Controllability and Stabilization of Kolmogorov Forward Equations for Robotic Swarms

by

Karthik Elamvazhuthi

A Dissertation Presented in Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy

Approved June 2019 by the
Graduate Supervisory Committee:

Spring Berman, Chair
Matthias Kawski
Hendrik Kuiper
Marc Mignolet
Matthew Peet

ARIZONA STATE UNIVERSITY

August 2019

ABSTRACT

Numerous works have addressed the control of multi-robot systems for coverage, mapping, navigation, and task allocation problems. In addition to classical *microscopic* approaches to multi-robot problems, which model the actions and decisions of individual robots, lately there has been a focus on *macroscopic* or *Eulerian* approaches. In these approaches, the population of robots is represented as a continuum that evolves according to a mean-field model, which is directly designed such that the corresponding robot control policies produce target collective behaviors.

This dissertation presents a control-theoretic analysis of three types of mean-field models proposed in the literature for modeling and control of large-scale multi-agent systems, including robotic swarms. These mean-field models are Kolmogorov forward equations of stochastic processes, and their analysis is motivated by the fact that as the number of agents tends to infinity, the empirical measure associated with the agents converges to the solution of these models. Hence, the problem of transporting a swarm of agents from one distribution to another can be posed as a control problem for the forward equation of the process that determines the time evolution of the swarm density.

First, this thesis considers the case in which the agents' states evolve on a finite state space according to a continuous-time Markov chain (CTMC), and the forward equation is an ordinary differential equation (ODE). Defining the agents' task transition rates as the control parameters, the finite-time controllability, asymptotic controllability, and stabilization of the forward equation are investigated. Second, the controllability and stabilization problem for systems of advection-diffusion-reaction partial differential equations (PDEs) is studied in the case where the control parameters include the agents' velocity as well as transition rates. Third, this thesis considers a controllability and optimal control problem for the forward equation in the more general case where the agent dynamics are given by a nonlinear discrete-time control system. Beyond these theoretical results, this thesis also

considers numerical optimal transport for control-affine systems. It is shown that finite-volume approximations of the associated PDEs lead to well-posed transport problems on graphs as long as the control system is controllable everywhere.

To Giants and their Shoulders

ACKNOWLEDGEMENTS

Firstly, I would like to thank my advisor, Dr. Spring Berman, for introducing me to the field of swarm robotics, her constant support throughout my time as a graduate student and allowing me the independence to pursue a wide range of research problems.

Secondly, I would like to thank Dr. Matthew Peet, who has played an important role for me, both for my education at ASU, and in being an inspiration as a control theorist. I would also like to thank Dr. Marc Mignolet for being on my committee and for helping me along critical points of during my graduate studies.

Special thanks to Dr. Hendrik Kuiper and Dr. Matthias Kawski. A large part of this dissertation is born out of a number of collaborations with them. Hank has been a constant source of guidance and confidence in all matters relating to PDEs throughout my last five years of research. The core of this dissertation would not have been possible without him. Dr. Kawski has played a major role in my mathematics education, from teaching me topology to introducing me to the wonderful world of differential geometry and geometric control theory. His passion for mathematics, and life in general, is infectious and inspiring.

An important collaboration during by dissertation research was with Dr. Piyush Grover of Mitsubishi Electric Research Labs (MERL), Cambridge, MA. My internship at MERL under Piyush's guidance, during the summer of 2016, gave me a completely new perspective on my research. I was exposed to a large number of interesting topics in engineering and mathematics during just the short period of three months, which was mostly due to his extraordinary range of technical knowledge and vision. The last two chapters of this dissertation are a result my work with him.

I would also like to thank my ex- and current ACS lab members, and friends: Sean Wilson, Ragesh Ramachandran, Zahi Kakish, Aniket Shirsat, Hamed Farivarnejad and Azadeh Dorouchi for the multiple collaborations, discussions, Friday sessions at Devils and con-

ferences that we have attended together. They made this entire journey a truly enjoyable one.

I would also like to thank my parents, Arunachalam Elam Vazhuthi and Kanaga Vazhuthi and my brother, Vignesh Elamvazhuthi, for their love and support. Lastly, I would like to thank my girlfriend and collaborator, Shiba Biswal, for enriching my life in every way, from our outdoor adventures to long discussions on Analysis. Her own Ph.D. journey has been a constant source of inspiration for mine.

This research was supported by National Science Foundation (NSF) Award No. CMMI-1436960, Office of Naval Research (ONR) Young Investigator Award N00014-16-1-2605, the Arizona State University Global Security Initiative, and Mitsubishi Electric Research Laboratories.

TABLE OF CONTENTS

Page

LIST OF FIGURES ix

CHAPTER

1 INTRODUCTION 1

1.1 Contribution 2

1.1.1 Controllability and Stabilization of Finite-Dimensional Forward Equations 2

1.1.2 Controllability and Stabilization of Partial Differential Equation Type Forward Equations 4

1.1.3 Controllability and Optimal Control of Discrete-time Nonlinear Systems to Target Measures 5

1.1.4 Computational Optimal Transport of Control-affine Systems 6

1.2 Literature Review 6

1.2.1 Finite-Dimensional Mean-Field Models 7

1.2.2 Infinite-Dimensional Mean-Field Models 19

2 CONTROLLABILITY AND STABILIZATION OF FINITE-DIMENSIONAL FORWARD EQUATIONS 27

2.1 Notation 29

2.2 Controllability of the Forward Equation of a CTMC 32

2.3 Stabilization of the Forward Equation of a CTMC 42

2.3.1 Stabilization of Distributions with Strongly Connected Supports using Open-loop Control 44

2.3.2 Stabilization of Distributions with Strongly Connected Supports using Linear Feedback Laws 45

CHAPTER Page

2.3.3 Stabilization of Probability Distributions with Disconnected Supports 52
2.4 Controllability and Stabilization of a Model for Herding a Swarm using a Leader 66
2.4.1 Controllability 67
2.4.2 Stabilization 72
3 CONTROLLABILITY AND STABILIZATION OF PARTIAL DIFFERENTIAL EQUATION TYPE FORWARD EQUATIONS 85
3.1 Notation 86
3.2 Controllability of an Advection-Diffusion Equation 91
3.2.1 Simulation 108
3.3 Controllability of a System of Advection-Diffusion-Reaction Equations 109
3.4 Stabilization of a System of Advection-Diffusion-Reaction Equations to Target Probability Densities 114
3.5 Weighted Hypoelliptic Laplacians and their Semigroups 121
3.6 Stabilization of a System of Hypoelliptic Reaction-Diffusion Equations to Target Probability Densities with Disconnected Supports 127
4 CONTROLLABILITY AND OPTIMAL CONTROL OF DISCRETE-TIME NONLINEAR SYSTEMS TO TARGET MEASURES 143
4.1 Notation 144
4.2 Controllability 148
4.3 Optimal Control 152
4.4 Numerical Optimization 154
4.5 Simulation Examples 155

CHAPTER Page

5 COMPUTATIONAL OPTIMAL TRANSPORT OF CONTROL-AFFINE SYSTEMS 159
5.1 Preliminaries 159
5.2 Problem Setup and Computational Approach 166
5.3 Construction of Approximate Feedback Control Laws 175
5.4 Numerical Implementation 176
5.5 Simulation Examples 177
6 CONCLUSION AND FUTURE WORK 183
REFERENCES 186

LIST OF FIGURES

Figure Page

1.1 Bidirected Graph with 3 Vertices, Representing Agent States. 8

2.1 (**Example 2.2.6**) Edges in $\mathscr{E}_0$ are Uncontrolled and Denoted by the Red Arrows. Controlled Edges in $\mathscr{E}_1$ are Denoted by the Green Arrows. 37

2.2 (**Example 2.2.8**) Edges in $\mathscr{E}_0$ are Denoted by the Red Arrows. Edges in $\mathscr{E}_1$ are Denoted by the Green Arrows. 37

2.3 Illustration of the Splitting of the Graph $\mathscr{G}$ in the Proof of Theorem 2.2.11. . 40

2.4 Six-vertex bidirected graph. 63

2.5 Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*. 64

2.6 Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*. 65

2.7 Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*. 65

2.8 Bidirected Graph With 4 Vertices, Representing Agent States. 80

2.9 Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*. 81

2.10 Snapshots at Three Times t of $N = 10^4$ Follower Agents Redistributing Over a 36-Vertex Graph During a Stochastic Simulation of the Closed-Loop System. 83

3.1 Simulated Agent Densities at Three Times t and the Underlying Scalar Field. 108

4.1 Solution of the Optimal Transport Problem at Several Times n for Unicycles in a Double-Gyre Flow Model 157

4.2 Solution of the Optimal Transport Problem at Several Times n for a Double-Integrator System 158

Figure Page

5.1 Analytical Solution of the Optimal Transport Problem for the Grushin Plane 179

5.2 Numerical Solution of the Optimal Transport Problem at Several Times for the Grushin Plane 180

5.3 Initial and Final Measures Shown on (x, y) Plane for Optimal Transport in the Unicycle Model 181

5.4 The Optimal Transport Solution of Unicycle Model Shown in the $x - y$ Plane. 182

Chapter 1

INTRODUCTION

There has been a significant amount of work on swarm robotic systems over the last two decades. A major challenge is to develop modeling and control techniques for these large-scale multi-robot systems that are scalable with the swarm population size (Brambilla *et al.*, 2013). One approach to address this issue, inspired by modeling methodologies used in the natural sciences such as fluid dynamics, statistical mechanics, and mathematical biology, is to treat the swarm as a continuum. The starting point of this approach is the *Kolmogorov forward equation* of a stochastic process, which describes the spatio-temporal evolution of the probability density associated with the process. For a finite number of agents that are each modeled using such a stochastic process, the state space of the forward equation, a linear dynamical system, is dependent on the number of agents N. On the other hand, in the limit as the number of agents tends to infinity, one can approximate the N-agent linear forward equation with a single, possibly nonlinear, forward equation with parameters that can be functions of the probability density. The resulting equation, known as the *mean-field model*, is defined on the set of probability densities that determine the probability of an agent being in a given state at a specific time. When the number of agents in the swarm is large, this approximation is valid if all agents follow the same control laws (i.e., the swarm is homogeneous) and the control laws of each agent are not dependent on other agents' identities, but only on the agent's own state or the local density of the swarm. This *identity-invariance* of the control laws implies that the dimension of the state space of the mean-field model depends on the dimension of the state space of a single agent, and hence is independent of the actual number of agents in the swarm. Therefore, the scalability of any controller design methodology that is based on mean-field

models is dependent on the number of admissible states of a single agent, rather than on the total number of agents in the swarm. While much work has been devoted to optimization-based computational tools that use mean-field models to synthesize control laws for large multi-agent systems, there has been very little investigation of fundamental properties of these models, such as solvability of control and problems of stabilization and estimation. Characterization of such system-theoretic properties are important because they enable an engineer to understand fundamental limitations on the ability to control such systems, and thus facilitate the effective design of multi-robot control laws. This dissertation makes significant contributions in these directions.

This chapter is organized as follows. In Section 1.1, we highlight the major contributions of this dissertation that are presented in Chapters 2-5. In Section 1.2, we present a detailed survey of the different types of mean-field models introduced in the literature on multi-robot systems and the application of these models to control and estimation problems for robotic swarms.

1.1 Contribution

The novel contributions of this work are summarized in this section.

1.1.1 Controllability and Stabilization of Finite-Dimensional Forward Equations

In Chapter 2, we provide several results on controllability and stabilizability properties of the Kolmogorov forward equation of a continuous-time Markov chain (CTMC) evolving on a finite state space, with the transition rates defined as the control parameters. First, we present a result on small-time local and global controllability of the system from and to strictly positive equilibrium distributions when the underlying graph is strongly connected. Then, we show that any target probability distribution can be reached asymptotically using time-varying control parameters. Second, we characterize all stationary distributions that

are stabilizable using time-independent control parameters. For bidirected graphs, we construct rational and polynomial density feedback laws that stabilize stationary distributions while satisfying the additional constraint that the feedback law takes zero value at equilibrium. Third, we extend our feedback stabilization results to stationary distributions that have a *strongly connected support.*

Then, we construct a class of density-feedback laws, i.e., control laws that are functions of the swarm population density, that achieve this stabilization of CTMCs to probability densities with disconnected supports. To execute these control laws, each agent only requires information on the population fraction of agents that are in its current state. Additionally, the control laws ensure that there are no state transitions by agents at equilibrium, which is a known drawback of stabilization using time- and density-independent control laws. We guarantee global asymptotic stability of the equilibrium distribution by analyzing the corresponding mean-field model. To admit feedback laws that take values only on a discrete set, we consider control laws that can be discontinuous functions of the agent densities. We validate the control laws using stochastic simulations of the CTMC model and numerical simulations of the mean-field model.

Lastly, we introduce a control model for herding a swarm of "follower" agents to a target distribution among a set of states using a single "leader" agent. The follower agents evolve on a finite state space that is represented by a graph and transition between states according to a CTMC, whose transition rates are determined by the location of the leader agent and the distribution of followers on the graph. The control problem is to define a sequence of states for the leader agent that steers the probability density of the forward equation of the Markov chain. For the case with inter-follower interactions, we prove approximate local controllability of the system about equilibrium configurations. If the followers are non-interacting, they exit to neighboring states with equal positive probabilities if the leader is present in their current state. For this case, we design two switching

control laws for the leader that drive the swarm of follower agents asymptotically to a target probability distribution that is positive for all states. The first strategy is open-loop in nature, and the switching times of the leader are independent of the follower distribution. The second strategy is of feedback type, and the switching times of the leader are functions of the follower density in the leader's current state. We validate our control approach using numerical simulations with varied numbers of follower agents that evolve on graphs of different sizes.

This chapter includes results from (**Elamvazhuthi** *et al.*, 2017a, 2018a).

1.1.2 Controllability and Stabilization of Partial Differential Equation Type Forward Equations

In Chapter 3, we investigate the exact controllability properties of an advection-diffusion equation on a bounded domain, using time- and space-dependent velocity fields as the control parameters. This partial differential equation (PDE) is the Kolmogorov forward equation for a reflected diffusion process that models the spatiotemporal evolution of a swarm of agents. We prove that if a target probability density has bounded first-order weak derivatives and is uniformly bounded from below by a positive constant, then it can be reached in infinite time using control inputs that are bounded in space and time. We then extend this controllability result to a class of advection-diffusion-reaction PDEs that corresponds to a hybrid-switching diffusion process (HSDP), in which case the reaction parameters are additionally incorporated as the control inputs. For the HSDP, we first constructively prove controllability of the associated CTMC system, in which the state space is finite. Then we show that our controllability results for the advection-diffusion equation and the CTMC can be combined to establish controllability of the forward equation of the HSDP. Third, we provide constructive solutions to the problem of asymptotically stabilizing an HSDP to a target non-negative stationary distribution using time-independent state feedback laws,

which correspond to spatially-dependent coefficients of the associated system of PDEs. Fourth, we consider a semilinear PDE model which is the closed-loop system for a HSDP with a mean-field feedback law that stabilizes the swarm to probability densities with disconnected supports. In the semilinear model, we relax the assumption made in earlier sections that the generator of the stochastic process is elliptic, and also consider processes associated with a class of hypoelliptic operators.

This chapter includes results from (**Elamvazhuthi** *et al.*, 2016, 2017b; **Elamvazhuthi** and Berman, 2018; **Elamvazhuthi** *et al.*, 2019).

1.1.3 Controllability and Optimal Control of Discrete-time Nonlinear Systems to Target Measures

Chapter 4 considers the relaxed version of the *transport problem* for general nonlinear control systems, where the objective is to design time-varying feedback laws that transport a given initial probability measure to a target probability measure under the action of the closed-loop system. To make the problem analytically tractable, we consider control laws that are *stochastic*, i.e., the control laws are maps from the state space of the control system to the space of probability measures on the set of admissible control inputs. Under some controllability assumptions on the control system as defined on the state space, we show that the transport problem, considered as a controllability problem for the lifted control system on the space of probability measures, is well-posed for a large class of initial and target measures. We use this to prove the well-posedness of a fixed-endpoint optimal control problem defined on the space of probability measures, where along with the terminal constraints, the goal is to optimize an objective functional along the trajectory of the control system. This optimization problem can be posed as an infinite-dimensional linear programming problem. This formulation facilitates numerical solutions of the transport problem for low-dimensional control systems, as we show in two numerical examples.

This chapter includes results from (**Elamvazhuthi** *et al.*, 2018b).

1.1.4 Computational Optimal Transport of Control-affine Systems

In Chapter 5, we numerically construct optimal control laws for steering a given initial distribution in phase space to a final distribution in prescribed finite time for the case of non-autonomous nonlinear control-affine systems, while minimizing a quadratic control cost. Toward this end, we introduce a Benamou-Brenier type fluid dynamics formulation on a graph, which is obtained from discretizing the space using gridding. This leads to a convex optimization problem despite the nonlinearity of the control problem. The well-posedness of the resulting numerical optimal control problem is shown to be a consequence of the graph being strongly connected, which in turn is shown to result from controllability of the underlying dynamical system.

This chapter includes results from (**Elamvazhuthi** and Grover, 2018).

1.2 Literature Review

In this section, we survey the application of mean-field models to different problems in swarm robotics such as coverage, task allocation, consensus, and distributed mapping. Many of these problems can be framed as problems of feedback stabilization or parameter identification for the corresponding mean-field model. There have been several surveys on swarm robotics (Brambilla *et al.*, 2013; Seeja *et al.*, 2018), multi-robot systems (Khamis *et al.*, 2015; Robin and Lacroix, 2016) and the broader field of multi-agent systems (Oh *et al.*, 2015); in this section, we limit our review to works that specifically use mean-field models to predict and control collective behaviors in robotic swarms. We note that the use of mean-field models in robotic swarm control has been previously discussed in the literature under different terminology, including macroscopic models (Agassounon *et al.*, 2004), Rate Equation models (Lerman *et al.*, 2006), and probabilistic swarm guidance (Açıkmeşe

and Bayard, 2015).

First, we describe finite-dimensional mean-field models in the form of ordinary differential equations and difference equations, in which case each agent has a finite number of states and the time variable is continuous or discrete. In the second section, we discuss infinite-dimensional mean-field models in the form of partial differential equations, for which the agents' state space is continuous and the time variable is continuous.

1.2.1 Finite-Dimensional Mean-Field Models

In this section, we introduce finite-dimensional mean-field models in which the time variable is continuous or discrete.

