_j^n(t)\rangle - \int_0^t \langle e_i, \Phi(s)e_j\rangle_2 ds\right| \to 0$$

in probability. A similar argument as in Lemma 15 shows that $\{\langle \bar{M}_i^n, \bar{M}_j^n\rangle_t\}_{n\in\mathbb{N}}$ is uniformly integrable for each $t\in[0,T]$ and $i,j\in\mathbb{N}_0$. Applying the above convergence and uniform integrability properties,

$$\begin{aligned}
&\mathbb{E}\mathcal{H}(\xi_1,\dots,\xi_m)[\langle \bar{M}_i, \bar{M}_j\rangle_t - \langle \bar{M}_i, \bar{M}_j\rangle_s - \int_s^t \langle e_i, \Phi(u)e_j\rangle_2 du] \\
&\quad = \lim_{n\to\infty} \mathbb{E}\mathcal{H}(\xi_1^n,\dots,\xi_m^n)[\langle \bar{M}_i^n, \bar{M}_j^n\rangle_t - \langle \bar{M}_i^n, \bar{M}_j^n\rangle_s - \int_s^t \langle e_i, \Phi(u)e_j\rangle_2 du] = 0.
\end{aligned}$$

Also from Lemma 15 and Fatou's lemma $\mathbb{E}\sup_{0\le t\le T}\sum_{j=0}^{\infty}|\bar{M}_j(t)|^2 < \infty$. Thus we have that $\bar{M} \doteq (\bar{M}_j)_{j\in\mathbb{N}_0}$ is a collection of square integrable $\{\mathcal{F}_t\}$-martingales with

$$\langle \bar{M}_i, \bar{M}_j\rangle(t) = \int_0^t \langle e_i, \Phi(s)e_j\rangle_2 ds, \qquad t\in[0,T].$$

From Theorem 8.2 of (Da Prato and Zabczyk, 2014) it now follows that there is a $\boldsymbol{\ell}_2$-cylindrical Brownian motion $\{(W_t(h))_{0\le t\le T} : h \in \boldsymbol{\ell}_2\}$ on some extension $(\bar{\Omega}, \bar{\mathcal{F}}, \bar{\mathbb{P}}, \{\bar{\mathcal{F}}_t\})$ of the filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ such that

$$\bar{M}(t) = \int_0^t a(s)dW(s). \tag{4.118}$$

Recall the representation of X^n in terms of $\bar{A}^n$ and $\bar{M}^n$ from (4.26). We now argue that together with X^n and $\bar{M}^n$, $\bar{A}^n(\cdot)$ converges to $\int_0^{\cdot} G(X(s),\pi(s))ds$ in $\mathbb{D}([0,T]:\boldsymbol{\ell}_2)$, in distribution, as $n\to\infty$ (along the chosen subsequence). The definition of $\bar{A}^n$ in (4.27) and the estimate in (4.49) imply that

$$\sup_{0\le t\le T}\left\|\bar{A}^n(t) - \int_0^t \sqrt{n}[F(\pi^n(s)) - F(\pi(s))]ds\right\|_2 \le \frac{\kappa_2}{\sqrt{n}}. \tag{4.119}$$

For $r,\tilde{r}\in\mathcal{S}$ such that $(r-\tilde{r})\in\tilde{\boldsymbol{\ell}}_2$, the i-th component of $F(r)-F(\tilde{r})$ can be written

$$\begin{aligned}
[F(r)-F(\tilde{r})]_i &= \int_0^1 \frac{\partial}{\partial u}F_i(ru+(1-u)\tilde{r})du \\
&= \int_0^1 G_i((r-\tilde{r}), ru+(1-u)\tilde{r})du \\
&= G_i(r-\tilde{r},\tilde{r}) + \int_0^1 [G_i((r-\tilde{r}, ru+(1-u)\tilde{r}) - G_i(r-\tilde{r},\tilde{r})]du.
\end{aligned}$$

Therefore, observing that $cG_i(x,r) = G_i(cx,r)$ for $c\in\mathbb{R}$ and $(x,r)\in\tilde{\boldsymbol{\ell}}_2\times\mathcal{S}$ and noting from (4.101) that $X^n(s)\in\tilde{\boldsymbol{\ell}}_2$ for every $s\in[0,T]$ almost surely, we can write

$$\sqrt{n}[F(\pi^n(s)) - F(\pi(s))]_i = G_i(X^n(s),\pi(s)) + R_i^n(s) \tag{4.120}$$

where

$$R_i^n(s) = \int_0^1 [G_i(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G_i(X^n(s),\pi(s))]du.$$