Continuous-time models

There are N autonomous agents whose states evolve in continuous time according to a Markov chain with a finite state space defined as the vertex set $\mathscr{V} = \{1,...,M\}$. For example, the vertices in $\mathscr{V}$ can represent a set of tasks that the agents must perform, or a set of spatial locations obtained by partitioning the agents' environment. The edge set $\mathscr{E} \subset \mathscr{V} \times \mathscr{V}$ defines the pairs of vertices between which the agents can transition. The directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ is assumed to be strongly connected. The agents' transition rules are determined by the control parameters $u_e : [0,\infty) \to \mathbb{R}_{\geq 0}$ for each $e \in \mathscr{E}$, and are known as the *transition rates* of the associated continuous-time Markov chain (CTMC). The state of each agent $i \in \{1,...,N\}$ at time t is defined by a stochastic process $X_i(t)$ that evolves on the state space $\mathscr{V}$ according to the conditional probabilities

$$\mathbb{P}(X_i(t+h) = T(e) | X_i(t) = S(e)) = u_e(t)h + o(h) \tag{1.1}$$

for each $e = (S(e), T(e)) \in \mathscr{E}$, where $S(e)$ and $T(e)$ denote the source and target vertices of the edge e, respectively. Here, $o(h)$ is the little-oh symbol and $\mathbb{P}$ is the underlying prob-

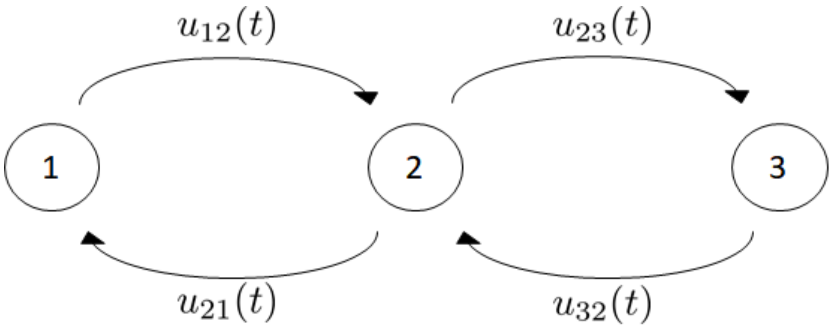


Figure 1.1: Bidirected Graph with 3 Vertices, Representing Agent States.

ability measure induced on the space of events Ω by the stochastic processes $\{X_i(t)\}_{i=1}^N$. Let $\mathscr{P}(\mathscr{V}) = \{y \in \mathbb{R}^M_{\geq 0};\ \sum_v y_v = 1\}$ be the simplex of probability densities on $\mathscr{V}$, and let int $\mathscr{P}(\mathscr{V})$ be the interior of this simplex. Corresponding to the CTMC is a system of ordinary differential equations (ODEs) that determines the time evolution of the probability densities $\mathbb{P}(X_i(t) = v) = x_v(t) \in \mathbb{R}_{\geq 0}$. If $X_i(0)$ are independent and identically distributed (IID), then the processes $\{X_i(t)\}_{i=1}^N$ are also IID, and the *Kolmogorov forward equation* can be represented by a single linear system of ODEs,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e \in \mathscr{E}} u_e(t) \mathbf{B}_e \mathbf{x}(t), \quad t \in [0, \infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{1.2}$$

where $\mathbf{x}^0$ represents the initial distribution of the random variables $X_i(0)$ and $\mathbf{B}_e \in \mathbb{R}^{M \times M}$ are control matrices whose entries at row i and column j are given by

$$B_e^{ij} = \begin{cases} -1 & \text{if } i = j = S(e), \\ 1 & \text{if } i = T(e),\ j = S(e), \\ 0 & \text{otherwise.} \end{cases}$$

For example, consider a 3-state Markov chain, for which the corresponding graph $\mathscr{G}$ is

illustrated in Fig. 1.1. The system of ODEs (1.2) in this case is given by:

$$\begin{aligned} \dot{x}_1(t) &= -u_{12}(t)x_1(t) + u_{21}(t)x_2(t) \\ \dot{x}_2(t) &= -(u_{21}(t) + u_{23}(t))x_2(t) + u_{12}(t)x_1(t) + u_{32}(t)x_3(t) \\ \dot{x}_3(t) &= -u_{32}(t)x_3(t) + u_{23}(t)x_2(t) \\ x_1(0) &= x_1^0, \;\; x_2(0) = x_2^0, \;\; x_3(0) = x_3^0. \end{aligned} \tag{1.3}$$

Let $\chi_v : \mathscr{V} \to \{0,1\}$ represent the indicator function of the vertex v. As $N \to \infty$, the population fraction of agents at a vertex v, given by $\frac{1}{N}\sum_{i=1}^{N} \chi_v(X_i(t))$, converges to $x_v(t)$ for each $t \in [0,\infty)$. This follows from the law of large numbers due to the random variables $X_i(t)$ being IID. Thus, instead of framing a control problem for the multi-agent system in terms of the random variables X_i, one can alternatively pose a control problem in terms of the deterministic quantity $\mathbf{x}(t)$, and hence control the *mean-field* behavior of the system. Therefore, control or estimation problems where the objectives are functions of the population fractions $\frac{1}{N}\sum_{i=1}^{N} \chi_v(X_i(t))$ can be replaced by problems where the objectives are functions of the probability distribution or *population density* $\mathbf{x}(t)$. An instance of this *mean-field control problem* is when the goal is to design the control inputs $u_e(t)$ such that $\mathbf{x}(T) = \mathbf{x}^d$ for a target distribution $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ and time $T > 0$. Another example of this type of control problem is the *mean-field stabilization problem*, where the goal is to design non-negative, possibly time-varying parameters k_e such that $u_e(t) = k_e$ for all $t \geq 0$ and a given $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ is an asymptotically stable equilibrium point of system (1.2). When the control inputs $u_e(t)$ are independent of time and the population density $\mathbf{x}(t)$, we will say that they are in *state-feedback form*. Here, the term *state-feedback* refers to the fact that agent i requires only knowledge of its current state $X_i(t)$ to execute the control action, and not the mean-field term $\mathbf{x}(t)$.

The following result is fundamental in analyzing the long-time behavior of Markov chains. It follows from the *Perron-Frobenius* theorem (Berman and Plemmons, 1994)

and plays an important role in the stabilization of the mean-field model (1.2) using time-independent state-feedback laws.

Theorem 1.2.1. *Suppose that $u_e(t) = k_e$ is a (time-independent) state-feedback law and is positive for each $e \in \mathscr{E}$. Then 0 is an eigenvalue of the matrix $\sum_{e\in\mathscr{E}} k_e \mathbf{B}_e$, and it has the largest real part of all the eigenvalues of this matrix. Moreover, this eigenvalue is simple. Hence, the solution $\mathbf{x}(t)$ of system (1.2) exponentially converges to a unique limit $\mathbf{x}^\infty \in \text{int } \mathscr{P}(\mathscr{V})$, which is a vector with all elements positive.*

Using the above theorem, the problem of designing state-feedback laws with the goal of achieving exponential stabilization with maximal decay rate was considered in (Berman *et al.*, 2009) for a multi-robot stochastic task allocation scenario. It was shown that this problem can be framed as a convex optimization problem. A drawback of using state-feedback laws is that the control inputs $u_e(t)$ remain non-zero at equilibrium and hence agents might continue switching between states at equilibrium; i.e., **the system being in** *macroscopic equilibrium* **does not imply that it is in** *microscopic equilibrium*. To reduce the frequency of switching at equilibrium, (Hsieh *et al.*, 2008) introduced control laws that are functions of the population density $\mathbf{x}(t)$. We will refer to such control laws as *mean-field feedback laws*. In particular, a mean-field feedback law is a family of functions $k_e : \mathscr{P}(\mathscr{V}) \to [0,\infty)$ such that the control inputs are defined as $u_e(t) = k_e(\mathbf{x}(t))$ for all $t \geq 0$ and all $e \in \mathscr{E}$. In (Hsieh *et al.*, 2008), the following mean-field feedback law k_e is considered,

$$k_e(\mathbf{x}) = k_e^* + \sigma_{S(e)}(x_{S(e)}, q_{S(e)})(\alpha - 1)k_e^*, \tag{1.4}$$

where for each $e \in \mathscr{E}$, $\sigma_{S(e)} = (1 + \exp[\gamma(q_{S(e)} - \frac{x_{S(e)}}{x^d_{S(e)}})])^{-1}$, and $q_{S(e)}$, γ, k_e^*, and α are suitably chosen parameters. It was shown in (Hsieh *et al.*, 2008) that for $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$, i.e. the set of probability distributions that are positive everywhere on $\mathscr{V}$, the solutions of system (1.2) converge to $\mathbf{x}^d$ as $t \to \infty$. Note that, when the control inputs $u_e(t)$ are

functions of the population fractions, which converge to the mean-field distribution $\mathbf{x}$ in the limit $N \to \infty$, the random variables X_i are not IID. Therefore, the validity of the limit $\lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} \chi_v(X_i(t)) = x_v(t)$ does not follow from the law of large numbers. Instead, one can apply the *dynamic law of large numbers*, which is proved in (Ethier and Kurtz, 2009).

Theorem 1.2.2. ***(Mean-field/Fluidic Limit) (Ethier and Kurtz, 2009)*** *Suppose that the transition rates $u_e(t)$ of each agent are given by*

$$u_e(t) = v_e\left(\frac{1}{N}\sum_{i=1}^{N}\chi_1(X_i(t)), ..., \frac{1}{N}\sum_{i=1}^{N}\chi_M(X_i(t))\right), \tag{1.5}$$

where $v_e : \mathscr{P}(\mathscr{V}) \to [0,\infty)$ is a Lipschitz-continuous function for each $e \in \mathscr{E}$. Consider the solution $\mathbf{x}(t)$ of the following system of ordinary differential equations,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e\in\mathscr{E}} v_e(x_1, ..., x_M)\mathbf{B}_e\mathbf{x}(t), \quad t \in [0,\infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}). \end{aligned} \tag{1.6}$$

Then for every $t \geq 0$,

$$\lim_{N\to\infty} \sup_{s\geq t} |\mathbf{Y}_N(s) - \mathbf{x}(s)| = 0 \quad \textit{almost surely} \tag{1.7}$$

where for each $s \geq 0$, the random variable $\mathbf{Y}_N(s)$ is given by

$$\mathbf{Y}_N(s) = \left[\frac{1}{N}\sum_{i=1}^{N}\chi_1(X_i(t)) \ ... \ \frac{1}{N}\sum_{i=1}^{N}\chi_M(X_i(t))\right]^T$$

and for each $\mathbf{y} \in \mathbb{R}^M$, $|\mathbf{y}| := \sum_{i=1}^{M} |y_i|$.

There has been an extensive amount of work on generalizing the above result to cases where the functions v_e are possibly discontinuous (Gast and Gaujal, 2012; Roth and Sandholm, 2013) or where the mean-field model is a hybrid system with continuous as well as discrete states (Bortolussi *et al.*, 2013).

Mean-field feedback laws require that agents can measure the population density $\mathbf{x}(t)$. For practical purposes, it is desirable that the mean-field feedback laws are *local*; that is, the control inputs u_e are functions of the population density at the source vertex $S(e)$, the target vertex $T(e)$, or both. The problem of reducing agent fluctuations at equilibrium is framed as a variance control problem in (Mather and Hsieh, 2014), using local mean-field feedback laws of the form $u_e(\mathbf{x}) = \alpha_e + \beta_e \frac{x_{S(e)}}{x_{T(e)}}$ for suitable choices of the parameters α_e and β_e.

Before one proceeds to design control laws, it is important to know which distributions are stabilizable. The works (Berman *et al.*, 2009; Hsieh *et al.*, 2008; Mather and Hsieh, 2014) require the assumption that $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$. When $\mathscr{G}$ is bidirected, it follows by construction from (Halász *et al.*, 2007) that, if $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$, then there exists a state-feedback law that asymptotically stabilizes $\mathbf{x}^d$. From Theorem 1.2.1, it can be seen that the assumption that $\mathscr{G}$ is bidirected can be relaxed in order for the stabilization result to still hold. Suppose that $\mathscr{G}$ is strongly connected, the parameters k_e are positive, and $\mathbf{x}^\infty$ is the unique (up to a scaling factor) eigenvector of the matrix $\sum_{e\in\mathscr{E}} k_e \mathbf{B}_e$ corresponding to 0. Then for the state-feedback law $\tilde{k}_e = k_e \frac{x^\infty_{S(e)}}{x^d_{S(e)}}$, we have that $\mathbf{x}^d$ is the unique eigenvector of the matrix $\sum_{e\in\mathscr{E}} \tilde{k}_e \mathbf{B}_e = \sum_{e\in\mathscr{E}} k_e \mathbf{B}_e \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix $diag(\frac{x^\infty_1}{x^d_1}, \frac{x^\infty_2}{x^d_2}, ..., \frac{x^\infty_M}{x^d_M})$. Thus, $\mathbf{x}^d$ is the globally asymptotically stable equilibrium point of system (1.2).

A method for computing optimal time-varying state-feedback laws in order to achieve a target distribution in finite time is shown in the work (Solomon *et al.*, 2016) on computational optimal transport. For certain cost functions, this optimal control problem can be treated in a convex optimization framework. For example, for a given $T > 0$ and $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$, consider the following optimization problem:

$$\inf_{u_e(t)\geq 0, x_v \geq 0} \sum_{e\in\mathscr{E}} \int_0^T u_e^2(t) x_{S(e)}(t) dt \tag{1.8}$$

subject to the bilinear constraints defined by system (1.2), with

$$\mathbf{x}(T) = \mathbf{x}^d. \tag{1.9}$$

This optimization problem is non-convex. However, it can be transformed into the following equivalent convex optimization problem:

$$\inf_{r_e(t)\geq 0, x_v(t)\geq 0} \sum_{e\in\mathscr{E}} \int_0^T \frac{r_e^2(t)}{x_{S(e)}(t)} dt \tag{1.10}$$

subject to the linear constraints

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e\in\mathscr{E}} r_e(t)\mathbf{B}_e\mathbf{1}, \quad t\in[0,\infty), \\ \mathbf{x}(0) &= \mathbf{x}^0, \ \mathbf{x}(T) = \mathbf{x}^d, \end{aligned} \tag{1.11}$$

where $\mathbf{1}\in\mathbb{R}^M$ is the vector with all elements equal to 1. This approach of convexifying optimization problems with objective functions such as the one in (1.8) and constraints (1.2), (1.9) was introduced in (Solomon *et al.*, 2016) in order to adapt the fluid-dynamic version of the optimal transport problem (Benamou and Brenier, 2000), where the state space is continuous, to the case of discrete state spaces. See Section 1.2.2 for more details. We note that the cost function in (1.8) has a simpler structure than the one considered in (Solomon *et al.*, 2016).

Numerical construction of mean-field feedback laws is a much more computationally challenging task, in comparison with the synthesis of state-feedback laws. Computational approaches based on Linear Matrix Inequalities (Boyd *et al.*, 1994) and Sum-of-squares methods (Chesi, 2011) are used to numerically construct decentralized mean-field feedback laws in (Deshmukh *et al.*, 2018). Execution of mean-field feedback strategies requires knowledge of the distribution of robots in each state. One approach to estimate the robot distribution is to use a centralized observer, such as an overhead camera (Deshmukh *et al.*, 2018). An alternative approach, which does not rely on a centralized authority to observe

the swarm, is to use encounter rates between agents to estimate population densities, as observed in natural swarms such as ant colonies (Pratt, 2005). A model for estimating population densities of swarms as a function of inter-agent encounter rates is proposed and experimentally validated in (Mayya *et al.*, 2019).

The work (Prorok *et al.*, 2017) considers the effect of heterogeneity in the robot populations on the optimal robot control policies. In this work, $\mathscr{V}$ denotes not only the states that robots can occupy, but also the types of different robots. The problem of identifying the minimum number of robots of each type in order to achieve a given goal is framed as an optimization problem.

In some scenarios, it is useful to consider mean-field models where different types of agents or agents in different states interact at particular probability rates and then physically bond or change their states. Such models are commonly used to describe the dynamics of chemical reaction networks (CRNs), and have been adopted in several works in swarm robotics. A CRN model of a swarm represents agents of different types or in different states as distinct *species* that are analogous to chemical species. A *reaction* occurs when a combination of *reactant* species converts into a combination of *product* species at a certain *reaction rate constant*. Suppose that a reaction r in a CRN has reactants $a_i \in \mathbb{R}_{>0}$, $i = 1,...,n$, that combine with probability $k_r(t)\Delta T$ in an infinitesimally small amount of time ΔT to form products $b_j \in \mathbb{R}_{>0}$, $j = 1,...,m$. Here, $k_r(t)$ is the reaction rate constant. We denote this reaction by $r = [(a_1,...,a_n),(b_1,...,b_m)]$. Let M be the total number of reactant and product species in the entire CRN; then the vector of agent population densities in each species is given by $\mathbf{x} \in \mathbb{R}^M$. Define a vector field $f_r : \mathbb{R}^M \to \mathbb{R}^M$ associated with reaction r that has entries $(f_r(\mathbf{x}))_{a_i} = -\prod_{i=1}^{n} x_{a_i}$ for $i \in \{1,...,n\}$, $(f_r(\mathbf{x}))_{b_j} = \prod_{i=1}^{n} x_{a_i}$ for $j \in \{1,...,m\}$, and 0 otherwise. Then the resulting mean-field model can be written as

follows, where $\mathscr{R}$ is the set of all reactions in the CRN:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{r\in\mathscr{R}} k_r(t) f_r(\mathbf{x}(t)), \quad t \in [0,\infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathbb{R}^M_{\geq 0}, \end{aligned} \tag{1.12}$$

The system of equations (1.12) simplifies to the form of system (1.2) when only *unimolecular reactions* are admissible; i.e, all reactions in the CRN are of the form $r = [a,b]$, where $a,b \in \mathscr{V}$.

The first application of this type of mean-field model to simulating the behavior of a robotic swarm was in (Lerman *et al.*, 2001), which introduced a CRN model for a stick-pulling experiment performed by a swarm of robots that do not explicitly communicate or coordinate with one another. Using the mean-field model, the authors identify optimal state-dependent control parameters to improve the system's performance. In (Lerman *et al.*, 2004), the authors study the application of these types of models to a number of tasks performed by a swarm of robots, including collaborative pulling, foraging, and aggregation. In (Lerman and Galstyan, 2002), the authors use a mean-field model to study the effect of spatial interference on the performance of robots in a collective foraging task. A mean-field model based on a CRN is used in (Matthey *et al.*, 2009) for a task in which a swarm of robots must assemble a collection of parts into target amounts of final products using stochastic control policies determined by the reaction rate constants. The authors optimize the reaction rate constants to improve the system's rate of convergence to the target numbers of products. In (Wilson *et al.*, 2014), the authors use a CRN-based mean-field model to design stochastic robot attachment-detachment policies that drive a swarm to specified spatial distributions around multiple payloads for a collective transport task. A CRN is used to model a stochastic self-assembly task in (Haghighat *et al.*, 2017), and methods are developed to estimate the reaction rates in the CRN model using high-fidelity physics-based simulations. In (Klavins *et al.*, 2006), the authors present an optimization-based method

to maximize the yield of a stochastic self-assembly process by finding the optimal reaction rates, and validate the method using a CRN model. These authors also introduce an integral feedback controller in (Napp *et al.*, 2011) to stabilize a CRN model of another stochastic self-assembly process. In (Mermoud *et al.*, 2010), the authors develop a CRN model of a scenario in which robots collaboratively screen an environment for undesirable agents, and use this model to find the optimal parameters to achieve the goal.

CRN models have also been used extensively to model collective decision-making problems in swarm robotics, where a group of robots must collectively decide among a number of available options using limited information and interactions. Collective decision-making leads to stabilization problems that differ from classical formulations: the target probability distribution to which the agents should stabilize is not predefined by a centralized authority, and this distribution is a non-local function of the states or the agent populations in the states, while the robot control laws are constrained to be local. In (de Oca *et al.*, 2011), the authors consider a modified form of the majority rule opinion dynamics, studied in the literature on opinion dynamics (Krapivsky and Redner, 2003), for a scenario where a swarm of robots must decide between two different actions with different execution times, but without any prior knowledge of the execution times. Similarly, CRN models that have been used to explain honeybee nest site selection strategies have found applications in swarm robotics (Reina *et al.*, 2015). See (Valentini, 2017; Valentini *et al.*, 2017) for extensive surveys on the topic of collective decision-making problems in swarm robotics with some applications of mean-field models. In (Albani *et al.*, 2018), CRN models are used to design unmanned aerial vehicle control policies for non-uniform spatial coverage. In this work, the states represent spatial sites as well as tasks.

Other recent work that uses a CRN-based mean-field framework for swarm applications considers the problem of keeping individual robot types private (Prorok and Kumar, 2016). A *privacy model* that uses notions from differential privacy is developed to understand the

privacy preservation capabilities of the swarm as a function of the reaction parameters.

Discrete-time models

In discrete-time mean-field models, the state of each agent $i \in \{1,...,N\}$ is defined by a discrete-time Markov chain (DTMC) $X_i(n)$, $n \in \mathbb{Z}_+$, that evolves on the state space $\mathscr{V}$ according to the conditional probabilities

$$\mathbb{P}(X_i(n+1) = T(e) | X_i(n) = S(e)) = u_e(n) \tag{1.13}$$

with control parameters $u_e(n) \in [0,1]$ that satisfy the constraint

$$\sum_{e \in \mathscr{E}, S(e)=v} u_e(n) = 1 \tag{1.14}$$

for all $v \in \mathscr{V}$ and all $n \in \mathbb{Z}_+$. The parameters $u_e(n)$ are the *transition probabilities* that are associated with each edge $\mathscr{E}$. The probability distribution $\mathbf{x}(n) \in \mathbb{R}^M$ of the DTMC $X_i(n)$, given by $\mathbb{P}(X_i(n) = v) = x_v(n) \in \mathbb{R}_{\geq 0}$ for all $v \in \mathscr{V}$, evolves according to the mean-field model

$$\begin{aligned} \mathbf{x}(n+1) &= \sum_{e \in \mathscr{E}} u_e(n) \mathbf{B}_e \mathbf{x}(n), \quad n \in \mathbb{Z}_+, \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{1.15}$$

where the entries of $\mathbf{B}_e \in \mathbb{R}^{M \times M}$ are given by

$$B_e^{ij} = \begin{cases} 1 & \text{if } i = T(e),\ j = S(e), \\ 0 & \text{otherwise.} \end{cases}$$

The above model is the discrete-time analogue of model (1.2). The problem of stabilizing the solution $\mathbf{x}(n)$ of the system (1.15) for swarm models was first considered in (Chattopadhyay and Ray, 2009). In this work, the authors develop an iterative scheme to construct a (time-independent) state feedback law u_e such that $\lim_{n \to \infty} \mathbf{x}(n) = \mathbf{x}^{eq}$, where $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$ is a target stationary probability distribution.

In (Açıkmeşe and Bayard, 2015), the authors investigate general conditions on the graph $\mathscr{G}$ under which time-independent state feedback laws $u_e \geq 0$ can be designed such that the solution of the system (1.15) converges to a given stationary distribution $\mathbf{x}^{eq}$. The authors construct a DTMC using a variant of the Metropolis-Hastings algorithm (Chib and Greenberg, 1995) and show that if the vector $\mathbf{x}^{eq}$ has a strongly connected support and the graph $\mathscr{G}$ is symmetric, then one can find parameters $u_e \geq 0$ such that this stabilization problem can be solved. The authors also provide a Linear Matrix Inequality based method for computing the parameters u_e such that a target $\mathbf{x}^{eq}$ is exponentially stable with a given decay rate. The following theorem is the discrete-time version of Theorem 1.2.1, and it provides a theoretical foundation for the results proved in (Açıkmeşe and Bayard, 2015).

Theorem 1.2.3. *Suppose that the transition probabilities u_e are positive and constant. Additionally, suppose that there exists a time $n \in \mathbb{Z}^+$ such that, for each $v, w \in \mathscr{V}$, there exists a directed path of length n from v to w. Then* 1 *is the eigenvalue of the matrix $\sum_{e \in \mathscr{E}} u_e \mathbf{B}_e$ with the largest modulus. Moreover, this eigenvalue is simple. Hence, the solution $\mathbf{x}(t)$ of system* (1.15) *exponentially converges to a unique limit $\mathbf{x}^{\infty} \in \text{int}\, \mathscr{P}(\mathscr{V})$ for which all the elements are positive.*

A drawback of using time-independent state-feedback laws is that, as for the case of CTMCs, the agents do not stop transitioning between states once the mean-field model (1.15) reaches equilibrium. In order to resolve this issue, the authors in (Bandyopadhyay *et al.*, 2017) consider the problem of constructing time-varying parameters $u_e(n)$ such that $\lim_{n \to \infty} u_e(n) = 1$ for all $e = (v, v) \in \mathscr{E}$, $v \in \mathscr{V}$. This problem is framed as a linear programming problem that each agent i must solve in order to compute its own optimal transition probabilities $u_e^i(n)$ at each time n so that the swarm reaches the target distribution while minimizing a particular objective functional. Strictly speaking, this linear programming approach is not a mean-field approach, since the problem is formulated for a finite number

of agents and it is not clear whether the transition probabilities $u_e(n)$ have well-defined limits as $N \to \infty$. The state-feedback laws constructed in (Bandyopadhyay *et al.*, 2017) depend on the distance of the swarm from the target distribution, and hence require global knowledge of the swarm distribution at each time n. This requirement is then relaxed by implementing a filtering algorithm that each agent uses to estimate the distribution of the swarm over all the states from local measurements of the agent distribution in its current state.

In (El Chamie *et al.*, 2019), the authors address a swarm stabilization problem in which the control laws must satisfy certain density constraints on the solution of the mean-field model. The authors adapt classical Markov decision process (MDP) theory (Puterman, 2014) to construct stochastic or *randomized* state-feedback laws with constraints on the probability distribution of the stochastic process that models agent motion, such as constraints on robot densities.

1.2.2 Infinite-Dimensional Mean-Field Models

In this section, we describe infinite-dimensional mean-field models in which the time variable is continuous. We start with the case where the state space Ω of each agent, indexed by $i \in \{1,2,...,N\}$, is a subset of the Euclidean space $\mathbb{R}^n$. The position of each agent i evolves according to a stochastic process $\mathbf{Z}_i(t) \in \Omega$, where t denotes time. We initially assume that the agents are non-interacting. Therefore, the random variables $\mathbf{Z}_i(t)$ are independent and identically distributed, and we can drop the subscript i and define the problem in terms of a single stochastic process $\mathbf{Z}(t) \in \Omega$. The deterministic motion of each agent is defined by a velocity vector field $\mathbf{v}(\mathbf{x},t) \in \mathbb{R}^n$, where $\mathbf{x} \in \Omega$. This motion is perturbed by an n-dimensional *Wiener process* $\mathbf{W}(t)$, which models noise. This process can be a model for stochasticity arising from inherent sensor and actuator noise. Alternatively, noise could be actively programmed into the agents' motion to implement more exploratory

agent behaviors and to take advantage of the smoothening effect of the process on the agents' probability densities. Given the velocity field $\mathbf{v}(\mathbf{x},t)$ and a diffusion coefficient $D>0$, the position of each agent evolves according to a *diffusion process* $\mathbf{Z}(t)$ that satisfies the following stochastic differential equation (SDE) (Gardiner, 2009):

$$\begin{aligned} d\mathbf{Z}(t) &= \mathbf{v}(\mathbf{Z}(t),t)dt + \sqrt{2D}d\mathbf{W}(t), \\ \mathbf{Z}(0) &= \mathbf{Z}_0. \end{aligned} \tag{1.16}$$

Given a final time $T>0$, the *Kolmogorov forward equation* corresponding to the SDE (1.16) is given by:

$$\begin{aligned} y_t = D\Delta y - \nabla\cdot(\mathbf{v}(\mathbf{x}(t),t)y) \qquad in\ \ \Omega\times[0,T], \\ y(\cdot,0) = y^0 \quad in\ \ \Omega. \end{aligned} \tag{1.17}$$

The solution $y(\mathbf{x},t)$ of this equation represents the probability density of a single agent occupying position $\mathbf{x}\in\Omega$ at time t, or alternatively, the density of a population of agents at this position and time. The PDE (1.17) is related to the SDE (1.16) through the relation $\mathbb{P}(\mathbf{Z}(t)\in\Gamma) = \int_\Gamma y(\mathbf{x},t)d\mathbf{x}$ for all $t\in[0,T]$ and all measurable $\Gamma\subset\Omega$.In Prorok *et al.* (2011), the authors use the model (3.5) to simulate a swarm of miniature robots performing an inspection task, and validate the model experimentally. In Kingston and Egerstedt (2011), the authors construct state-feedback laws $\mathbf{v}$ that are piecewise constant with respect to space for the model (3.5) with $D=0$, using the Helmholtz-Hodge decomposition of a vector field.

The work Mesquita *et al.* (2008) considers a PDE model of the form

$$\begin{aligned} y_t(\mathbf{x},\mathbf{v}) = \ & -\mathbf{v}\cdot\nabla_{\mathbf{x}}\cdot(y(\mathbf{x},\mathbf{v})) - \lambda(\mathbf{x},\mathbf{v})y((\mathbf{x},\mathbf{v})) \\ & + \int T_{\mathbf{v}'}(v,v')\lambda(\mathbf{x},\mathbf{v})y(\mathbf{x},\mathbf{v}',t)d\mathbf{v}', \end{aligned} \tag{1.18}$$

where $\mathbf{x}$ denotes the position coordinates and $\mathbf{v}$ denotes the velocity coordinates. The parameter λ denotes the rate at which a robot jumps to a random value of $\mathbf{v}$ according to

the parameter $T_{\mathbf{v}'}$, a function known as the *jump pdf*. The authors design suitable λ and $T_{\mathbf{v}'}$ such that the robots converge to a target probability density that is positive everywhere. This result is generalized to a larger class of controllable nonlinear systems in Mesquita and Hespanha (2012).

There have been a number of works on numerical construction of state-feedback laws for a swarm of agents that follow the dynamics (3.4). In (Foderaro *et al.*, 2014), the authors consider the problem of designing a time-varying, state-dependent velocity $u_1(\mathbf{x},t)$ and turning rate $u_2(\mathbf{x},t)$ with the vector field $\mathbf{v}$ in (1.17) given by

$$\mathbf{v}(\mathbf{x},t) = \begin{bmatrix} u_1(\mathbf{x},t)\cos(x_1) \\ u_1(\mathbf{x},t)\sin(x_2) \\ u_2(\mathbf{x},t) \end{bmatrix}.$$

The authors use optimal control to compute the control inputs $u_1(\mathbf{x},t)$ and $u_2(\mathbf{x},t)$ that transport a swarm from an initial probability density to a target density. The optimal control of PDEs that govern stochastic processes has received considerable attention in the mathematics literature (Annunziato and Borzì, 2010, 2013; Annunziato and Borzi, 2018; Fleig and Guglielmi, 2017). Similar optimal control problems have also been investigated in the mathematics and control theory literature on *mean-field games* (Lasry and Lions, 2007; Huang *et al.*, 2007; Bensoussan *et al.*, 2013; Caines *et al.*, 2017; Carmona and Delarue, 2018). The application of mean-field games to swarm robotics problems has begun only recently (Liu *et al.*, 2018). A promising approach to numerically constructing state-feedback laws comes from optimal transport theory. While this approach has thus far not been applied to control swarms of robots, we mention it here due to its applicability in this domain. Consider the following optimization problem:

$$\inf_{\mathbf{v}} \int_0^T \int_\Omega |\mathbf{v}(\mathbf{x},t)|^2 y(\mathbf{x},t) d\mathbf{x} dt \tag{1.19}$$

subject to the constraints

$$y_t = -\nabla \cdot (\mathbf{v}(\mathbf{x},t)y),$$
$$y(0) = y^0, \;\; y(T) = y^d, \tag{1.20}$$

where y^0 and y^d are the initial and target probability densities, respectively. The optimization problem (1.19)-(1.20) was introduced to develop a computationally tractable approach to calculating the 2-Wasserstein distance (Villani, 2008). In swarm robotics applications, this can be viewed as an optimal control problem that computes a state-feedback law $\mathbf{v}(x,t)$ which drives a swarm from an initial probability density y^0 to a target probability density y^d in time T. However, this optimization problem is non-convex in the decision variables $\mathbf{v}$ and ρ. If we perform the change of variable $\mathbf{m} = \frac{\mathbf{v}}{\rho}$, we can instead consider the equivalent convex optimization problem,

$$\inf_{\mathbf{m},\rho \geq 0} \int_0^1 \int_\Omega \frac{|\mathbf{m}(\mathbf{x},t)|^2}{y(\mathbf{x},t)} d\mathbf{x}dt \tag{1.21}$$

subject to the constraints

$$y_t = -\nabla \cdot (\mathbf{m}(\mathbf{x},t)),$$
$$y(0) = y^0, \;\; y(1) = y^d. \tag{1.22}$$

Due to this convexification, one can guarantee that any locally optimal solution of the optimization problem (1.21)-(1.22) is also globally optimal. This offers an advantage over objective functionals that are more commonly used in optimal control of PDEs (Tröltzsch, 2010), for which global optimality of locally optimal solutions is much more difficult to guarantee.

In (**Elamvazhuthi** *et al.*, 2016) considers the problem of stabilizing the PDE (1.17) to a target probability density y_∞. It is shown that if the diffusion coefficient is defined as the spatially-dependent function $c/\sqrt{y_\infty}$ for any positive constant c, then the solution of

the PDE converges to ρ_∞. The effectiveness of this control law is experimentally verified with robot experiments in (Li *et al.*, 2017). This strategy is extended to the case where agents evolve on compact manifolds in (**Elamvazhuthi** and Berman, 2018). An alternative approach to stabilize a swarm to a target distribution is to set D to a positive constant and $\mathbf{v} = D\frac{\nabla\rho_\infty}{\rho_\infty}$, which also results in the solution converging to ρ_∞ (Breiten *et al.*, 2018; **Elamvazhuthi** *et al.*, 2019). The long-time behavior of SDEs with gradient drift has been extensively treated in the mathematics and physics literature (Stroock, 1993; Markowich and Villani, 1999; Ambrosio *et al.*, 2009). In applications beyond swarm robotics, the problem of controlling the PDE (1.17) to a target probability density using a time-dependent state-feedback law $\mathbf{v}(\mathbf{x},t)$ has been investigated in optimal transport theory (Benamou and Brenier, 2000) and stochastic control (Blaquiere, 1992) for the case where $\Omega = \mathbb{R}^n$, and in the theory of mean-field games (Porretta, 2014) when Ω is a torus.

While models of the form (1.17), with control parameters that are functions of the swarm density, have been extensively analyzed in the mathematics literature (Bodnar and Velazquez, 2005; Topaz *et al.*, 2006; Bertozzi *et al.*, 2011; Carrillo *et al.*, 2014, 2010), there has been very little work on using such models to construct mean-field feedback laws for stabilization of robotic swarms. In (Kingston and Egerstedt, 2010), the authors design mean-field feedback laws where the vector field $\mathbf{v}$ in (1.17) is set to a suitable integral functional of the density so that the agents achieve consensus. A similar approach for the analysis of consensus in swarms is also considered in (Canuto *et al.*, 2008). In (Eren and Açıkmeşe, 2017), the authors construct a mean-field feedback law by interpreting the linear heat equation as a nonlinear advection equation with a density-dependent velocity field as follows. The diffusion coefficient D is set to zero, and the control law is defined as $\mathbf{v}(\mathbf{x},t) = -\frac{\nabla e(\mathbf{x},t)}{y(\mathbf{x},t)}$ for all $\mathbf{x} \in \Omega$ and all $t \geq 0$, where $e(\mathbf{x},t) = y(\mathbf{x},t) - y^d(\mathbf{x})$ and y^d is the

target probability density. Then model (1.17) becomes

$$\begin{aligned} e_t &= \Delta e \qquad in\ \ \Omega \times [0,T], \\ e(\cdot,0) &= e^0 \qquad in\ \ \Omega. \end{aligned} \tag{1.23}$$

Using the relation between models (1.17) and (1.23), one can show that the swarm density $y(\cdot,t)$ converges to the target probability density y^d as $t \to \infty$.

While the analysis of the closed-loop system (1.23) is straightforward due to its linearity, the solutions of these PDEs make sense only for initial conditions that are positive everywhere on Ω; otherwise, the control law $\mathbf{v}$ is unbounded. An alternative is to set $\mathbf{v}(\mathbf{x},t) = -b(\mathbf{x})\frac{\nabla y(\mathbf{x},t)}{y^d(\mathbf{x},t)}$, where $b(\mathbf{x})$ is a positive function. The resulting closed-loop system is a weighted variation of a well-known nonlinear PDE called the porous media equation (Vázquez, 2007). According to results established in the mathematics literature (Grillo *et al.*, 2013), it is known that under particular technical assumptions on $b(\mathbf{x})$ and $y^d(\mathbf{x})$, the swarm density $y(\cdot,t)$ converges to the target probability density y^d as $t \to \infty$. These types of control laws are used for stabilizing swarms to target probability densities in the recent works (**Elamvazhuthi** and Berman, 2018), for robots evolving on compact manifolds without boundary, and (Krishnan and Martínez, 2018), for robots evolving on a subset of a Euclidean space with boundary.

In models of robotic swarms, it is useful to consider *hybrid* variants of the SDE (3.4) to account for the fact that each robot, in addition to a continuous spatial state $\mathbf{Z}(t)$, can be associated with a discrete state $Y(t) \in \mathscr{V}$ at each time t. For such scenarios, we can define a hybrid switching diffusion process $(\mathbf{Z}(t), Y(t))$ as a system of SDEs of the form

$$\begin{aligned} d\mathbf{Z}(t) &= \mathbf{v}(Y(t),\mathbf{Z}(t),t)dt + \sqrt{2\mathbf{D}} \cdot d\mathbf{W}(t), \\ \mathbf{Z}(0) &= \mathbf{Z}_0, \end{aligned} \tag{1.24}$$

where $\mathbf{v} : \mathscr{V} \times \Omega \times [0,T] \to \mathbb{R}^n$ is the state- and time-dependent velocity vector field, and $\mathbf{D} \in \mathbb{R}^M_+$ is a vector of positive elements D_k, the diffusion coefficient associated with discrete

state $k \in \mathscr{V}$. Let $\mathbf{v}_k$ denote the velocity field associated with discrete state $k \in \mathscr{V}$ Then the forward equation for this system of SDEs is given by the system of PDEs

$$\begin{aligned}(y_k)_t &= D_k \Delta y_k - \nabla \cdot (\mathbf{v}_k(\mathbf{x},t) y_k) + \mathscr{F}_k \quad \text{in } \Omega \times [0,t_f], \\ y_k(\cdot,0) &= y_k^0 \quad \text{in } \Omega, \end{aligned} \tag{1.25}$$

where $k \in \mathscr{V}$ and $\mathscr{F}_k = \sum_{e \in \mathscr{E}} \sum_{j \in \mathscr{V}} u_e(t) B_e^{kj} y_j$, with $\mathbf{B}_e$ defined as in Subsection 1.2.1. The PDE (1.25) is related to the SDE (1.24), for each $k \in \mathscr{V}$, through the relation $\mathbb{P}(Y(t) = k, \mathbf{Z}(t) \in \Gamma) = \int_\Gamma y_k(\mathbf{x},t) d\mathbf{x}$ for all $t \in [0,T]$ and all measurable $\Gamma \subset \Omega$.