Thus

$$\sqrt{n}[F(\pi^n(s)) - F(\pi(s))] = G(X^n(s), \pi(s)) + R^n(s)$$

where $R^n(s) \doteq (R_i^n(s))_{i\in\mathbb{N}_0}$. We now show that $\int_0^T \|R^n(s)\|_2 ds \to 0$ in probability as $n \to \infty$. Since $\sum_{m=0}^{j} X_m^n(s) = -\sum_{m=j+1}^{\infty} X_m^n(s)$, it follows from (4.37) that for $r, \tilde{r} \in \mathcal{S}$

$$\begin{aligned}
&\|\xi^i(X^n(s), r) - \xi^i(X^n(s), \tilde{r})\|_2^2 \\
&\quad \le \kappa_3' \sum_{j=0}^{\infty} \left(\sum_{m=0}^{j} |X_m^n(s)| \right)^2 \left[[r-\tilde{r}]_j^2 + \tilde{r}_j \left(\sum_{i=0}^{j-1} [r-\tilde{r}]_i \right)^2 + \tilde{r}_j \left(\sum_{i=j+1}^{\infty} [r-\tilde{r}]_i \right)^2 \right] \\
&\quad \le \kappa_3 \left(\sum_{j=0}^{\infty} j^2 |X_j^n(s)|^2 \right) \sum_{j=0}^{\infty} \left[j\tilde{r}_j \|r-\tilde{r}\|_2^2 + [r-\tilde{r}]_j^2 \right]
\end{aligned}$$

for $i = 1, 2, 3$. The triangle inequality, (4.115), and the observation that $\sup_{0\le s\le T} \sum_{j=0}^{\infty} j\pi_j(s) < \infty$ (see (4.98)) then implies that

$$\|G(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G(X^n(s), \pi(s))\|_2^2 \le \kappa_3 \left(\sum_{j=0}^{\infty} j^2 |X_j^n(s)|^2 \right) \|\pi^n(s) - \pi(s)\|_2^2.$$

Since $\sup_{0\le s\le T} \|\pi^n(s) - \pi(s)\|_2 \to 0$ in probability and, from (4.101), $\sup_{n\in\mathbb{N}} \mathbb{E} \sup_{0\le s\le T} \sum_{j=0}^{\infty} j^2 |X_j^n(s)|^2 < \infty$, it follows that

$$\sup_{0\le u\le 1} \sup_{0\le s\le T} \|G(X^n(s), \pi^n(s)u + (1-u)\pi(s)) - G(X^n(s), \pi(s))\|_2 \to 0$$

in probability, as $n \to \infty$ and thus

$$\int_0^T \|R_n(s)\|_2 ds \to 0, \text{ in probability.} \tag{4.121}$$

In view of (4.119), (4.120), and (4.121) it now suffices to show that, along the subsequence

$$\left(X^n, \bar{M}^n, \int_0^{\cdot} G(X^n(s), \pi(s)) ds \right) \Rightarrow \left(X, \bar{M}, \int_0^{\cdot} G(X(s), \pi(s)) ds \right)$$

in $\mathbb{D}([0,T] : (\ell_2)^3)$. By appealing to the Skorohod representation theorem we can assume without loss of generality that $(X^n, \bar{M}^n)$ converges almost surely in $\mathbb{D}([0,T] : (\ell_2)^2)$ to $(X, \bar{M})$.

From (4.101) and Fatou's lemma we also have

$$\sup_{0\le t\le T}\sum_{j=0}^{\infty} j^2(X_j(t))^2 < \infty \quad \text{a.s.}$$