The class of models (1.25) is used in (Galstyan *et al.*, 2005) to model microscopic robots that reside in a fluid. In this work, some components of the vector are used to model robot densities, and some model them densities of chemicals that the robots follow. In (Milutinovic and Lima, 2006, 2007), the authors consider a 3-state model, with diffusion coefficients equal to 0, in which the time-dependent transition rates are optimized using infinite-dimensional optimal control theory (Fattorini, 1999). Each state is associated with an uncontrolled velocity vector field, corresponding to left-translation, right-translation, and remaining stationary. In (Hamann and Wörn, 2008; Hamann, 2010), these models are applied to study collective migration and collective perception tasks in swarms. To simulate the phenomenon of emergent taxis, the authors construct *mean-field feedback laws* in the sense that the diffusion coefficients are functions of the population densities, as in biological models of chemotaxis.

In (Berman *et al.*, 2011), the authors use model (1.25) to simulate the coverage activity of a swarm of robotic bees in a commercial pollination problem. The framework presented in (Berman *et al.*, 2011) is used in (**Elamvazhuthi** *et al.*, 2018c) to optimize time-dependent (and state-independent) robot velocities and state transition rates using optimal control theory of PDEs (Tröltzsch, 2010). Additionally, (**Elamvazhuthi** *et al.*, 2018c) considers the problem of identifying the spatial distribution of resources in the environment

from temporal robot data and frames this as a problem of identifying coefficients in model (1.25) using PDE-constrained optimization. Following a similar approach, (Ramachandran *et al.*, 2018) addresses the problem of mapping the boundaries of regions of interest in an environment from temporal robot data. In (**Elamvazhuthi** *et al.*, 2019), the authors analytically construct control laws $\mathbf{v}_k(\mathbf{x},t)$ and $u_e(t)$ to transport a swarm modeled by (1.25) from an initial probability density to a target density, thus establishing the controllability of the system (1.25).

When the parameters $\mathbf{v}_k(\mathbf{x},t)$ and $u_e(t)$ are independent of the density $\mathbf{y}$, the convergence of the solution of the mean-field model (3.32) to the density of a swarm with a finite number of agents can be concluded from the law of large numbers. However, such convergence results thus far have been mostly qualitative. A more quantitative convergence analysis of the model presented in (**Elamvazhuthi** *et al.*, 2018c) is performed in (Zhang *et al.*, 2018), where the density of the finite-agent model is shown to converge to the solution of the mean-field model as the number of agents tends to infinity. Using this convergence result, performance bounds are derived in (Zhang *et al.*, 2018) for the optimal control strategies constructed in (**Elamvazhuthi** *et al.*, 2018c) as a function of the approximation error due to the finiteness of the agent population.

Chapter 2

CONTROLLABILITY AND STABILIZATION OF FINITE-DIMENSIONAL FORWARD EQUATIONS

In this chapter, we present novel results on the controllability and stabilizability of the mean-field control problem for CTMCs described in Section 1.2.1. We study local and global controllability properties of the forward equation when the control inputs are required to be zero at equilibrium. The case when control inputs are not constrained to be zero at equilibrium is comparatively much easier, since local controllability follows directly from linearization-based arguments, so we do not consider this case here. We also demonstrate that it is possible to compute density-independent transition rates of a CTMC that make any probability distribution with a strongly connected support (to be defined later) invariant and globally stable. Similar work in (Acikmese and Bayard, 2012) has characterized the class of stabilizable stationary distributions for DTMCs with control parameters that are time- and density-invariant; we characterize this class of distributions for CTMCs with the same type of control parameters (see Theorem 2.3.4). We show that this result can be further strengthened by employing time-varying control parameters that make the system asymptotically controllable to any feasible probability distribution.

In addition, we address the stabilization of mean-field models using decentralized density feedback laws under the constraint that the transition rates are required to be zero at equilibrium. Such a constraint is needed in swarm robotic applications to prevent robots from constantly switching between states at equilibrium. The problem of unnecessary state-switching was previously addressed for CTMCs in (Mather and Hsieh, 2014) as a variance control problem, and for DTMCs in (Bandyopadhyay *et al.*, 2017) using a decentralized density estimation strategy that implements centralized feedback laws and ensures that the

transition matrix is the identity matrix at equilibrium. In this chapter, we investigate the CTMC case in more detail. In contrast to (Mather and Hsieh, 2014), we explicitly show that any distribution with a *strongly connected support* is stabilizable using a decentralized feedback law, and we impose the additional constraint that transition rates must be zero at equilibrium. Moreover, the controller in (Mather and Hsieh, 2014) was proved to be stabilizing with the assumption that negative transition rates are admissible, and was then implemented with a saturation condition in order to avoid negative rates, in which case the stability guarantees are lost. We show how this issue can be resolved with a linear controller by interpreting a negative flow from one state to another as a positive flow of appropriate magnitude in the opposite direction. While the algorithmic construction of linear controllers has low computational complexity, these controllers violate positivity constraints on the control inputs. To realize linear controllers in practice for our problem, we show that for bidirected graphs, we can implement linear controllers with rational feedback laws that mimic their behavior.

Lastly, we extend the stabilization results on density feedback-based stabilization to the more general case in which agents are not required in some states at equilibrium. In this case, the target distribution possibly has a *disconnected support*, meaning that the underlying subgraph induced by the vertices that are associated with positive target densities is disconnected. Stabilization of target distributions with disconnected supports is not possible using time- and density-independent control laws. If a desired distribution with disconnected support is a stationary distribution of a CTMC for a given set of time- and density-independent transition rates, then multiple other stationary distributions can be constructed from the disconnected components of the support of the desired distribution, thus obstructing global stability of this distribution. To bridge this gap, we propose a general class of decentralized control laws that can globally asymptotically stabilize any probability distribution. These feedback laws require each agent to know the density of agents

only in its current state, and thus rely only on information that can be locally acquired. The works (Hsieh *et al.*, 2008; Mather and Hsieh, 2011) also propose density-dependent feedback laws to address the swarm redistribution problem that we consider. However, the feedback laws in (Hsieh *et al.*, 2008), which are implemented using a quorum-sensing approach, stabilize a swarm only to positive target distributions, with a nonzero desired agent density in each state. In addition, while the control laws in (Hsieh *et al.*, 2008) are designed to yield a low rate of agent transitions between states at equilibrium, the transitions do not stop completely since the equilibrium control inputs are nonzero.

2.1 Notation

We first define some notation that will be used to formally state the problems addressed in this chapter. We will use the following definitions from graph theory. We denote by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ a directed graph with a set of M vertices, $\mathcal{V} = \{1, ..., M\}$, and a set of $N_{\mathcal{E}}$ edges, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$, where $e = (i, j) \in \mathcal{E}$ if there is an edge from vertex $i \in \mathcal{V}$ to vertex $j \in \mathcal{V}$. We define a source map $S : \mathcal{E} \to \mathcal{V}$ and a target map $T : \mathcal{E} \to \mathcal{V}$ for which $S(e) = i$ and $T(e) = j$ whenever $e = (i, j) \in \mathcal{E}$. There is a *directed path* of length s from a vertex $i \in \mathcal{V}$ to a vertex $j \in \mathcal{V}$ if there exists a sequence of edges $\{e_i\}_{i=1}^{s}$ in $\mathcal{E}$ such that $S(e_1) = i$, $T(e_s) = j$, and $S(e_k) = T(e_{k-1})$ for all $2 \leq k < s$. A directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is called *strongly connected* if for every pair of distinct vertices $v_0, v_T \in \mathcal{V}$, there exists a *directed path* of edges in $\mathcal{E}$ connecting v_0 to v_T. We will assume that $(i, i) \notin \mathcal{E}$ for all $i \in \mathcal{V}$. We will denote the set of outgoing edges from a vertex $v \in \mathcal{V}$ by $\mathcal{N}^{\text{out}}(v)$. The set of incoming edges to a vertex $v \in \mathcal{V}$ will be denoted by $\mathcal{N}^{\text{in}}(v)$. Throughout this chapter, we will assume that the graph $\mathcal{G}$ is strongly connected. We will also assume that $(i, i) \notin \mathcal{E}$ for all $i \in \mathcal{V}$. The graph $\mathcal{G}$ is said to be *bidirected* if $e \in \mathcal{E}$ implies that $\tilde{e} = (T(e), S(e))$ also lies in $\mathcal{E}$. We say that a vector $\mathbf{x}^d \in \mathbb{R}^M$ has a *strongly connected support* if the subgraph $\mathcal{G}_{sub} = (\mathcal{V}_{sub}, \mathcal{E}_{sub})$, defined by $\mathcal{V}_{sub} = \{v \in \mathcal{V} : \mathbf{x}^d_v > 0\}$ and $\mathcal{E}_{sub} = (\mathcal{V}_{sub} \times \mathcal{V}_{sub}) \cap \mathcal{E}$, is

strongly connected. Moreover, $\mathscr{V}_{sub}$ is called the *support* of the vector $\mathbf{x}^d$.

We denote the M-dimensional Euclidean space by $\mathbb{R}^M$. $\mathbb{R}^{M\times N}$ is the space of $M\times N$ matrices, and $\mathbb{R}_{\geq 0}$ is the set of non-negative real numbers. Given a vector $\mathbf{x}\in\mathbb{R}^M$, x_i will refer to the i^{th} coordinate value of $\mathbf{x}$. The $2-$norm of the vector $\mathbf{x}\in\mathbb{R}^M$ is denoted by $\|\mathbf{x}\|_2=\sqrt{\sum_i x_i^2}$. For a matrix $\mathbf{A}\in\mathbb{R}^{M\times N}$, A^{ij} will refer to the element in the i^{th} row and j^{th} column of $\mathbf{A}$. The spectrum of a matrix $\mathbf{A}$ will be denoted by $\text{spec}(\mathbf{A})$. Given a vector $\mathbf{y}\in\mathbb{R}^M$, for each vertex $i\in\mathscr{V}$, the set $\sigma_{\mathbf{y}}(i)\subset\mathscr{V}$ consists of all vertices j for which there exists a directed path $\{e_k\}_{k=1}^f$ of some length f from j to i such that $\mathbf{y}_{S(e_k)}=0$ for each $k=1,...,f-1$.

A matrix is *non-negative* if all its elements are non-negative, and it is *essentially non-negative* if all its off-diagonal elements are non-negative. A real eigenvalue λ_m of a matrix $\mathbf{A}$ will be called the *maximal eigenvalue* of $\mathbf{A}$ if $\lambda_m\geq|\lambda|$ for all $\lambda\in\text{spec}(\mathbf{A})$. We will denote the *conical span* of a set C of m vectors $\mathbf{x}_i\in\mathbb{R}^M$, $i=1,...,m$, by $\text{co span}(C)=\{\sum_{i=1}^m\alpha_i\mathbf{x}_i:\mathbf{x}_i\in C,\ \alpha_i\in\mathbb{R}_{\geq 0},\ i=1,...,m\}$.

The matrix $\mathscr{L}_{out}(\mathscr{G})=\mathbf{D}_{out}(\mathscr{G})-\mathbf{A}(\mathscr{G})\in\mathbb{R}^{M\times M}$ denotes the *out-Laplacian* of the graph $\mathscr{G}$, where $\mathbf{D}_{out}(\mathscr{G})$ is the out-degree matrix of $\mathscr{G}$ and $\mathbf{A}(\mathscr{G})$ is the adjacency matrix of $\mathscr{G}$. $\mathbf{D}_{out}(\mathscr{G})$ is a diagonal matrix for which $(\mathbf{D}_{out}(\mathscr{G}))^{ii}$ is the total number of edges e such that $S(e)=i$. The entries of $\mathbf{A}(\mathscr{G})$ are defined as $(\mathbf{A}(\mathscr{G}))^{ij}=1$ if $(j,i)\in\mathscr{E}$, and 0 otherwise. When $\mathscr{G}$ is bidirected, $\mathscr{L}_{out}(\mathscr{G})$ is the usual Laplacian of the graph, and we will drop the subscript and denote it by $\mathscr{L}(\mathscr{G})$. For a subset $B\subset\mathbb{R}^M$, int B and $\text{Bd}(B)$ will refer to the interior and the boundary, respectively, of B.

We will also need some basic notions from set-valued analysis (Aubin and Frankowska, 2009). We will use $\mathbf{F}:X\rightrightarrows Y$ to denote a *set-valued map*, i.e., a map $\mathbf{F}$ from a metric space X to the power set of a metric space Y. Let $\mathscr{B}_\eta(\mathbf{x})$ denote the open ball with center $\mathbf{x}\in X$ and radius $\eta>0$. Then the set-valued map $\mathbf{F}$ will be called *upper semi-continuous* at $\mathbf{x}\in X$ if and only if for any neighborhood $\mathscr{U}$ of $\mathbf{F}(\mathbf{x})$, there exists $\eta>0$ such that for all

$\mathbf{x}' \in B_X(\mathbf{x},\eta)$, $\mathbf{F}(\mathbf{x}') \subset \mathscr{U}$. If A is a subset of $\mathbb{R}^M$, we define the distance between a point $\mathbf{x} \in \mathbb{R}^M$ and the set A using the notation $\text{dist}(\mathbf{x},A) = \inf_{\mathbf{y}\in A} \|\mathbf{x}-\mathbf{y}\|$. The notation $\bar{\text{co}}\, A$ will denote the convex closure of the set A in X. The notation $\bar{\text{co}}\, \mathbf{F}$ will denote the set-valued map that is defined by setting $(\bar{\text{co}}\, \mathbf{F})(\mathbf{x}) = \bar{\text{co}}\, \mathbf{F}(\mathbf{x})$ for all $\mathbf{x} \in X$. A function $\mathbf{f} : \mathbb{R} \to \mathbb{R}^M$ is said to be absolutely continuous if $\forall \varepsilon > 0$, there exists $\delta > 0$ such that for any finite set of disjoint intervals $(a_1,b_1),...,(a_N,b_N)$, $\sum_{j=1}^N (b_j - a_j) < \delta \implies \sum_{j=1}^N \|\mathbf{f}(b_j) - \mathbf{f}(a_j)\| < \varepsilon$. More generally, $\mathbf{f}$ is said to be absolutely continuous on $[a,b]$ if this condition is satisfied whenever the intervals (a_j,b_j), $j = 1,...,N$, all lie in $[a,b]$.

In this chapter, we will analyze the forward equation of a CTMC presented in Section 1.2.1. For the reader's convenience, we recall the description of this model from Section 1.2.1. There are N autonomous agents whose states evolve in continuous time according to a Markov chain with a finite state space defined as the vertex set $\mathscr{V} = \{1,...,M\}$. For example, the vertices in $\mathscr{V}$ can represent a set of tasks that the agents must perform, or a set of spatial locations obtained by partitioning the agents' environment. The edge set $\mathscr{E} \subset \mathscr{V} \times \mathscr{V}$ defines the pairs of vertices between which the agents can transition. The directed graph $\mathscr{G} = (\mathscr{V},\mathscr{E})$ is assumed to be strongly connected. The agents' transition rules are determined by the control parameters $u_e : [0,\infty) \to \mathbb{R}_{\geq 0}$ for each $e \in \mathscr{E}$, and are known as the *transition rates* of the associated CTMC. The state of each agent $i \in \{1,...,N\}$ at time t is defined by a stochastic process $X_i(t)$ that evolves on the state space $\mathscr{V}$ according to the conditional probabilities

$$\mathbb{P}(X_i(t+h) = T(e) | X_i(t) = S(e)) = u_e(t)h + o(h) \tag{2.1}$$

for each $e = (S(e),T(e)) \in \mathscr{E}$, where $S(e)$ and $T(e)$ denote the source and target vertices of the edge e, respectively. Here, $o(h)$ is the little-oh symbol and $\mathbb{P}$ is the underlying probability measure induced on the space of events Ω by the stochastic processes $\{X_i(t)\}_{i=1}^N$. Let $\mathscr{P}(\mathscr{V}) = \{y \in \mathbb{R}^M_{\geq 0};\ \sum_v y_v = 1\}$ be the simplex of probability densities on $\mathscr{V}$, and let

int $\mathscr{P}(\mathscr{V})$ be the interior of this simplex. Corresponding to the CTMC is a system of ordinary differential equations (ODEs) that determines the time evolution of the probability densities $\mathbb{P}(X_i(t) = v) = x_v(t) \in \mathbb{R}_{\geq 0}$. If $X_i(0)$ are independent and identically distributed (IID), then the processes $\{X_i(t)\}_{i=1}^N$ are also IID, and the *Kolmogorov forward equation* can be represented by a single linear system of ODEs.

We recall the definition of the forward equation of a CTMC,

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e \in \mathscr{E}} u_e(t) \mathbf{B}_e \mathbf{x}(t), \quad t \in [0, \infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{2.2}$$

where $\mathbf{B}_e$ are control matrices whose entries are given by

$$B_e^{ij} = \begin{cases} -1 & \text{if } i = j = S(e), \\ 1 & \text{if } i = T(e),\ j = S(e), \\ 0 & \text{otherwise.} \end{cases} \tag{2.3}$$

We note that $\mathscr{P}(\mathscr{V})$ is an invariant set for system (2.2) because $\mathbf{B}_e$ has off-diagonal positive entries, the columns sum to 0, and the control inputs $u_e(t)$ are constrained to be non-negative. This fact will be used throughout the chapter.

2.2 Controllability of the Forward Equation of a CTMC

The focus of this section is to study the controllability of the control system (2.2). Toward this end, we will address the following problems.

Problem 2.2.1. ***(Global controllability)*** *Given $\mathbf{x}^0, \mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ and $T > 0$, determine if there exist bounded time-dependent non-negative control parameters $\{u_e\}_{e \in \mathscr{E}}$ for system (2.2) such that $\mathbf{x}(T) = \mathbf{x}^d$.*

Problem 2.2.2. ***(Local controllability of underactuated forward equation)*** *Given $\mathbf{x}^0, \mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ and $T > 0$, let $\mathscr{E} = \mathscr{E}_0 \cup \mathscr{E}_1$ be a partition of the set of edges $\mathscr{E}$. Determine if there*

exists $r > 0$ *such that each point in the neighborhood* $B(\mathbf{x}^d, r) \cap \mathscr{P}(\mathscr{V})$ *for system* (2.2) *is reachable within a finite time using time-varying control inputs* $\{u_e\}_{e \in \mathscr{E}_2}$ *with* $u_{\hat{e}} = 1$ *for all* $\hat{e} \in \mathscr{E}_1$.

When $\mathscr{E}_0$ is empty, that is, when all the transition rates can be specified and hence the system is **fully actuated**, Problem 2.2.1 and Problem 2.2.2 are equivalent. The above generalized problem (Problem 2.2.2) might be relevant in control problems where the control input can act only locally on the graph. For example, this could be the case when the Markov chain represents a traffic flow model (Yu *et al.*, 2003), where it might not be possible for an external supervisor to control the flow along all the edges.

Note that these controllability results for Problem 2.2.1 and Problem 2.2.2 could not be directly concluded from classical tests of controllability such as the Kalman rank condition or the Lie Algebra Rank conditions (Bloch, 2015) due to the positivity constraints on the control inputs. Here, we prove a general result which implies that the classical rank conditions for controllability have a simple generalization to non-negative control inputs. These generalized rank conditions can be used to establish the controllability result for system (2.2) that was proved in (**Elamvazhuthi** *et al.*, 2019), in the case where all control inputs can be specified. More importantly, we will apply these conditions to establish the local controllability for the underactuated case in which only a subset of the control inputs can be designed (Example 2.2.6). A result such as the one we present is already known for general nonlinear control systems with control constraints for the case when the linearized control system with the same constraints is also controllable (Klamka, 1996). On the other hand, the following result also applies to the larger class of controllable nonlinear systems when the linearized system is not controllable, but controllability follows from Lie rank conditions (Bloch, 2015). Moreover, our arguments are more elementary, and it will be less cumbersome to address Problem 2.2.2 directly from our generalization of the rank condition, rather than to invoke the result from (Klamka, 1996).