Also, since $\sum_{j=0}^{\infty} X_j^n(t) = 0$ for all $t \in [0,T]$ and $n \in \mathbb{N}$, by Lemma 16 and (4.101), we have that $\sum_{j=0}^{\infty} X_j(t) = 0$ for all $t \in [0,T]$ almost surely as well. It then follows that $X^n(t), X(t) \in \tilde{\boldsymbol{\ell}}_2$ for all $t \in [0,T]$ almost surely for all $n \in \mathbb{N}$. From the Lipschitz property in (4.114) it now follows that, as $n \to \infty$,

$$\int_0^T \|G(X^n(s),\pi(s)) - G(X(s),\pi(s))\|_2 ds \le C\int_0^T \|X^n(s) - X(s)\|_2 ds \to 0, \tag{4.122}$$

which proves the desired convergence. Together with (4.26) and the representation (4.118) we now have that the limit point $(X,\bar{M})$ satisfies

$$X(t) = x_0 + \int_0^t G(X(s),\pi(s))ds + \int_0^t a(s)dW(s)$$

almost surely for all $t \in [0,T]$. Since $X(t) \in \tilde{\boldsymbol{\ell}}_2$ for all $t \in [0,T]$ almost surely, this in particular proves the existence part of Proposition 4. Finally the uniqueness part of Proposition 4 (which was established earlier in this section) now says that X^n converges in distribution along the full sequence to the unique weak solution of (4.14) with values in $\tilde{\boldsymbol{\ell}}_2$. The result follows. □

4.5 Numerical Results

In this section, we present some simulation results comparing the pre-limit n-server system with results of the corresponding law of large number and central limit approximations. We consider a system with $n = 10,000$ servers. For all combinations of L and k in the set $\{(L,k) \in \mathbb{N}\times \mathbb{N} : 2 \le L \le 5, k < L\}$, we simulate 1,000 realizations of both the n-server system and the diffusion approximation given through Theorem 14 using parameters $T = 10$, $\lambda = .9$, and $c = 1$. Note that since the limiting processes are infinite dimensional we must truncate to a finite dimensional approximation in order to perform simulations. In our numerical approximations, we truncate to the first 20 coordinates. All computations were performed in Matlab. A numerical ODE solver (ode45) was used to compute the ODE corresponding to the law of large number limit.

The limit diffusion was simulated using Euler's method with step sizes of .1. The realizations of the diffusion were used to create 95% confidence intervals for the following metrics at time T; the number of empty queues, the number of "large" queues (queues with more than 5 jobs), and the mean queue length. The coverage rates (i.e. the proportion of the n-server system simulations which fall within the 95% confidence interval estimated by the diffusion approximation) can be found in Tables 4.1, 4.2, and 4.3. As is seen in these results, the diffusion approximation

		L			
		2	3	4	5
k	1	95.1%	96.3%	97.7%	95.9%
	2	-	96.5%	95.3%	95.6%
	3	-	-	96.8%	97.5%
	4	-	-	-	97.1%

Table 4.1. Empty Queue Coverage Rate

		L			
		2	3	4	5
k	1	97.1%	100%	100%	100%
	2		94.9%	95.6%	100%
	3	-	-	96.7%	96.4%
	4	-	-	-	95.0%

Table 4.2. Large Queue Coverage Rate

		L			
		2	3	4	5
k	1	95.2%	94.8%	94.8%	95.4%
	2		94.7	92.9%	94.9%
	3	-	-	96.8%	95.1%
	4	-	-	-	94.8%

Table 4.3. Mean Queue Length Coverage Rate

based confidence intervals, in general, contain approximately 95% of the n-server simulated observations, as desired.

The goal of this paper was to develop reliable approximations of the n-server system that are much quicker to simulate. Table 4.4 presents the average time (in seconds) required to simulate one trial of the finite system (left) and diffusion approximation (right). As is seen from these tables, the time required to simulate the diffusion approximations is substantially smaller than for the underlying n-server jump-Markov process. In addition, increasing n will further increase the amount of time required to simulate the n-server system. Indeed, $n = 10,000$ is a small number compared to the size of typical data centers and server farms that have machines which number in the hundreds of thousands.

		L			
		2	3	4	5
k	1	22.6	23.8	25.4	19.2
	2	-	39.1	38.0	33.1
	3	-	-	44.5	45.5
	4	-	-	-	57.4

(a) Average Time for Finite System

		L			
		2	3	4	5
k	1	.29	.50	.79	.79
	2	-	2.4	3.7	4.6
	3	-	-	6.0	10.0
	4	-	-	-	16.3

(b) Average Time for Limit Diffusion

Table 4.4. Average Simulation Times

APPENDIX A: TIGHTNESS CRITERIA

In this appendix we collect several tightness criteria that are used in this dissertation.