Finally, we will also address the following problem in this section.

Problem 2.2.3. ***(Asymptotic controllability)*** *Given $\mathbf{x}^0, \mathbf{x}^d \in \mathscr{P}(\mathscr{V})$, determine if there exist globally bounded time-dependent non-negative control parameters $\{u_e\}_{e\in\mathscr{E}}$ for system* (2.2) *such that* $\lim_{t\to\infty} \|\mathbf{x}(t) - \mathbf{x}^d\| = 0$.

Having defined the problems that will be addressed in this section, we will start by addressing Problem 2.2.1 and Problem 2.2.2.

Theorem 2.2.4. *Consider the control-affine system*

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}_0(\mathbf{x}(t)) + \sum_{i=1}^{N} u_i(t)\mathbf{f}_i(\mathbf{x}(t)) \\ \mathbf{x}(0) &= \mathbf{x}^0 \end{aligned} \tag{2.4}$$

with smooth vector fields $\mathbf{f}_i : \mathbb{R}^M \to \mathbb{R}^M$ for $i = 0, ..., N$. Suppose $\mathbf{x}^f \in \mathbb{R}^M$ and there exist control inputs $u_i : [0,T] \to \mathbb{R}$ such that a unique solution of system (2.4) *exists and satisfies $\mathbf{x}(T) = \mathbf{x}^f$. Additionally, suppose that the following condition holds for all $t \in [0,T]$:*

$$\operatorname{span}\{\mathbf{f}_i(\mathbf{x}(t)) : \ \ i = 1, ..., N\} = \operatorname{co}\operatorname{span}\{\mathbf{f}_i(\mathbf{x}(t)) : i = 1, ..., N\}. \tag{2.5}$$

Then there exist measurable control inputs $\tilde{u}_i : [0,T] \to \mathbb{R}_{\geq 0}$ such that the state $\mathbf{x}(t)$ evolves according to the following system for almost every (a.e.) $t \in [0,T]$:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{f}_0(\mathbf{x}(t)) + \sum_{i=1}^{N} \tilde{u}_i(t)\mathbf{f}_i(\mathbf{x}(t)) \\ \mathbf{x}(0) &= \mathbf{x}^0 \end{aligned} \tag{2.6}$$

Proof. The proof is a simple application of a representation theorem due to Filippov. We consider the set-valued map $\mathbf{F} : [0,T] \rightrightarrows \mathbb{R}^M$ defined by

$$\mathbf{F}(t) = \mathbb{R}^M_{\geq 0}, \ \ \forall t \in [0,T] \tag{2.7}$$

Here, by a set-valued map, we mean that $\mathbf{F}$ takes values from $[0,T]$ to the power set of $\mathbb{R}^M$. We define the map $\mathbf{g}:[0,T]\times\mathbb{R}^N\to\mathbb{R}^M$ by

$$\mathbf{g}(t,\mathbf{v})=\mathbf{f}_0(\mathbf{x}(t))+\sum_{i=1}^{N}v_i\mathbf{f}_i(\mathbf{x}(t)) \tag{2.8}$$

for all $t\in[0,T]$ and for all $\mathbf{v}\in\mathbb{R}^N$. The map $\mathbf{g}$ is a *Carathéodory map*; that is, $\mathbf{g}(\cdot,\mathbf{v})$ is measurable for every $\mathbf{v}\in\mathbb{R}^N$, and the map $\mathbf{g}(t,\cdot)$ is continuous for every $t\in[0,T]$. From the definitions of $\mathbf{g}$ and $\mathbf{F}$, it follows that $\dot{\mathbf{x}}(t)\in\mathbf{g}(t,\mathbf{F}(t)):=\{g(t,\mathbf{z});\ \mathbf{z}\in\mathbf{F}(t)\}$ for a.e. $t\in[0,T]$ due to assumption (2.5). Therefore, by (Aubin and Frankowska, 2009)[Theorem 8.2.10], there exists a measurable function $\tilde{\mathbf{u}}:[0,T]\to\mathbb{R}^N_{\geq 0}$ such that

$$\dot{\mathbf{x}}(t)=\mathbf{f}_0(\mathbf{x}(t))+\sum_{i=1}^{N}\tilde{u}_i(t)\mathbf{f}_i(\mathbf{x}(t)) \tag{2.9}$$

for almost every $t\in[0,T]$. □

The above theorem can be used to establish the following controllability result for system (2.2).

Theorem 2.2.5. ***(Global controllability for fully actuated system)*** *Let $\mathbf{x}^d\in\text{int }\mathscr{P}(\mathscr{V})$ and $T>0$ be given. Suppose $\mathbf{x}^0\in\text{int }\mathscr{P}(\mathscr{V})$. Then there exist control inputs $\{u_e(t)\}_{e\in\mathscr{E}}$ such that the solution of system* (2.2) *satisfies $\mathbf{x}(T)=\mathbf{x}^d$.*

Proof. To conclude the above result from Theorem 2.2.4, one only needs to observe that the conical span of the set $\{B_e\mathbf{y}\}_{e\in\mathscr{E}}=T_{\mathbf{y}}\mathscr{P}(\mathscr{V})=\{\mathbf{x}\in\mathbb{R}^M;\ \sum_i x_{i=1}^M=1\}$, the *tangent space* of $\mathscr{P}(\mathscr{V})$ at $\mathbf{y}$, for all $\mathbf{y}\in\text{int }\mathscr{P}(\mathscr{V})$. To see this explicitly, note that since the graph $\mathscr{G}$ is strongly connected, we know from the Perron-Frobenius theorem (Minc, 1988) that span $\{B_e\mathbf{1}\}_{e\in\mathscr{E}}=T_{\mathbf{1}}\mathscr{P}(\mathscr{V})$. Moreover, due to the strongly connected nature of the graph $\mathscr{G}$, it follows that if $e\in\mathscr{E}$, then there exists a directed path $(e_i)_{i=1}^p$ of length p such that $S(e_1)=T(e)$ and $T(e_p)=S(e)$. Hence, $\sum_{i=1}^p B_{e_i}\mathbf{1}=-B_e\mathbf{1}$. And hence, we also have that co span $\{B_e\mathbf{1}\}_{e\in\mathscr{E}}=T_{\mathbf{1}}\mathscr{P}(\mathscr{V})$. For general $\mathbf{y}\in\text{int }\mathscr{P}(\mathscr{V})$, the result follows. Therefore,

given any path $\gamma(t) \in \text{int } \mathscr{P}(\mathscr{V})$ that is differentiable, which implies that $\dot{\gamma}(t) \in T_{\gamma(t)}\mathscr{P}(\mathscr{V})$, the path can be realized by system (2.2) using an appropriate choice of measurable control inputs $\{u_e(t)\}_{e\in\mathscr{E}}$. □

Now we address our main motivation for proving Theorem 2.2.4. The following example demonstrates the possibility of achieving local controllability of system (2.2) even when the system is underactuated, in contrast with the requirement in Theorem 2.2.5.

Example 2.2.6. *Let* $\mathscr{V} = \{1,...,4\}$, $\mathscr{E}_0 = \{(1,2),(2,1)\}$, *and* $\mathscr{E}_1 = \{(2,3),(3,4),(4,2)\}$ *(see Fig. 2.1). We set* $\mathscr{G} = (\mathscr{V},\mathscr{E})$, *where* $\mathscr{E} = \mathscr{E}_0 \cup \mathscr{E}_1$. *We consider a variant of the control system* (2.2) *in which the control inputs* $u_e(t)$, $e \in \mathscr{E}_0$ *are each set to* 1, *and the inputs* $u_e(t)$, $e \in \mathscr{E}_1$ *can be designed:*

$$\dot{\mathbf{x}}(t) = \sum_{e\in\mathscr{E}_0} \mathbf{B}_e\mathbf{x}(t) + \sum_{e\in\mathscr{E}_1} u_e(t)\mathbf{B}_e\mathbf{x}(t), \;\; \mathbf{x}(0) = \mathbf{x}_0 \in \mathscr{P}(\mathscr{V}). \tag{2.10}$$

Let $\mathbf{x}^d = [\frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4} \;\; \frac{1}{4}]^T \in \mathscr{P}(\mathscr{V})$. *The Kalman rank condition can be used to verify that the control system* (2.10) *linearized about the point* $\mathbf{x}^d$ *is controllable. Hence, system* (2.10) *is locally controllable (Bloch, 2015); that is, given* $T > 0$, *there exists* $r > 0$ *and a neighborhood* $B(\mathbf{x}^d, r) \cap \mathscr{P}(\mathscr{V})$ *of* $\mathbf{x}^d$ *such that for each* $\mathbf{x}^d \in B(\mathbf{x}^d, r) \cap \mathscr{P}(\mathscr{V})$, *there exist measurable control inputs* $\mathbf{u}_e(t)$ *for* $e \in \mathscr{E}_1$, *possibly with negative entries at some time* t, *such that* $\mathbf{x}(T) = \mathbf{x}^d$. *Moreover, a straightforward computation of* $\mathbf{B}_e\mathbf{y}$ *confirms that* $\text{span}\{\mathbf{B}_e\mathbf{y} : e \in \mathscr{E}_1\} = \text{co span}\{\mathbf{B}_e\mathbf{y} : e \in \mathscr{E}_1\}$ *for all* $\mathbf{y} \in \text{int } \mathscr{P}(\mathscr{V})$. *Hence, by Theorem 2.2.4, system* (2.10) *is locally controllable at* $\mathbf{x}^d$ *using only the non-negative control inputs corresponding to* $\mathscr{E}_1$, *a subset of the edges in* $\mathscr{E}$.

The above example can be generalized to give the following sufficient condition for local controllability.

Theorem 2.2.7. *Let* $\mathscr{G}$ *be a strongly connected graph with* $\mathscr{E} = \mathscr{E}_0 \cup \mathscr{E}_1$. *Let system* (2.2) *be small-time locally controllable at* $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$ *without non-negativity constraints on the*

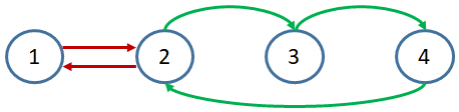


Figure 2.1: (**Example 2.2.6**) Edges in $\mathscr{E}_0$ are Uncontrolled and Denoted by the Red Arrows. Controlled Edges in $\mathscr{E}_1$ are Denoted by the Green Arrows.

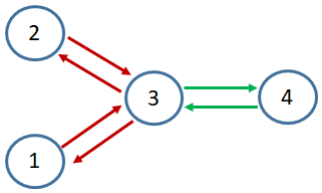


Figure 2.2: (**Example 2.2.8**) Edges in $\mathscr{E}_0$ are Denoted by the Red Arrows. Edges in $\mathscr{E}_1$ are Denoted by the Green Arrows.

control inputs. Then the system (2.2) *is small-time locally controllable at* $\mathbf{x}^d \in \text{int}\,\mathscr{P}(\mathscr{V})$ *if* $e \in \mathscr{E}_1$ *implies that there exists a directed path* $(e_i)_{i=1}^p$ *of length* p *from the vertex* $T(e)$ *to the vertex* $S(e)$ *such that* $e_i \in \mathscr{E}_1$ *for all* $i \in \{1,...,p\}$. *In particular, with this assumption on* $\mathscr{E}_1$, *if the linearization of control system* (2.10) *is controllable, then system* (2.10) *is small-time locally controllable.*

The controllability result in Theorem 2.2.5 was proved for the case where all control inputs can be specified, in contrast to Example 2.2.6, in which only a subset of these inputs can be designed. To prove Theorem 2.2.5, it was sufficient to assume that the graph $\mathscr{G}$ is strongly connected. The following example shows that when only a small subset of the control inputs can be designed, strong connectivity of the graph is not a sufficient condition for proving local controllability of system (2.2).

Example 2.2.8. *Let* $\mathscr{V} = \{1,...,4\}$, $\mathscr{E}_0 = \{(1,3),(3,1),(2,3),(3,2)\}$, *and* $\mathscr{E}_1 = \{(3,4),(4,3)\}$ *(see Fig. 2.2). We set* $\mathscr{G} = (\mathscr{V},\mathscr{E})$, *where* $\mathscr{E} = \mathscr{E}_0 \cup \mathscr{E}_1$. *Note that* $\mathscr{G}$ *is strongly connected.*

We consider the control system (2.10) *with this graph. If* $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$ *is such that* $x_1^0 = x_2^0$*, then the solution* $\mathbf{x}(t)$ *of system* (2.10) *satisfies* $x_1(t) = x_2(t)$ *for all* $t \geq 0$ *for any choice of control inputs* $u_e(t)$*,* $e \in \mathscr{E}_1$*. Hence, although* $\mathscr{G}$ *is strongly connected, system* (2.10) *is not locally controllable at any point* $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ *that satisfies* $x_1^d = x_2^d$*. The nature of this obstruction to controllability is similar to the one in leader-based control of linear consensus protocols (Mesbahi and Egerstedt, 2010), where inputs act at the vertices rather than the edges, and symmetries in the network with respect to input locations have detrimental effects on the controllability of the system.*

It would be desirable to extend the above controllability results to include target distributions that lie on the boundary of $\mathscr{P}(\mathscr{V})$. However, one cannot expect to reach target distributions on the boundary in finite time. The boundary of $\mathscr{P}(\mathscr{V})$ is unreachable if the initial condition of system (2.2) starts from the interior of $\mathscr{P}(\mathscr{V})$, even if one uses possibly unbounded but measurable inputs with finite Lebesgue integral. The following counterexample clarifies this point.

Example 2.2.9. *Consider system* (2.2) *for a bidirected graph* $\mathscr{G}$ *with two vertices:*

$$\begin{aligned} \dot{x}_1(t) &= -u_{(1,2)}(t)x_1(t) + u_{(2,1)}x_2(t), \\ \dot{x}_2(t) &= u_{(1,2)}(t)x_1(t) - u_{(2,1)}x_2(t), \end{aligned} \tag{2.11}$$

$$x_1(0) = x_1^0, \quad x_2(0) = x_2^0.$$

Let $u_{(1,2)}, u_{(2,1)} \in L_+^1(0,1)$*, the set of positive-valued measurable inputs with finite integrals over the time interval* $(0,1)$*. Then the solution,* $\mathbf{x}(t) = [x_1(t) \;\; x_2(t)]^T$*, satisfies:*

$$x_1(t) = x_1^0 - \int_0^t (u_{(1,2)}(\tau)x_1(\tau) - u_{(2,1)}(\tau)x_2(\tau))d\tau, \tag{2.12}$$

$$x_2(t) = x_2^0 + \int_0^t (u_{(1,2)}(\tau)x_1(\tau) - u_{(2,1)}(\tau)x_2(\tau))d\tau, \tag{2.13}$$

such that $x_1^0 \in (0,1)$ *and* $x_2^0 = 1 - x_1^0$. *We assume, without loss of generality, that* $x_1(t) > 0$ *for all* $t \in [0,1)$. *Then for each* $T \in [0,1)$, *Equations* (2.12) *and* (2.13) *imply that:*

$$x_1(T) = x_1^0 - \int_0^T \left(u_{(1,2)}(\tau) + u_{(2,1)}(\tau) - \frac{u_{(2,1)}(\tau)}{x_1(\tau)} \right) x_1(\tau) d\tau.$$

From this equation, we can conclude that

$$\begin{aligned} x_1(1) &\geq x_1^0 - \int_0^1 (u_{(1,2)}(\tau) + u_{(2,1)}(\tau)\tilde{x}_1(\tau)) d\tau \\ &= \exp\left(- \int_0^1 (u_{(1,2)}(\tau) + u_{(2,1)}(\tau)) d\tau \right) x_1^0, \end{aligned} \tag{2.14}$$

where $\tilde{x}_1$ *is the solution of the differential equation*

$$\begin{aligned} \dot{\tilde{x}}_1(t) &= -(u_{(1,2)}(t) + u_{(2,1)}(t))\tilde{x}_1(t), \\ \tilde{x}_1(0) &= x_1^0. \end{aligned} \tag{2.15}$$

Therefore, it must be true that $\exp(-\int_0^1 (u_{(1,2)}(\tau) + u_{(2,1)}(\tau)) d\tau) x_1^0 \leq 0$, *which yields a contradiction since* $x_1^0 \neq 0$.

In the following theorem, we establish a general negative controllability result for the case where the control inputs are restricted to be bounded.

Theorem 2.2.10. *Let* $\mathbf{x}^0 \in \operatorname{int} \mathscr{P}(\mathscr{V})$ *and* $T \geq 0$. *Suppose that the control inputs* $u_e(t)$ *are essentially bounded over the time interval* $[0,T]$. *Then the solution* $\mathbf{x}(t)$ *of the control system* (2.2) *satisfies* $\mathbf{x}(t) \in \operatorname{int} \mathscr{P}(\mathscr{V})$ *for all* $t \in [0,T]$.

Proof. For the sake of contradiction, suppose that there exist bounded piecewise control inputs $u_e(t)$ such that the solution $\mathbf{x}(t)$ of the control system (2.2) satisfies $\mathbf{x}(T) = \mathbf{x}^d \in \operatorname{Bd}(\mathscr{P}(\mathscr{V}))$. Since $\mathbf{x}^d \in \operatorname{Bd}(\mathscr{P}(\mathscr{V}))$, there exists $i \in \mathscr{V}$ such that $x_i^d = 0$. Note that $x_i(t) = x_i^0 + \sum_{e \in \mathscr{E}_0} \int_0^t u_e(\tau) x_{S(e)}(\tau) d\tau - \sum_{e \in \mathscr{E}_1} \int_0^t u_e(\tau) x_i(\tau) d\tau$ for all $t \in [0,T]$, where $\mathscr{E}_0$ is the set of edges e such that $T(e) = i$ and $\mathscr{E}_1$ is the set of edges e such that $S(e) = i$. Then $x_i(t) \geq \hat{x}_i(t) =$

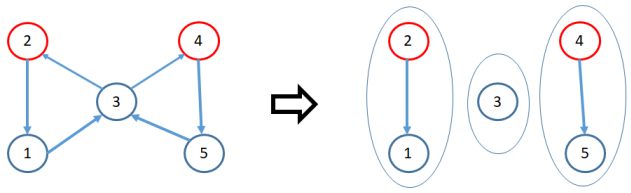


Figure 2.3: Illustration of the splitting of the graph $\mathscr{G}$ in the proof of Theorem 2.2.11. The graph on the left is the original graph $\mathscr{G}$. The target densities at the red vertices are equal to zero. The graphs on the right show the splitting of the graph into 3 disjoint graphs with rooted in-branches whose root vertices, shown in blue, have positive target densities.

$\exp\left(-\sum_{e\in\mathscr{E}_1}\int_0^t\sum_{e\in\mathscr{E}_1}\|u_e\|_\infty d\tau\right)x_i^0 = x_i^0 - \sum_{e\in\mathscr{E}_1}\int_0^t\|u_e\|_\infty\hat{x}_i(\tau)d\tau$ for all $t\in[0,T]$. Otherwise, due to continuity of the solution $\mathbf{x}(t)$ with respect to time t, there would exist a time $t_{in}\in(0,T)$ at which $\dot{x}_i(t_{in}) < \dot{\hat{x}}_i(t_{in})$ and $x_i(t_{in}) = \hat{x}_i(t_{in})$. Hence, the inequality $x_i(t)\geq\hat{x}_i(t)$ must hold for all $t\in[0,T]$. However, given the initial assumption that $x_i(T) = x_i^d(T) = 0$, this inequality leads to a contradiction since $\exp\left(-\sum_{e\in\mathscr{E}_1}\int_0^t\sum_{e\in\mathscr{E}_1}\|u_e\|_\infty d\tau\right)x_i^0 > 0$. Therefore, the boundary set $\mathrm{Bd}(\mathscr{P}(\mathscr{V}))$ is not reachable in finite time, using piecewise constant control inputs that are bounded from above by $\max_{e\in\mathscr{E}}\|u_e\|_\infty$ and bounded from below by $-\max_{e\in\mathscr{E}}\|u_e\|_\infty$. This implies that $\mathrm{Bd}(\mathscr{P}(\mathscr{V}))$ is not reachable in finite time using essentially bounded control inputs, since any essentially bounded function can be approximated uniformly using piecewise constant functions. □

In contrast with the above result, which shows that the boundary points of $\mathscr{P}(\mathscr{V})$ are not reachable in finite time, the next theorem proves that these points can be reached asymptotically as $t\to\infty$.

Theorem 2.2.11. *Suppose that $\mathbf{x}^0\in\mathscr{P}(\mathscr{V})$ is the initial distribution, and $\mathbf{x}^d\in\mathscr{P}(\mathscr{V})$ is the desired distribution. Then for each $e\in\mathscr{E}$, there exists a set of time-dependent control inputs $u_e:\mathbb{R}_{\geq 0}\to\mathbb{R}_{\geq 0}$, $e\in\mathscr{E}$, such that the solution $\mathbf{x}(t)$ of the control system (2.2) satisfies $\lim_{t\to\infty}\mathbf{x}(t) = \mathbf{x}^d$.*

Proof. We define the set $\mathscr{R} = \{i : x_i^d > 0,\ i = 1,...,M\}$ with cardinality $N_{\mathscr{R}}$. Let $\mathscr{I} : \{1,2,...,N_{\mathscr{R}}\} \to \mathscr{R}$ be a bijective map that defines an ordering on $\mathscr{R}$. Then we recursively define a collection $\{\mathscr{V}_n\}$ of disjoint subsets of $\mathscr{V}$ as follows:

$$\mathscr{V}_1 = \{\mathscr{I}(1)\} \cup \{i \in \mathscr{V} : x_i^d = 0 \ \ s.t. \ \ i \in \sigma_{\mathbf{x}^d}(\mathscr{I}(1))\}$$

$$\mathscr{V}_n = \{\mathscr{I}(n)\} \cup \{i \in \mathscr{V} : x_i^d = 0 \ \ s.t. \ \ i \in \sigma_{\mathbf{x}^d}(\mathscr{I}(n)) \text{ and } i \notin \cup_{k=1}^{n-1} \mathscr{V}_k\}$$

for each $n \in \{2,3,...,N_{\mathscr{R}}\}$. We note that $\mathscr{V} = \cup_{n=1}^{N_{\mathscr{R}}} \mathscr{V}_n$. Let $\mathbf{x}^{in} \in \text{int } \mathscr{P}(\mathscr{V})$ be some element such that $\sum_{k \in \mathscr{V}_n} x_k^{in} = x^d_{\mathscr{I}(n)}$ for each $n \in \{1,2,...,N_{\mathscr{R}}\}$. From (**Elamvazhuthi** *et al.*, 2019)[Theorem IV.17], we know that there exists a control $u_e^1 : [0,T] \to \mathbb{R}_{\geq 0}$ for each $e \in \mathscr{E}$ such that the solution $\mathbf{x}(t)$ of system (2.2) satisfies $\mathbf{x}(T) = \mathbf{x}^{in}$. Now we will design $\{u_e\}_{e \in \mathscr{E}}$ such that $u_e(t) = u_e^1(t)$ for each $t \in [0,T]$ and $u_e(t) = a_e$ for each $t \in (T,\infty]$, where a_e is defined as follows:

$$a_e = \begin{cases} 0 & \text{if } S(e) \in \mathscr{V}_n \text{ and } T(e) \notin \mathscr{V}_n, \ \ n \in \{1,...,N_{\mathscr{R}}\}, \\ 0 & \text{if } S(e) = \mathscr{I}(n) \ \text{ for some } n \in \{1,...,N_{\mathscr{R}}\}, \\ 1 & \text{otherwise.} \end{cases}$$

Then the solution of system (2.2) for $t > T$ can be constructed from the solution of the following decoupled set of ODEs:

$$\begin{aligned} \dot{\mathbf{y}}_n(t) &= -\mathscr{L}_{out}(\tilde{\mathscr{G}}_n)\mathbf{y}_n(t), \quad t \in [T,\infty) \\ \mathbf{y}_n(T) &= \mathbf{y}_n^0 \in \mathscr{P}(\mathscr{V}_n) \end{aligned} \tag{2.16}$$

for $n = 1,...,N_{\mathscr{R}}$. Here, $\mathscr{G}_n = (\mathscr{V}_n, \mathscr{E}_n)$ for each $n \in \{1,...,N_{\mathscr{R}}\}$, where $e \in \mathscr{E}_n$ if $S(e), T(e) \in \mathscr{V}_n$ and $a_e = 1$. See Fig. 2.3 for an illustration. The solution of system (2.16) is related to the solution of system (2.2) with $\mathbf{x}(T) = \mathbf{x}^{in}$ through a suitable permutation matrix $\mathbf{P}$, defined such that $\mathbf{P}\mathbf{x}(t) = [\mathbf{y}_1(t)\ \mathbf{y}_2(t)\ ...\ \mathbf{y}_{N_{\mathscr{R}}}(t)]$. Since each graph $\mathscr{G}_n$ has a rooted in-

branching subgraph, the process generated by $-\mathscr{L}_{out}(\tilde{\mathscr{G}}_n)^T$ has a unique stationary distribution. Moreover, by construction, this unique, globally stable stationary distribution is the vector $[x^d_{\mathscr{I}(n)}\ \mathbf{0}_{1\times(|\mathscr{V}_n|-1)}]^T$, where $|\mathscr{V}_n|$ is the cardinality of the set $\mathscr{V}_n$. This implies that $\lim_{t\to\infty}\mathbf{P}^{-1}\mathbf{y}(t) = \lim_{t\to\infty}\mathbf{x}(t) = \mathbf{x}^d$. By concatenating the control inputs $\{u^1_e\}_{e\in\mathscr{E}}$ and $\{a_e\}_{e\in\mathscr{E}}$, we obtain the desired asymptotic controllability result. □

An interesting aspect of the above proof is its implication that asymptotic controllability is achievable with piecewise constant control inputs with a finite number of pieces. From the above result, it follows that any point in $\mathscr{P}(\mathscr{V})$ can be stabilized using a full-state feedback controller (Clarke *et al.*, 1997). However, for a general target equilibrium distribution, a stabilizing controller with a decentralized structure might not exist.

2.3 Stabilization of the Forward Equation of a CTMC

Now we investigate the stabilizability properties of system (2.2). Note that stabilizability using centralized feedback follows from the asymptotic controllability result in Theorem 2.2.11 and a result in (Clarke *et al.*, 1997) which states that asymptotic controllability implies feedback stabilizability. In contrast, our focus in this section is to establish stabilizability using *decentralized control laws*, which does not follow from (Clarke *et al.*, 1997).

The problems in Section 2.2 allow the control inputs to be time-varying. We now pose a problem in which the control inputs are constrained to be time-independent. The motivation for the following problem is that time-invariant control laws are easier to implement. However, as a trade-off, only a smaller class of target distributions can be reached using such control laws as compared to the time-varying case (see Theorem 2.3.4).

Problem 2.3.1. ***(Open-loop stabilization)*** *Given $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$, determine if there exist globally bounded time-dependent non-negative control parameters $\{u_e\}_{e\in\mathscr{E}}$ for system* (2.2)

such that $\lim_{t\to\infty} \|\mathbf{x}(t) - \mathbf{x}^d\| = 0$ *for all* $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$.

By addressing Problem 2.2.3 and Problem 2.3.1, we provide a complete characterization of the stationary distributions that are stabilizable for CTMCs with forward equation (2.2) and transition rates u_e that may be either time-independent or time-dependent. Although time-independent transition rates of CTMCs have been previously computed in an optimization framework (Berman *et al.*, 2009), the question of which equilibrium distributions are feasible has remained unresolved for the case where the target distribution is not strictly positive on all vertices. While only strictly positive target distributions have been considered in previous work on control of swarms governed by CTMCs (Berman *et al.*, 2009), we address the more general case in which the target densities of some states can be zero. This question was addressed in (Acikmese and Bayard, 2012) for swarms governed by DTMCs. The problem has also been investigated in the context of consensus protocols (Chapman, 2015) for strictly positive distributions. In our controller synthesis, we will relax the assumption of strict positivity for desired target distributions.

Next, we address the feedback stabilization problem for system (2.2). Consider the following system:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e\in\mathscr{E}} k_e(\mathbf{x}(t))\mathbf{B}_e\mathbf{x}(t), \quad t\in[0,\infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}). \end{aligned} \tag{2.17}$$

Problem 2.3.2. ***(Closed-loop stabilization)*** *Given* $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$*, determine whether there exists a decentralized feedback law, defined as a collection of maps* $\tilde{k}_e : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ *where* $k_e(\mathbf{y}) = \tilde{k}_e(y_{S(e)}, y_{T(e)})$ *for each* $e \in \mathscr{E}$ *and* $\mathbf{y} \in \mathbb{R}^M$*, such that for the closed-loop system* (2.17)*,* $\mathbf{x}^d$ *is asymptotically stable and* $\tilde{k}_e(x^d_{S(e)}, x^d_{T(e)}) = 0$ *for each* $e \in \mathscr{E}$ *whenever* $x^d_{S(e)} > 0$.

Remark 2.3.3. ***(The significance of the condition*** $\tilde{k}_e(x^d_{S(e)}, x^d_{T(e)}) = 0$***)*** *Solutions to Problem 2.3.2 can be inferred from solutions of Problem 2.3.1* ***only if*** *the constraint* $\tilde{k}_e(x^d_{S(e)}, x^d_{T(e)}) =$

0 *is not imposed on the control laws. If this constraint were not imposed, then agents' transition rates would not necessarily be equal to* 0 *when the forward equation* (2.17) *reaches equilibrium. Hence, even when the system reaches equilibrium from a "macroscopic point of view," at the microscopic level, agents would still keep switching between vertices of the graph.*

2.3.1 Stabilization of Distributions with Strongly Connected Supports using Open-loop Control

We now address the feasibility of Problem 2.3.1.

Proposition 2.3.4. *Let $\mathscr{G}$ be a strongly connected graph. Suppose that $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$ is an initial distribution and $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ is a desired distribution. Additionally, assume that $\mathbf{x}^d$ has strongly connected support. Then there is a set of parameters, $a_e \in [0,\infty)$ for each $e \in \mathscr{E}$, such that if $u_e(t) = a_e$ for all $t \in [0,\infty)$ and for each $e \in \mathscr{E}$ in system* (2.2)*, then the solution $\mathbf{x}(t)$ of this system satisfies $\|\mathbf{x}(t) - \mathbf{x}^d\| \leq Me^{-\lambda t}$ for all $t \in [0,\infty)$ and for some positive parameters M and λ that are independent of $\mathbf{x}^0$.*

Proof. Let $\mathscr{V}_s \subset \mathscr{V}$ be the support of $\mathbf{x}^d$. From this vertex set, we construct a new graph $\tilde{\mathscr{G}} = (\mathscr{V}, \tilde{\mathscr{E}})$, where $e = (i,j) \in \mathscr{E}$ implies that $e \in \tilde{\mathscr{E}}$ if and only if $i \in \mathscr{V}_s$ implies that $j \notin \mathscr{V} \backslash \mathscr{V}_s$. Then it follows from (Chapman, 2015)[Proposition 10] that the process generated by the transition rate matrix $-\mathscr{L}_{out}(\tilde{\mathscr{G}})^T$ has a unique, globally stable invariant distribution if we can establish that $\tilde{\mathscr{G}}$ has a *rooted in-branching* subgraph. This implies that $\tilde{\mathscr{G}}$ must have a subgraph $\tilde{\mathscr{G}}_{sub} = (\mathscr{V}, \mathscr{E}_{sub})$ which has no directed cycles and for which there exists a root node, v_r, such that for every $v \in \mathscr{V}$ there exists a directed path from v to v_r. This is indeed true for the graph $\tilde{\mathscr{G}}$, which can be shown as follows. First, let $r \in \mathscr{V}$ such that $x_r^d > 0$. From the assumption that $\mathscr{G}$ is strongly connected and the construction of $\tilde{\mathscr{G}}$, it can be concluded that there exists a directed path in $\tilde{\mathscr{E}}$ from any $v \in \mathscr{V}$ to r. Now, for each

$n \in \mathbb{Z}_+$, the set of positive integers, let $\mathscr{N}_n(r)$ be the set of all vertices for which there exists a directed path of length n to r. For each $n > 1$, let $\tilde{\mathscr{N}}_n(r) = \mathscr{N}_n(r) \setminus \cup_{m=1}^{n-1} \mathscr{N}_m(r)$. We define $\mathscr{E}_{sub}$ by setting $e \in \mathscr{E}_{sub}$ if and only if $e \in \tilde{\mathscr{E}}$, $S(e) \in \tilde{\mathscr{N}}_n(r)$, and $T(e) \in \tilde{\mathscr{N}}_{n-1}(r)$ for some $n > 1$. Then $\tilde{\mathscr{G}}_{sub} = (\mathscr{V}, \mathscr{E}_{sub})$ is the desired rooted in-branching subgraph.

The matrix $-\mathscr{L}_{out}(\tilde{\mathscr{G}})^T$ is the generator of a CTMC, since $\mathscr{L}_{out}(\tilde{\mathscr{G}})^T \mathbf{1} = \mathbf{0}$ and its off-diagonal entries are positive. Moreover, as we have shown, $\tilde{\mathscr{G}}$ has a rooted in-branching subgraph. Hence, there exists a unique vector $\mathbf{z}$ such that $-\mathscr{L}_{out}(\tilde{\mathscr{G}})\mathbf{z} = \mathbf{0}$ and $\mathbf{z} \in \mathscr{P}(\mathscr{V})$. The vector $\mathbf{z}$ is nonzero only on $\mathscr{V}_s$, since the subgraph corresponding to $\mathscr{V}_s$ is strongly connected. Then we consider a positive definite diagonal matrix $\mathbf{D} \in \mathbb{R}^{M \times M}$ such that $D^{ii} = x_i / x_i^d$ if $i \in \mathscr{V}_s$ and an arbitrary strictly positive value for any other $i \in \mathscr{V}$. The matrix $-\mathbf{D}\mathscr{L}_{out}(\tilde{\mathscr{G}})^T$ is also the generator of a CTMC. Moreover, $\mathbf{x}^d$ is the unique stationary distribution of the process generated by $-\mathbf{D}\mathscr{L}_{out}(\tilde{\mathscr{G}})^T$, since $\mathbf{x}^d$ lies in the null space of $\mathbf{G} = -\mathscr{L}_{out}(\tilde{\mathscr{G}})\mathbf{D}$ by construction. The simplicity of the principal eigenvalue at 0 for the matrix $-\mathbf{D}\mathscr{L}_{out}(\tilde{\mathscr{G}})^T$ is inherited by the same eigenvalue of the matrix $\mathbf{G}$. Then the result follows by setting $a_e = G^{T(e)S(e)}$ for each $e \in \mathscr{E}$ and by noting that since $\mathbf{G}^T$ is the generator of a CTMC, and its eigenvalue at zero has the aforementioned properties and is simple, then the rest of the spectrum of $\mathbf{G}$ lies in the open left half of the complex plane. □

Next, we address Problem 2.3.2.

2.3.2 Stabilization of Distributions with Strongly Connected Supports using Linear Feedback Laws

In this subsection, we consider the possibility of stabilization using linear feedback laws. The motivation behind considering linear feedback laws is that this type of controller is a well-studied class of feedback laws for which there exists a rich literature on computational approaches for synthesis and design. Moreover, in contrast to stabilization using

open-loop controls in Section 2.3.1, the controls take zero value at equilibrium and thus prevent unnecessary switching of agents between states at equilibrium.

Lemma 2.3.5. *Define* $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$. *For each* $e \in \mathscr{E}$ *and each* $\mathbf{y} \in \mathbb{R}^M$, *let* $k_e : \mathbb{R}^M \to (-\infty, \infty)$ *be given by* $k_e(\mathbf{y}) = x^d_{T(e)} y_{S(e)} - x^d_{S(e)} y_{T(e)}$ *in system* (2.17). *Then,* $\mathbf{x}^d$ *is locally exponentially stable on the space* $\mathscr{P}(\mathscr{V})$. *That is, there exists* $r > 0$ *such that* $\|\mathbf{x}^0 - \mathbf{x}^d\|_2 < r$ *and* $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$ *imply that the solution* $\mathbf{x}(t)$ *of system* (2.17) *satisfies the following inequality,*

$$\|\mathbf{x}(t) - \mathbf{x}^d\|_2 \leq M_0 e^{-\lambda t}, \tag{2.18}$$

for all $t \in [0, \infty)$ *and for some parameters* $M_0 > 0$ *and* $\lambda > 0$ *that depend only on r. If* $\mathscr{G}$ *is bidirected, then* $\mathbf{x}^d$ *is also asymptotically stable.*

Proof. We use the linearization of system (2.2) about $\mathbf{x}^d$ to establish local exponential stability. Consider the vector field $\mathbf{f}^e = [f^e_1 \ f^e_2 \ \ldots \ f^e_M]^T$ given by

$$f^e_i(\mathbf{y}) = \begin{cases} -(x^d_{T(e)} y_{S(e)} - x^d_{S(e)} y_{T(e)}) y_{S(e)} & \text{if } i = S(e), \\ (x^d_{T(e)} y_{S(e)} - x^d_{S(e)} y_{T(e)}) y_{S(e)} & \text{if } i = T(e), \\ 0 & \text{otherwise} \end{cases}$$

for each $\mathbf{y} \in \mathbb{R}^M$. Then for each $e \in \mathscr{E}$, we define the matrix $\mathbf{A}_e \in \mathbb{R}^M \times \mathbb{R}^M$ as follows:

$$A^{ij}_e = \begin{cases} \left.\frac{\partial f^e_{S(e)}}{\partial y_{S(e)}}\right|_{\mathbf{y}=\mathbf{x}^d} = -x^d_{T(e)} x^d_{S(e)} & \text{if } i = j = S(e), \\ \left.\frac{\partial f^e_{S(e)}}{\partial y_{T(e)}}\right|_{\mathbf{y}=\mathbf{x}^d} = (x^d_{S(e)})^2 & \text{if } i = S(e), j = T(e), \\ \left.\frac{\partial f^e_{T(e)}}{\partial y_{T(e)}}\right|_{\mathbf{y}=\mathbf{x}^d} = -(x^d_{S(e)})^2 & \text{if } i = j = T(e), \\ \left.\frac{\partial f^e_{T(e)}}{\partial y_{S(e)}}\right|_{\mathbf{y}=\mathbf{x}^d} = x^d_{T(e)} x^d_{S(e)} & \text{if } i = T(e), j = S(e), \\ 0 & \text{otherwise.} \end{cases}$$

Now we define the matrix $\mathbf{G} \in \mathbb{R}^{M \times M}$ as $\mathbf{G} = \sum_{e \in \mathscr{E}} \mathbf{A}_e$. Note that $G^{S(e)T(e)} > 0$ for each $e \in \mathscr{E}$, since $\mathbf{x}^d \in \text{int } \mathscr{P}(\mathscr{V})$. Moreover, $\mathbf{1}^T \mathbf{G} = \mathbf{0}$, and the off-diagonal terms of $\mathbf{G}$ are

positive. Hence, $\mathbf{G}$ is an irreducible transition rate matrix. It is a classical result that this implies that $\mathbf{G}$ has its principal eigenvalue at 0, which is simple. The other eigenvalues of $\mathbf{G}$ lie in the open left half of the complex plane. However, note that the equilibrium point $\mathbf{x}^d$ is *non-hyperbolic*, since the principal eigenvalue of $\mathbf{G}$ is at 0. Hence, local exponential stability of the original nonlinear system does not immediately follow. However, it follows that there exists an $(M-1)-$dimensional local stable manifold of the system that is tangential to $\mathscr{P}(\mathscr{V})$ at $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$. Noting that the set $\{\mathbf{y} \in \mathbb{R}^M; \sum_{i=1}^M y_i = c\}$ is invariant for solutions of the system (2.17) for any $c \in \mathbb{R}$, it follows that the stable manifold is in fact in $\mathscr{P}(\mathscr{V})$. From this, the result follows.