A.1 Conditions [A] and [T_1] of (Joffe and Métivier, 1986)

For the sake of the reader's convenience we present Theorem 2.3.2 and Conditions [A] and [T_1] of (Joffe and Métivier, 1986) in a form that are used to prove Theorem 4 and Proposition 6. Let $\{M^n\}$ be a sequence of $\mathbb{R}^k$-valued processes which are RCLL (right continuous with left limit) square-integrable local martingales, defined on their own filtered probability space $\{(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n), \mathbb{P}^n)\}$. Consider the following two conditions for a sequence of k-dimensional RCLL processes $\{X_n\}$, with X_n defined on $(\Omega^n, \mathcal{F}^n, (\mathcal{F}_t^n), \mathbb{P}^n)$.

[A] For each $\varepsilon > 0, \eta > 0$ there exists a $\delta > 0$ and $n_0 \in \mathbb{N}$ with the property that for every family of stopping times $\{\tau_n\}_{n\in\mathbb{N}}$ (τ_n being an $\mathcal{F}^n$-stopping time on Ω^n) with $\tau_n \leq T - \delta$,

$$\sup_{n\geq n_0} \sup_{\theta\leq\delta} P^n\{\|X^n_{\tau_n} - X^n_{\tau_n+\theta}\| \geq \eta\} \leq \varepsilon.$$

[T_1] For every t in some dense subset of $[0,T]$, $\{X_t^n\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{R}^k$ valued random variables.

Theorem 15 (Theorem 2.3.2 of (Joffe and Métivier, 1986) originally in (Rebolledo, 1979))**.** *Let $\langle M^n\rangle \doteq \sum_{i=1}^k \langle M_i^n, M_i^n\rangle$ be the predictable quadratic variation process associated with the k-dimensional local martingale M^n. Then if the sequence $\{\langle M^n\rangle\}_{n\in\mathbb{N}}$ of $\mathbb{R}$-valued stochastic processes satisfies condition [A], the same condition holds for the sequence $\{M^n\}_{n\in\mathbb{N}}$ of $\mathbb{R}^k$-valued stochastic processes. Futhermore if $\{\langle M^n\rangle\}_{n\in\mathbb{N}}$ satisfies [T_1] then the same condition holds for $\{M^n\}_{n\in\mathbb{N}}$. In particular if $\{\langle M^n\rangle\}_{n\in\mathbb{N}}$ satisfies [A] and [T_1], the sequence $\{\{\langle M_i^n, M_i^n\rangle, i = 1,\ldots,k\}\}_{n\in\mathbb{N}}$ and $\{M^n\}_{n\in\mathbb{N}}$ are tight in $\mathcal{D}([0,T]:\mathbb{R}^k)$.*

A.2 Criterion for Tightness of Hilbert Space-Valued Random Variables

The following theorem gives sufficient conditions for tightness of a sequence of random variables taking values in a (possibly infinite-dimensional) Hilbert space. For a proof see Corollary 2.3.1 of (Kallianpur and Xiong, 1995).

Theorem 16. *Let $\mathbb{H}$ be a separable Hilbert Space with inner product $\langle \cdot, \cdot \rangle$ and complete orthonormal system $\{e_i\}_{i=1}^{\infty}$. Suppose $\{Y_n\}_{n\in\mathbb{N}}$ is a sequence of $\mathbb{H}$-valued random variables satisfying the following conditions:*

a) For each $N \in \mathbb{N}$, $\lim_{A\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(\max_{1\le i\le N} \langle Y_n, e_i\rangle^2 > A\right) = 0$

b) For every $\delta > 0$, $\lim_{N\to\infty} \sup_{n\in\mathbb{N}} \mathbb{P}\left(\sum_{j=N}^{\infty} \langle Y_n, e_j\rangle^2 > \delta\right) = 0$.

Then $\{Y_n\}_{n\in\mathbb{N}}$ is a tight sequence of $\mathbb{H}$-valued random variables.

A.3 Aldous-Kurtz Criterion for Tightness of RCLL Processes

The following theorem gives a criterion for tightness of a sequence of RCLL processes with values in a Polish space, see (Kurtz, 1981).

Theorem 17. *Let $\mathbb{S}$ be a Polish Space and $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence of $\mathbb{D}([0,T]:\mathbb{S})$-valued RCLL $\{\mathcal{F}_t^n\}$-adapted satisfying the following conditions:*

(T_1) *$\{Y_n(t)\}_{n\in\mathbb{N}}$ is tight for every t in a dense subset of $[0,T]$.*

(A) *For each $\varepsilon > 0$, $\eta > 0$ and $N \in [0, T-\varepsilon]$ there exists a $\delta > 0$ and n_0 with the property that for every collection of stopping times $(\tau_n)_{n\in\mathbb{N}}$ (τ_n being an $\mathcal{F}_t^n \doteq \sigma\{Y_n(s) : s \le t\}$-stopping time) with $\tau_n \le N$,*

$$\sup_{n\ge n_0} \sup_{0\le\theta\le\delta} \mathbb{P}\{d(Y_n(\tau_n+\theta), Y_n(\tau_n)) \ge \eta\} \le \varepsilon,$$

where $d(\cdot,\cdot)$ is the distance on $\mathbb{S}$.

Then $\{Y_n\}_{n\in\mathbb{N}}$ is tight in $\mathbb{D}([0,T]:\mathbb{S})$.

APPENDIX B: HILBERT-SCHMIDT AND TRACE CLASS OPERATORS

We collect here some elementary facts about trace class and Hilbert-Schmidt operators. We refer the reader to (Reed and Simon, 1980) for details. For a separable Hilbert space $\mathbb{H}$ (with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$), let $\mathcal{L}(\mathbb{H})$ be the collection of all bounded linear operators on $\mathbb{H}$. An operator $A \in \mathcal{L}(\mathbb{H})$ is called non-negative if $\langle u, Au \rangle \geq 0$ for all $u \in \mathbb{H}$. Such an operator is called trace class if for some CONS $\{e_i\}$ in $\mathbb{H}$, $\sum_i \langle Ae_i, e_i \rangle < \infty$ in which case the quantity is finite (and is the same) for every CONS $\{e_i\}$. An operator $A \in \mathcal{L}(\mathbb{H})$ is called Hilbert-Schmidt if there exists a CONS $\{e_i\}$ in $\mathbb{H}$ such that $\sum_j \langle Ae_j, Ae_j \rangle = \sum_j \|Ae_j\|^2 < \infty$. In that case, this quantity is the same for all CONS $\{e_i\}$ and its square root is called the Hilbert-Schmidt norm of A, denoted as $\|A\|_{\text{HS}}$. For a non-negative operator $A \in \mathcal{L}(\mathbb{H})$, there is a unique non-negative $B \in \mathcal{L}(\mathbb{H})$ referred to as the non-negative square root of A such that $B^2 = A$. If A is a trace class operator, then B is a Hilbert-Schmidt operator.

APPENDIX C: CYLINDRICAL BROWNIAN MOTION

In this appendix we present the definition of a cylindrical Brownian motion (CBM) and introduce the stochastic integral with respect to a CBM. A collection of continuous real stochastic processes $\{(W_t(h))_{0\le t\le T} : h \in \boldsymbol{\ell}_2\}$ given on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \{\mathcal{F}_t\})$ is called a $\boldsymbol{\ell}_2$-cylindrical Brownian motion if for every $h \in \boldsymbol{\ell}_2$, $(W_t(h))_{0\le t\le T}$ is a $\{\mathcal{F}_t\}$-Brownian motion with variance $\|h\|_2^2$ and for $h, k \in \boldsymbol{\ell}_2$

$$\langle W(h), W(k)\rangle_t = \langle h, k\rangle_2 t,\ 0 \le t \le T.$$

For a measurable map a from $[0,T]$ to the space of Hilbert-Schmidt operators from $\boldsymbol{\ell}_2$ to $\boldsymbol{\ell}_2$ such that $\int_0^T \|a(s)\|_{\mathrm{HS}}^2 ds < \infty$, we denote by $\int_0^t a(s)dW(s)$ the $\boldsymbol{\ell}_2$-valued martingale defined as the limit of

$$\sum_{i=1}^{n}\sum_{j=1}^{n} \phi_i \int_0^t \langle \phi_i, a(s)\phi_j\rangle_2 dW_s(\phi_j)$$

as $n \to \infty$ where $\{\phi_i\}_{i\in\mathbb{N}}$ is a complete orthonormal system (CONS) in $\boldsymbol{\ell}_2$. For the fact that the limit exists and is independent of the choice of CONS, we refer the reader to Chapter 4 of (Da Prato and Zabczyk, 2014).

BIBLIOGRAPHY

Anderson, D. and Kurtz, T. (2015). *Stochastic analysis of biochemical systems*, volume 1. Springer.

Antunes, N., Fricker, C., Robert, P., and Tibi, D. (2008). Stochastic networks with multiple stable points. *The Annals of Probability*, 36(1):255–278.