To prove asymptotic stability of $\mathbf{x}^d$ for bidirected graphs, consider the continuously differentiable function $V : \mathbb{R}^M \to \mathbb{R}_{\geq 0}$ given by

$$V(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{x}^d)^T \mathbf{D}(\mathbf{y} - \mathbf{x}^d) \tag{2.19}$$

for all $\mathbf{y} \in \mathbb{R}^M$, where $\mathbf{D} \in \mathbb{R}^{M \times M}$ is defined as $\mathbf{D} = [\mathrm{diag}(\mathbf{x}^d)]^{-1}$. Then

$$\dot{V}(\mathbf{x}(t)) = \sum_{e \in \mathscr{E}} x_{S(e)}(t)(x^d_{T(e)} x_{S(e)}(t) - x^d_{S(e)} x_{T(e)}(t))^2.$$

Thus, $\dot{V}(\mathbf{x}(t)) \leq 0$ for all $t \in [0, \infty)$, with the equality $\dot{V}(\mathbf{x}(t)) = 0$ holding only when $\mathbf{x}(t) = \mathbf{x}^d$. Then, the asymptotic stability of $\mathbf{x}^d$ follows from *LaSalle's invariance principle* (Khalil, 2001) by noting that the set $\mathscr{P}(\mathscr{V})$ is invariant for the system (2.2). □

The above lemma implies that *if negative transition rates are admissible*, then there exists a linear feedback law, $\{k_e\}_{e \in \mathscr{E}}$, such that $k_e(\mathbf{x}^d) = 0$ for each $e \in \mathscr{E}$ and the desired equilibrium point is locally exponentially stable.

A desirable property of the control system (2.2) is that stabilization of the target equilibrium can be achieved using a linear feedback law that satisfies positivity constraints away from equilibrium and is zero at equilibrium. However, any stabilizing linear control law that is zero at equilibrium and is additionally non-negative everywhere must in

fact be zero everywhere. To see this explicitly, suppose that $\varepsilon = [\varepsilon_1 \ ... \ \varepsilon_M]^T \in \mathbb{R}^M$ is a nonzero element such that $\mathbf{x}^{eq} \pm \varepsilon \in \mathscr{P}(\mathscr{V})$, and suppose that $k_e(\mathbf{x})$ is a linear control law. Then the control law has the form $k_e(\mathbf{x}) = \sum_{i\in\mathscr{V}} a_e^i x_i + b_e$, where a_e^i and b_e are gain parameters. Since the control inputs must take the value 0 at equilibrium, we must have that $b_e = -\sum_{i\in\mathscr{V}} a_e^i x_i^{eq}$. Suppose, for the sake of contradiction, that this linear control law satisfies the positivity constraints; that is, the range of $k_e(\mathbf{x})$ is $[0,\infty)$ for some $e \in \mathscr{E}$. Then we must have that $k_e(\mathbf{x}^{eq} + \varepsilon) = \sum_{i\in\mathscr{V}} a_e^i (x_i^{eq} + \varepsilon_i) + b_e = \sum_{i\in\mathscr{V}} a_e^i \varepsilon_i > 0$. This must imply that $k_e(\mathbf{x}^{eq} - \varepsilon) = \sum_{i\in\mathscr{V}} a_e^i (x_i^{eq} - \varepsilon_i) + b_e = -\sum_{i\in\mathscr{V}} a_e^i \varepsilon_i < 0$, which contradicts the original assumption that the control law $k_e(\mathbf{x})$ satisfies the positivity constraints. Hence, to ensure that the control laws satisfy the positivity constraints, we replace them with rational feedback control laws $c_e(\mathbf{x})$ that produce the same closed-loop system trajectories but respect the positivity constraints, as desired.

On the other hand, in the next theorem we show that whenever $\mathscr{G}$ is bidirected, any feedback control law that violates positivity constraints can be implemented using a rational feedback law of the form $k(\mathbf{x}) = a(\mathbf{x}) + b(\mathbf{x})\frac{f(\mathbf{x})}{g(\mathbf{x})}$, such that $k(\mathbf{x})$ satisfies the positivity constraints and is zero at equilibrium.

Theorem 2.3.6. *Let $\mathscr{G}$ be a bidirected graph. Let $k_e : \mathbb{R}^M \to (-\infty,\infty)$ be a map for each $e \in \mathscr{E}$ such that there exists a unique global solution of the system* (2.17)*. Additionally, assume that $\mathbf{x}(t) \in \operatorname{int} \mathscr{P}(\mathscr{V})$ for each $t \in [0,\infty)$. Consider the functions $m_e^p : \mathbb{R}^M \to \{0,1\}$ and $m_e^n : \mathbb{R}^M \to \{0,1\}$, defined as follows for each $e \in \mathscr{E}$:*

$$\begin{aligned} m_e^p(\mathbf{y}) &= 1 \quad \text{if} \quad k_e(\mathbf{y}) \geq 0, \quad 0 \ \text{otherwise;} \\ m_e^n(\mathbf{y}) &= 1 \quad \text{if} \quad k_e(\mathbf{y}) \leq 0, \quad 0 \ \text{otherwise.} \end{aligned}$$

Let $c_e : \mathbb{R}^M \to [0,\infty)$ be given by

$$c_e(\mathbf{y}) = m_e^p(\mathbf{y}) k_e(\mathbf{y}) - m_{\tilde{e}}^n(\mathbf{y}) k_{\tilde{e}}(\mathbf{y}) \frac{y_{T(e)}}{y_{S(e)}}. \tag{2.20}$$

Then the solution $\tilde{\mathbf{x}}(t)$ *of the following system,*

$$\begin{aligned} \dot{\tilde{\mathbf{x}}} &= \sum_{e\in\mathscr{E}} c_e(\tilde{\mathbf{x}}(t))\mathbf{B}_e\tilde{\mathbf{x}}(t), \quad t\in[0,\infty), \\ \tilde{\mathbf{x}}(0) &= \mathbf{x}^0 \in \text{int } \mathscr{P}(\mathscr{V}), \end{aligned} \tag{2.21}$$

is unique, defined globally, and satisfies $\tilde{\mathbf{x}}(t) = \mathbf{x}(t)$ *for all* $t \in [0,\infty)$.

Proof. This follows by noting that the right-hand sides of systems (2.2) and (2.17) are equal for all $t \geq 0$. □

We now extend the stabilization results in Lemma 2.3.5 and Theorem 2.3.6 to the more general case where the target distribution has a strongly connected support and is not necessarily strictly positive everywhere on $\mathscr{V}$. We will need the following preliminary results to prove these extensions.

Proposition 2.3.7. *Let* $\mathbf{A} \in \mathbb{R}^{M\times M}$ *be an essentially non-negative matrix. Let* $\mathscr{S}$ *be the set of elements* k *in* $\mathscr{V}$ *such that* $\sum_{i\in\mathscr{V}} A^{ik} < 0$. *Assume that* $\mathscr{S}$ *is non-empty and that* $\sum_{i\in\mathscr{V}} A^{ij} \leq 0$ *for all* $j \in \mathscr{V}$. *Additionally, suppose that for each* $j \in \mathscr{V}\backslash\mathscr{S}$, *there exists a sequence* $(i_n)_{n=1}^m \in \mathscr{V}$ *of length* m *such that* $i_1 = j$, $i_m \in \mathscr{S}$, *and* $A^{i_k i_{k-1}} > 0$ *for all* $i_k \neq i_{k-1}$ *with* $k = 2,...,m$. *Then* spec$(\mathbf{A})$ *lies in the open left half of the complex plane.*

Proof. First, we will confirm that spec$(\mathbf{A})$ lies in the closed left half of the complex plane. Toward this end, let $\lambda > 0$ be large enough such that $\lambda\mathbf{I}+\mathbf{A}$ is a non-negative matrix, where $\mathbf{I}$ is the $M \times M$ identity matrix. Since each column sum of the matrix $\lambda\mathbf{I}+\mathbf{A}$ is less than or equal to λ, it follows from (Minc, 1988)[Theorem 4.2] and (Minc, 1988)[Theorem 1.1] that the maximal eigenvalue r of $\lambda\mathbf{I}+\mathbf{A}$ exists and is bounded from above by λ. Next, we will establish that $r \neq \lambda$. Suppose, for the sake of contradiction, that the maximal eigenvalue of $\lambda\mathbf{I}+\mathbf{A}$ is λ, and hence that $\mathbf{A}$ has an eigenvalue at 0. Then, by (Minc, 1988)[Theorem 4.2], there exists a nonzero element of $\mathbf{v} \in \mathbb{R}^M_{\geq 0}$ such that $\mathbf{Av} = \mathbf{0}$. Therefore, $\mathbf{1}^T\mathbf{Av} =$

$\sum_{i\in V}\sum_{k\in\mathscr{S}} A^{ik}v_k = 0$. Hence, since each column of $\mathbf{A}$ corresponding to $\mathscr{V}\backslash\mathscr{S}$ sums to 0, we can conclude that $(\mathbf{Av})_k = 0$ for all $k \in \mathscr{S}$. Additionally, we assumed that if $j \notin \mathscr{S}$, then there exists a sequence $(i_n)_{n=1}^m \in \mathscr{V}$ of length m such that $i_1 = j$, $i_m \in \mathscr{S}$, and $A^{i_k i_{k-1}} > 0$ for all $i_k \neq i_{k-1}$ with $k = 2,...,m$. Moreover, all the off-diagonal elements of $\mathbf{A}$ are non-negative, and $\mathbf{Av} = \mathbf{0}$. Thus, it must be the case that $v_i = 0$ for all $i \in \mathscr{N}(j)$, the set of vertices that are adjacent to any vertex $j \in \mathscr{S}$. The non-negativity of the off-diagonal elements of $\mathbf{A}$ and the fact that $\mathbf{Av} = \mathbf{0}$ also imply that $v_i = 0$ for all $i \in \mathscr{N}(p)$, for all $p \in \mathscr{N}(k)$ with $k \in \mathscr{S}$. Using a similar argument, we can show that since the graph $\mathscr{G}$ is strongly connected, $v_i = 0$ for all $i \in \mathscr{V}$. This implies that $r \neq \lambda$. Therefore, the matrix $\mathbf{A}$ is Hurwitz. This concludes the proof. □

Theorem 2.3.8. *Let $\mathbf{f}: \mathbb{R}^{M_1} \to \mathbb{R}^{M_1}$ be a Lipschitz-continuous vector field, where M_1 is the cardinality of a set $\mathscr{V}_1 \subset \mathscr{V}$. Also, let M_2 be the cardinality of $\mathscr{V}_2 = \mathscr{V}\backslash\mathscr{V}_1$. Suppose there exists a continuously differentiable positive semidefinite function $U: \mathbb{R}^{M_2} \to \mathbb{R}_{\geq 0}$ such that $\frac{\partial U}{\partial \mathbf{y}}\mathbf{f}(\mathbf{y}) \leq 0$, with the equalities $U(\mathbf{y}) = \frac{\partial U}{\partial \mathbf{y}}\mathbf{f}(\mathbf{y}) = 0$ holding only at a unique fixed point of $\mathbf{f}(\mathbf{x})$ given by $\mathbf{y} = \mathbf{x}^d \in \mathscr{P}(\mathscr{V}_1)$. Now consider the following system with solution $\mathbf{z}(t) \in \mathbb{R}^M$,*

$$\begin{aligned} \dot{\mathbf{z}}_1(t) &= \mathbf{f}(\mathbf{z}_1(t)) + \mathbf{G}_2\mathbf{z}_2(t), \\ \dot{\mathbf{z}}_2(t) &= \mathbf{A}\mathbf{z}_2(t), \\ \mathbf{z}(0) &= \mathbf{z}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{2.22}$$

where $\mathbf{z}(t) = [\mathbf{z}_1(t)^T\ \mathbf{z}_2(t)^T]^T$, $\mathbf{G}_2 \in \mathbb{R}^{M\times M_2}$, $\mathbf{A} \in \mathbb{R}^{M_2\times M_1}$, $\mathscr{V}$ has cardinality $M > M_1$, and $\mathscr{P}(\mathscr{V})$ is invariant for the system. Lastly, assume that the matrix $\mathbf{A}$ satisfies the sufficient conditions in Proposition (2.3.7) *for* spec($\mathbf{A}$) *to lie in the open left half of the complex plane. Then $\mathbf{z}^d = [(\mathbf{x}^d)^T\ \ \mathbf{0}^T]^T$ is the globally asymptotically stable equilibrium point of the system* (2.22).

Proof. From the proof of Proposition (2.3.7), the matrix $\mathbf{A}$ is Hurwitz. This implies that $\lim_{t\to\infty}\mathbf{z}_2(t) = \mathbf{0}$. Hence, $\lim_{t\to\infty}\sum_{i\in\mathscr{V}_1}(\mathbf{z}_1)_i(t) = 1$, since $\mathscr{P}(\mathscr{V})$ is invariant for the system

(2.22). We can extend the function U to a function $\hat{U}$ on $\mathbb{R}^M$ by defining $\hat{U}(\mathbf{y}) = U(\mathbf{y}_1)$, where $\mathbf{y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T]^T$, $\mathbf{y}_1 \in \mathbb{R}^{M_1}$, and $\mathbf{y}_2 \in \mathbb{R}^{M_2}$. Consider the set $\Delta_c = \{\mathbf{y} \in \mathscr{P}(\mathscr{V}) : \sum_{i \in \mathscr{V}_2} y_i \leq c\}$. From the assumptions made on U, we have that $\frac{\partial U}{\partial \mathbf{y}_1}\mathbf{f}(\mathbf{y}_1) + \frac{\partial U}{\partial \mathbf{y}_1}\mathbf{G}_2\mathbf{y}_2 \leq 0$ on the set Δ_0, with the equality holding only at $\mathbf{y} = [(\mathbf{x}^d)^T \ \ \mathbf{0}^T]^T$. Now fix $c_1 > 0$. By the continuity of the function $\hat{U}_d(\mathbf{y}) \coloneqq \frac{\partial U}{\partial \mathbf{y}_1}\mathbf{f}(\mathbf{y}_1) + \frac{\partial U}{\partial \mathbf{y}_1}\mathbf{G}_2\mathbf{y}_2$, there exist $\varepsilon > 0$ and $c_2 > 0$ such that $\hat{U}_d(\mathbf{y}) \leq -\varepsilon$ for all $\mathbf{y} \in U^{-1}\big((c_1, \infty]\big) \cap \Delta_{c_2}$. Due to the assumption on the matrix $\mathbf{A}$ that $\sum_{i \in \mathscr{V}} A^{ij} \leq 0$ for all $j \in \mathscr{V}$, it follows that $U^{-1}\big([0, c_1]\big) \cap \Delta_{c_2}$ is invariant for the system (2.22). This implies that the equilibrium $\mathbf{x}^d$ is Lyapunov stable for the system (2.22). Next, we will establish that the distribution $\mathbf{x}^d$ is also globally attractive. We know that $\lim_{t \to \infty} \mathbf{z}_2(t) = \mathbf{0}$. Since $\mathscr{P}(\mathscr{V})$ is compact, we can conclude that there exists $t_0 \geq 0$ such that $\mathbf{z}(t) \in U^{-1}\big([0, c_1]\big) \cap \Delta_{c_2}$ for all $t \geq t_0$. The constant c_1 can be chosen to be arbitrarily small. This implies that $\lim_{\to \infty}$ $\mathbf{z}(t) = \mathbf{x}^d$. □

Using the results in Proposition 2.3.7 and Theorem 2.3.8, we prove the following result, which generalizes Lemma 2.3.5 to target distributions that have a strongly connected support.

Theorem 2.3.9. *Let $\mathscr{G}$ be a bidirected graph. Suppose that $\mathbf{x}^d \in \mathscr{P}(\mathscr{V})$ has a strongly connected support. Let $\mathscr{V}_1$ be the support of $\mathbf{x}^d$ and $\mathscr{V}_2 = \mathscr{V} \setminus \mathscr{V}_1$. Let $k_e : \mathbb{R}^M \to [0, \infty)$ be defined as*

$$k_e(\mathbf{x}) = \begin{cases} a_1(x^d_{T(e)} y_{S(e)} - x^d_{S(e)} y_{T(e)}) & \text{if } S(e), T(e) \in \mathscr{V}_1, \\ g_e \in (0, \infty) & \text{if } S(e) \in \mathscr{V}_2, \\ 0 & \text{if } S(e) \in \mathscr{V}_1, T(e) \in \mathscr{V}_2. \end{cases}$$

Then $\mathbf{x}^d$ is the globally asymptotically stable equilibrium point of the system (2.17).