Atar, R. and Shifrin, M. (2014). An asymptotic optimality result for the multiclass queue with finite buffers in heavy traffic. *Stochastic Systems*, 4(2):556–603.

Bell, S. and Williams, R. (2001). Dynamic scheduling of a system with two parallel servers in heavy traffic with resource pooling: Asymptotic optimality of a threshold policy. *The Annals of Applied Probability*, 11(3):608–649.

Bensoussan, A. (2011). *Stochastic control by functional analysis methods*. Elsevier.

Billingsley, P. (1999). *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition. A Wiley-Interscience Publication.

Bonald, T., Borst, S., Hegde, N., and Proutière, A. (2004). Wireless data performance in multi-cell scenarios. *ACM SIGMETRICS Performance Evaluation Review*, 32(1):378–380.

Borkar, V. S. (1989). Optimal control of diffusion processes. In *Pitman Research Notes in Math., 203. 36 Borkar V.(2005):Controlled diffusion processes, Probability surveys*. Citeseer.

Bramson, M., Lu, Y., and Prabhakar, B. (2012). Asymptotic independence of queues under randomized load balancing. *Queueing Systems*, 71(3):247–292.

Budhiraja, A. and Friedlander, E. (2017). Diffusion approximations for load balancing mechanisms in cloud storage systems. *arXiv preprint arXiv:1706.09914*.

Budhiraja, A., Friedlander, E., et al. (2018). Diffusion approximations for controlled weakly interacting large finite state systems with simultaneous jumps. *The Annals of Applied Probability*, 28(1):204–249.

Budhiraja, A. and Ghosh, A. (2012). Controlled stochastic networks in heavy traffic: Convergence of value functions. *The Annals of Applied Probability*, pages 734–791.

Budhiraja, A., Ghosh, A., and Lee, C. (2011). Ergodic rate control problem for single class queueing networks. *SIAM Journal on Control and Optimization*, 49(4):1570–1606.

Chen, Z. and Huan, Z. (1997). On the continuity of themth root of a continuous nonnegative definite matrix-valued function. *Journal of Mathematical Analysis and Applications*, 209(1):60–66.

Da Prato, G. and Zabczyk, J. (2014). *Stochastic equations in infinite dimensions*. Cambridge university press.

Dai, J. and Lin, W. (2008). Asymptotic optimality of maximum pressure policies in stochastic processing networks. *The Annals of Applied Probability*, 18(6):2239–2299.

Eschenfeldt, P. and Gamarnik, D. (2015). Join the shortest queue with many servers. the heavy traffic asymptotics. *arXiv preprint arXiv:1502.00999*.

Ethier, S. and Kurtz, T. (2009). *Markov processes: Characterization and convergence*, volume 282. John Wiley & Sons.

Fleming, W. H. and Rishel, R. W. (1976). *Deterministic and stochastic optimal control*. Springer-Verlag, New York, NY.

Friedlander, E. (2018). Steady-state behavior of some load balancing mechanisms in cloud storage systems. *arXiv preprint arXiv:1801.02979*.

Ganesh, A. J., Kermarrec, A.-M., and Massoulié, L. (2003). Peer-to-peer membership management for gossip-based protocols. *IEEE transactions on computers*, 52(2):139–149.

Gibbens, R., Hunt, P., and Kelly, F. (1990). Bistability in communication networks. *Disorder in physical systems*, pages 113–128.

Graham, C. (2000). Chaoticity on path space for a queueing network with selection of the shortest queue among several. *Journal of Applied Probability*, 37(1):198–211.

Graham, C. and Robert, P. (2009). Interacting multi-class transmissions in large stochastic networks. *The Annals of Applied Probability*, 19(6):2334–2361.

Gupta, P. and Kumar, P. (2003). Towards an information theory of large networks: An achievable rate region. *IEEE Transactions on Information Theory*, 49(8):1877–1894.

Harrison, J. (1988). Brownian models of queueing networks with heterogeneous customer populations. In *Stochastic differential systems, stochastic control theory and applications*, pages 147–186. Springer.

Hunt, P. and Kurtz, T. (1994). Large loss networks. *Stochastic Processes and their Applications*, 53(2):363–378.

Ikeda, N. and Watanabe, S. (1989). *Stochastic differential equations and diffusion processes*, volume 24 of *North-Holland Mathematical Library*. North-Holland Publishing Co., Amsterdam, second edition.

Joffe, A. and Métivier, M. (1986). Weak convergence of sequences of semimartingales with applications to multitype branching processes. *Advances in Applied Probability*, pages 20–65.

Kallianpur, G. and Xiong, J. (1995). *Stochastic differential equations in infinite-dimensional spaces*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 26. Institute of Mathematical Statistics, Hayward, CA.

Kang, H.-W., Kurtz, T., and Popovic, L. (2014). Central limit theorems and diffusion approximations for multiscale Markov chain models. *The Annals of Applied Probability*, 24(2):721–759.

Krylov, N. V. (2008). *Controlled diffusion processes*, volume 14. Springer Science & Business Media.

Kurtz, T. (1970). Solutions of ordinary differential equations as limits of pure jump Markov processes. *Journal of applied Probability*, 7(1):49–58.

Kurtz, T. (1971). Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability*, 8(2):344–356.

Kurtz, T. (1980). Representations of Markov processes as multiparameter time changes. *The Annals of Probability*, pages 682–715.

Kurtz, T. (1981). *Approximation of population processes*. SIAM.

Kurtz, T. and Protter, P. (1991). Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, pages 1035–1070.

Kushner, H. (2013). *Heavy traffic analysis of controlled queueing and communication networks*, volume 47. Springer Science & Business Media.

Kushner, H. and Dupuis, P. (2013). *Numerical methods for stochastic control problems in continuous time*, volume 24. Springer Science & Business Media.

Li, B., Ramamoorthy, A., and Srikant, R. (2016). Mean-field-analysis of coding versus replication in cloud storage systems. In *IEEE INFOCOM*.

Lin, S. and Costello, D. (2004). *Error control coding*. Pearson Education India.

Métivier, M. (1982). *Semimartingales: A course on stochastic processes*, volume 2. Walter de Gruyter.

Mitzenmacher, M. (2001). The power of two choices in randomized load balancing. *IEEE Transactions on Parallel and Distributed Systems*, 12(10):1094–1104.

Mukherjee, D., Borst, S. C., and van Leeuwaarden, J. S. (2017). Asymptotically optimal load balancing topologies. *arXiv preprint arXiv:1707.05866*.

Mukherjee, D., Borst, S. C., van Leeuwaarden, J. S., and Whiting, P. A. (2016a). Asymptotic optimality of power-of-d load balancing in large-scale systems. *arXiv preprint arXiv:1612.00722*.

Mukherjee, D., Borst, S. C., Van Leeuwaarden, J. S., and Whiting, P. A. (2016b). Universality of load balancing schemes on the diffusion scale. *Journal of Applied Probability*, 53(4):1111–1124.

Protter, P. (2005). *Stochastic integration and differential equations*, volume 21 of *Stochastic Modelling and Applied Probability*. Springer-Verlag, Berlin. Second edition. Version 2.1, Corrected third printing.

Rebolledo, R. (1979). La méthode des martingales appliquée à l'étude de la convergence en loi de processus (1979).

Reed, M. and Simon, B. (1980). Functional analysis, vol. I.

Rudin, W. (1986). *Real and complex analysis (3rd)*. New York: McGraw-Hill Inc.

Shiga, T. and Tanaka, H. (1985). Central limit theorem for a system of markovian particles with mean field interactions. *Probability Theory and Related Fields*, 69(3):439–459.

Stolyar, A. (2015). Pull-based load distribution in large-scale heterogeneous service systems. *Queueing Systems*, 80(4):341–361.

Stroock, D. W. and Varadhan, S. S. (2007). *Multidimensional diffusion processes*. Springer.

Sznitman, A.-S. (1991). Topics in propagation of chaos. In *Ecole d'été de probabilités de Saint-Flour XIX1989*, pages 165–251. Springer.

van der Boor, M., Borst, S. C., van Leeuwaarden, J. S., and Mukherjee, D. (2017). Scalable load balancing in networked systems: Universality properties and stochastic coupling methods. *arXiv preprint arXiv:1712.08555*.

Vvedenskaya, N., Dobrushin, R., and Karpelevich, F. (1996). Queueing system with selection of the shortest of two queues: An asymptotic approach. *Problemy Peredachi Informatsii*, 32(1):20–34.

Whitt, W. (2002). *Stochastic-process limits: An introduction to stochastic-process limits and their application to queues*. Springer Science & Business Media.