Proof. Without loss of generality, we can assume that $\mathscr{V}_1$ is of the form $\{1, ..., M_1\}$ for

some $M_1 \geq M$. We rewrite system (2.17) as

$$\dot{\mathbf{x}}(t) = \mathbf{G}(\mathbf{x}(t))\mathbf{x}(t), \;\; \mathbf{x}(0) = \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}) \tag{2.23}$$

with $\mathbf{G} : \mathbb{R}^M \to \mathbb{R}^{M \times M}$ given by $\mathbf{G}(\mathbf{y}) = \sum_{e \in \mathscr{E}} k_e(\mathbf{y}) B_e$ for all $\mathbf{y} \in \mathbb{R}^M$. Since $k_e(\mathbf{y}) = 0$ whenever $S(e) \in \mathscr{V}_1, \; T(e) \in \mathscr{V}_2$, the state-dependent matrix $\mathbf{G}$ can be factorized into the form

$$\mathbf{G}(\mathbf{y}) = \begin{bmatrix} \mathbf{G}_1(\mathbf{y}_1) & \mathbf{G}_2 \\ \mathbf{0} & \mathbf{A} \end{bmatrix}, \tag{2.24}$$

where $\mathbf{G}_1 : \mathbb{R}^{M_1} \to \mathbb{R}^{M_1 \times M_1}$ and $\mathbf{G}_2 \in \mathbb{R}^{M_1 \times M_2}$. Moreover, since the graph $\mathscr{G}$ is strongly connected and bidirected, from the definition of k_e, it follows that $\mathbf{A}$ satisfies the sufficient conditions of Proposition 2.3.7; therefore, spec($\mathbf{A}$) lies in the open left half of the complex plane. In addition, since each column of the matrix $\mathbf{G}(\mathbf{y})$ sums to 0 and this matrix is essentially non-negative for each $\mathbf{y} \in \mathscr{P}(\mathscr{V})$, the set $\mathscr{P}(\mathscr{V})$ is invariant for the system (2.23). Let M_1 be the cardinality of the set $\mathscr{V}_1$. Additionally, define the function $U : \mathbb{R}^{M_1} \to \mathbb{R}_{\geq 0}$ by $U(\mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{y}^d)^T \mathbf{D}(\mathbf{y} - \mathbf{y}^d)$ for all $\mathbf{y} \in \mathbb{R}^{M_1}$, where $\mathbf{y}^d \in$ int $\mathscr{P}(\mathscr{V}_1)$ such that $\mathbf{x}^d = [(\mathbf{y}^d)^T \; \mathbf{0}^T]^T \in \mathscr{P}(\mathscr{V})$, and $\mathbf{D} \in \mathbb{R}^{M_1 \times M_1}$ is given by $\mathbf{D} = [\text{diag}(\mathbf{x}^d)]^{-1}$. By Lemma 2.3.5, this function satisfies the conditions of Theorem 2.3.8 with respect to the vector field $\mathbf{f}(\mathbf{z}) = \mathbf{G}_1^T(\mathbf{z})\mathbf{z}$ on the set $\mathscr{P}(\mathscr{V}_1)$. Then the result follows from Theorem 2.3.8. □

2.3.3 Stabilization of Probability Distributions with Disconnected Supports

In the previous subsection, we were able to only stabilize probability distributions that have a strongly connected support. The goal in this subsection is to consider the case when the target distribution is an arbitrary element of $\mathscr{P}(\mathscr{V})$, thus including the possibility that the support of the probability distribution is not strongly connected. Toward this end, we define a general class of control laws under which the resulting closed-loop system (2.17) will have the desired probability distribution as a globally asymptotically stable equilibrium point.

Define $k_e : [0,1] \to [0, u_{\max}]$ as

$$k_e(y) = \begin{cases} c_e(y - x^{eq}_{S(e)}) & \text{if } y > x^{eq}_{S(e)} \\ 0 & \text{otherwise} \end{cases} \tag{2.25}$$

where $c_e : [0, 1 - x^{eq}_{S(e)}] \to [0, u_{\max}]$ is a positive-valued function for each $e \in \mathscr{E}$, and $u_{\max} > 0$ is the upper bound on the transition rate parameters. For each $e \in \mathscr{E}$, we make the following **assumptions** on the function c_e:

1. The inequality $c_e(y) > 0$ is satisfied for all $y \in (0, 1 - x^{eq}_{S(e)}]$.
2. The function c_e is non-decreasing on $[0, 1 - x^{eq}_{S(e)}]$.
3. The function c_e is locally Lipschitz continuous at every point in $[0, 1 - x^{eq}_{S(e)}]$, except for a finite number of points, and right-continuous with left limits at every point in $[0, 1 - x^{eq}_{S(e)}]$.
4. The set of points in $[0, 1 - x^{eq}_{S(e)}]$ at which c_e is discontinuous is finite.
5. If $c_{e_1}(0) > 0$ for some $e_1 \in \mathscr{E}$, then $c_{e_2}(0) > 0$ for all $e_2 \in \mathscr{E}$ such that $S(e_1) = S(e_2)$.

Due to the above assumptions on the function c_e, the right-hand side of the ODE (2.17) can be discontinuous. Hence, the classical solution of the ODE (2.17) might not exist in general. Therefore, we will consider a generalized notion of solutions using Filippov's theory for ODEs with discontinuous right-hand sides (Filippov, 2013). Toward this end, we define the set-valued map $\mathbf{F} : \mathscr{P}(\mathscr{V}) \rightrightarrows \mathbb{R}^M$, also known as the *Krasovskii regularization* of the vector field $\mathbf{f}(\mathbf{x}) = \sum_{e \in \mathscr{E}} k_e(x_{S(e)}) \mathbf{B}_e \mathbf{x}$, as:

$$\mathbf{F}(\mathbf{x}) = \cap_{\delta > 0} \, \bar{\text{co}} \, \{\mathbf{f}(\mathbf{y}) : \ \mathbf{y} \in \mathbb{R}^M \ \& \ \|\mathbf{x} - \mathbf{y}\| \leq \delta\} \tag{2.26}$$

for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$. We will also need the set-valued map $\tilde{\mathbf{F}} : \mathscr{P}(\mathscr{V}) \rightrightarrows \mathbb{R}^M$ defined by

$$\tilde{\mathbf{F}}(\mathbf{x}) = \left\{ \lim_{j\to\infty} \mathbf{f}(\mathbf{x}^j) : \ \lim_{j\to\infty} \mathbf{x}^j \to \mathbf{x} \ \& \ \lim_{j\to\infty} \mathbf{f}(\mathbf{x}^j) \text{ exists} \right\} \tag{2.27}$$

for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$. Then $\tilde{\mathbf{F}}$ and $\mathbf{F} = \bar{\text{co}}\ \tilde{\mathbf{F}}$ are upper-semicontinuous, closed, and bounded at each $\mathbf{x} \in \mathscr{P}(\mathscr{V})$ (Filippov, 2013)[Lemma 1, Pg. 67]. Let $\mathscr{L} = \{+,-\}^M$. With each $\ell \in \mathscr{L}$, we associate the set-valued map $\tilde{\mathbf{F}}_\ell : \mathscr{P}(\mathscr{V}) \rightrightarrows \mathbb{R}^M$,

$$\tilde{\mathbf{F}}_\ell(\mathbf{x}) = \left\{ \mathbf{f}_\ell(\mathbf{x}) \right\} = \left\{ \sum_{e\in\mathscr{E}} k_e^{\ell_{S(e)}}(x_{S(e)}) \mathbf{B}_e \mathbf{x} \right\} \tag{2.28}$$

for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$, where $k_e^+(y)$ and $k_e^-(y)$ denote the right limit and left limit, respectively, of $k_e(y)$ at $y \in [0,1]$. Since the function k_e accepts $x_{S(e)}$ as its argument, the directional limits of the vector field $\mathbf{f}$ at $\mathbf{x} \in \mathscr{P}(\mathscr{V})$ are determined completely by the right and left limits of the function k_e at $x_{S(e)}$. Moreover, due to the assumption of right-continuity of the functions c_e at every $x \in [0, 1 - x_{S(e)}^{eq}]$, we can infer that $\tilde{\mathbf{F}}(\mathbf{x}) = \cup_{\ell\in\mathscr{L}} \tilde{\mathbf{F}}_\ell(\mathbf{x})$ for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$. From the definition of the set-valued map $\mathbf{F}$, it follows that $\mathbf{F}(\mathbf{x})$ is convex for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$. Note that $\mathscr{P}(\mathscr{V})$ is a convex and closed set. Whenever the limits $\lim_{j\to\infty} \mathbf{x}^j \to \mathbf{x}$ and $\lim_{j\to\infty} \mathbf{f}(\mathbf{x}^j)$ exist for some $\mathbf{x} \in \mathscr{P}(\mathscr{V})$ and sequence $\{\mathbf{x}^j\}$ in $\mathscr{P}(\mathscr{V})$, $\lim_{j\to\infty} \mathbf{f}(\mathbf{x}^j)$ lies in $T_{\mathbf{x}}\mathscr{P}(\mathscr{V})$, the *tangent space* of $\mathscr{P}(\mathscr{V})$ at $\mathbf{x}$,

$$T_{\mathbf{x}}\mathscr{P}(\mathscr{V}) = \left\{ \mathbf{y} \in \mathbb{R}^M : \ \sum_{v\in\mathscr{V}} y_v = 0 \ \ \& \ \ y_w \geq 0 \text{ whenever } x_w = 0 \text{ for } w \in \mathscr{V} \right\}. \tag{2.29}$$

This leads to the following observation.

Proposition 2.3.10. *Let* $\mathbf{F}$ *be the set-valued map defined in Equation* (2.26)*. Then,*

$$\begin{aligned} \mathbf{F}(\mathbf{x}) = & \ \cap_{\delta>0} \ \bar{\text{co}} \ \{\mathbf{f}(\mathbf{y}) : \ \mathbf{y} \in \mathscr{P}(\mathscr{V}) \ \& \ \|\mathbf{x}-\mathbf{y}\| \leq \delta\} \\ = & \ \bar{\text{co}} \ \{ \lim_{h\to 0^+} \mathbf{f}(\mathbf{x}+h\mathbf{y}) : \mathbf{y} \in T_{\mathbf{x}}\mathscr{P}(\mathscr{V})\} \end{aligned}$$

for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$.

For a given $T > 0$, a *generalized solution* or simply *solution* of the ODE (2.17) will refer to an absolutely continuous function $\mathbf{x} : [0,T] \to \mathbb{R}^M$ such that the following *Differential Inclusion* (DI) is satisfied,

$$\dot{\mathbf{x}}(t) \in \mathbf{F}(\mathbf{x}(t)), \tag{2.30}$$

for almost every (a.e.) $t \in [0,T]$ and $\mathbf{x}(0) = \mathbf{x}^0$. We will be interested only in those solutions $\mathbf{x}(t)$ that are *viable* in $\mathscr{P}(\mathscr{V})$, meaning that $\mathbf{x}(t) \in \mathscr{P}(\mathscr{V})$ for all $t \geq 0$. In the context of this subsection, only viable solutions are physically meaningful since the density of agents in any state (vertex) cannot be negative. Hence, we will first establish that for a given $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$, at least one global viable solution of the system (2.30) (and hence a generalized solution of system (2.17)) exists.

Theorem 2.3.11. ***(Viability)*** *Given $\mathbf{x}^0 \in \mathscr{P}(\mathscr{V})$, there exists at least one global viable solution of the system* (2.17)*.*

Proof. We define the *contingent cone* (Aubin and Frankowska, 2009) of the set $\mathscr{P}(\mathscr{V})$ at a point $\mathbf{z} \in \mathscr{P}(\mathscr{V})$ as

$$T_-(\mathbf{z}) = \left\{ \mathbf{y} \in \mathbb{R}^M : \liminf_{h \to 0^+} \frac{\operatorname{dist}(\mathbf{z} + h\mathbf{y}, \mathscr{P}(\mathscr{V}))}{h} = 0 \right\}. \tag{2.31}$$

where $\operatorname{dist}(\mathbf{x}, A) := \sup_{\mathbf{p} \in A} \{\|\mathbf{x} - \mathbf{p}\|\}$ for each $\mathbf{x} \in \mathbb{R}^M$ and $A \subseteq \mathbb{R}^M$ Then, from (Aubin and Frankowska, 2009)[Lemma 4.2.4], we know that $T_-(\mathbf{z}) = T_{\mathbf{z}}\mathscr{P}(\mathscr{V})$ for all $\mathbf{z} \in \mathscr{P}(\mathscr{V})$. Moreover, $\mathbf{F}$ is upper-semicontinuous, closed, and compact-valued, and it is defined on a closed domain $\mathscr{P}(\mathscr{V})$. From Proposition 2.3.10, it follows that $\mathbf{F}(\mathbf{z}) \subset T_-(\mathbf{z})$ for all $\mathbf{z} \in \mathscr{P}(\mathscr{V})$. Hence, it follows from the *Local Viability Theorem* (Aubin and Frankowska, 2009)[Theorem 10.1.4] that there exists a solution $\mathbf{x} : [0, t_f] \to \mathscr{P}(\mathscr{V})$ of the DI (2.30) that is viable in $\mathscr{P}(\mathscr{V})$ for some $t_f > 0$, i.e., a *local viable* solution exists. Since $\mathbf{F}(\mathbf{z})$ is uniformly bounded for all $\mathbf{z} \in \mathscr{P}(\mathscr{V})$ and $\mathscr{P}(\mathscr{V})$ is a compact subset of $\mathbb{R}^M$, we can take $t_f = \infty$ (Aubin and Frankowska, 2009)[Theorem 10.1.4], and hence $\mathbf{x}(t)$ can be extended to a global viable solution. □

In the following theorem, we note that the derivative of any solution of the DI (2.30) can be expressed as a convex combination of elements in $F(\mathbf{x}(t))$ for a.e. $t \geq 0$ and that this representation can be constructed using measurable functions. The theorem and its proof are minor modifications of the statement and proof of the *Carathéodory representation theorem* (Aubin and Frankowska, 2009)[Theorem 8.2.15], and are adapted for our purposes.

Lemma 2.3.12. *Let* $\mathbf{x} : [0,\infty) \to \mathscr{P}(\mathscr{V})$ *be a global viable solution of the DI* (2.30)*. Then there exist measurable functions* $\lambda_v^+ : [0,\infty) \to \mathbb{R}_{\geq 0}$, $\lambda_v^- : [0,\infty) \to \mathbb{R}_{\geq 0}$ *for each* $v \in \mathscr{V}$ *such that*

$$\dot{\mathbf{x}}(t) = \sum_{e\in\mathscr{E}} \lambda^+_{S(e)}(t) k_e^+(x_{S(e)}(t))\mathbf{B}_e\mathbf{x}(t) + \sum_{e\in\mathscr{E}} \lambda^-_{S(e)}(t) k_e^-(x_{S(e)}(t))\mathbf{B}_e\mathbf{x}(t)$$

and

$$\sum_{v\in\mathscr{V}} \lambda_v^+(t) + \sum_{v\in\mathscr{V}} \lambda_v^-(t) = 1 \tag{2.32}$$

for a.e. $t \in [0,\infty)$.

Proof. Suppose that $\mathbf{x} : [0,\infty) \to \mathscr{P}(\mathscr{V})$ is a solution of the DI (2.30). We define the set $Q = \{\mathbf{y} \in \mathbb{R}^{2^M}_{\geq 0} : \ \sum_{i=1}^{2^M} y_i = 1\}$. Let $\mathscr{I} : \{1,...,2^M\} \to \{+,-\}^M$ be a bijective map, i.e., an ordering on $\{+,-\}^M$. Then consider the map $h : \mathbb{R}^{2^M}_{\geq 0} \times (\mathbb{R}^M)^{2^M} \to \mathbb{R}^M$ defined by

$$h(\gamma_1,...,\gamma_{2^M},\mathbf{y}_1,...,\mathbf{y}_{2^M}) = \sum_{i=1}^{2^M} \gamma_i \mathbf{y}_i \tag{2.33}$$

and the measurable set-valued map $\mathbf{H} : [0,\infty) \rightrightarrows \mathbb{R}^{2^M}_{\geq 0} \times (\mathbb{R}^M)^{2^M}$ defined by

$$\mathbf{H}(t) = Q \times \tilde{\mathbf{F}}_{\mathscr{I}(1)}(\mathbf{x}(t)) \times ... \times \tilde{\mathbf{F}}_{\mathscr{I}(2^M)}(\mathbf{x}(t)) \tag{2.34}$$

for all $t \in [0,\infty)$. We recall that $\mathbf{F}(\mathbf{x}(t)) = \bar{\text{co}}\ \tilde{\mathbf{F}}(\mathbf{x}(t)) = \cup_{\ell\in\mathscr{L}} \tilde{\mathbf{F}}_\ell(\mathbf{x}(t))$ for all $t \in [0,\infty)$. Hence, $\dot{\mathbf{x}}(t) \in g(t,\mathbf{H}(t))$ for a.e. $t \in [0,\infty)$, where $g(t,\mathbf{z}) = h(\mathbf{z})$ for all $\mathbf{z} = (\gamma_1,...,\gamma_{2^M},\mathbf{y}_1,...,\mathbf{y}_{2^M})^T \in \mathbb{R}^{2^M}_{\geq 0} \times (\mathbb{R}^M)^{2^M}$. The map $g(t,\mathbf{z})$ is a *Carathéodory map*, i.e.,

for every $t \in [0,\infty)$ the map $\mathbf{z} \mapsto g(t,\mathbf{z})$ is continuous and for every $\mathbf{z} \in \mathbb{R}^{2^M}_{\geq 0} \times (\mathbb{R}^M)^{2^M}$ the map $t \mapsto g(t,\mathbf{z})$ is measurable. Then it follows from the *inverse image theorem* (Aubin and Frankowska, 2009)[Theorem 8.2.9] that there exists a measurable map $t \mapsto (\gamma_1(t),...,\gamma_{2^M}(t),\mathbf{y}_1(t),...,\mathbf{y}_{2^M}(t))^T$ such that

$$\dot{\mathbf{x}}(t) = g(t,(\gamma_1(t),...,\gamma_{2^M}(t),\mathbf{y}_1(t),...,\mathbf{y}_{2^M}(t))^T) = h(\gamma_1(t),...,\gamma_{2^M}(t),\mathbf{y}_1(t),...,\mathbf{y}_{2^M}(t))$$

for a.e. $t \in [0,\infty)$. From this the result follows. □

Remark 2.3.13. *Henceforth, in the following results, when we refer to the functions* $\lambda_v^+ : [0,\infty) \to \mathbb{R}_+$, $\lambda_v^- : [0,\infty) \to \mathbb{R}_+$ *for* $v \in \mathscr{V}$, *we will mean measurable functions such that equation* (2.32) *in Lemma 2.3.12 is satisfied for a given solution* $\mathbf{x}(t)$ *of the DI* (2.30).

In the following lemma, we establish some monotonicity properties of the solutions of the DI (2.30). In particular, if the agent density in a given state is below the desired value over a certain time interval, then it is non-decreasing since the outflow of agents from the state is zero over that time interval. This lemma lies at the heart of the proof of the main stability theorem (Theorem 2.3.17).

Lemma 2.3.14. *Suppose that* $\mathbf{x} : [0,T] \to \mathscr{P}(\mathscr{V})$ *is a local viable solution of the DI* (2.30) *for a given* $T > 0$, *and that* $x_v(t) < x_v^{eq}$ *for all* $t \in [0,T]$. *Then* $x_v(t)$ *is non-decreasing over the time interval* $[0,T]$.

Proof. Let $\mathbf{x} : [0,T] \to \mathscr{P}(\mathscr{V})$ be a local viable solution of the DI (2.30). Then $x_v(t)$ is differentiable almost everywhere on $t \in [0,T]$. Suppose $\dot{x}_v(s)$ exists for some $s \in [0,T]$. Note that $k_e^+(x_v(t)) = k_e^-(x_v(t)) = 0$ for all $t \in [0,T]$ and for all e such that $S(e) = v$. This fact, along with the assumption that $x_v(t) < x_v^{eq}$ for all $t \in [0,T]$, implies that $\dot{x}_v(s) \geq 0$. The

result that $x_v(t)$ is non-decreasing for $t \in [0,T]$ follows by noting that

$$\begin{aligned} x_v(t) &= x_v^0 + \int_0^t \dot{x}_v(s)ds \\ &= \sum_{p\in\{+,-\}} \sum_{w\in\mathscr{N}^{\text{in}}(v)} \int_0^t \lambda_w^p(\tau) k_{(w,v)}^p(x_w(\tau)) x_w(\tau) d\tau \\ &\quad - \sum_{p\in\{+,-\}} \sum_{w\in\mathscr{N}^{\text{out}}(v)} \int_0^t \lambda_v^p(\tau) k_{(v,w)}^p(x_v(\tau)) x_v(\tau) d\tau \end{aligned}$$

for all $t \in [0,T]$. □

If the function k_e is continuous at the origin, then the stability theorem (Theorem 2.3.17) can be directly proved using the above lemma. To account for the possibility of discontinuity of $k_e(x_v)$ at $x_v = 0$ for some $e \in \mathscr{E}$, we prove the following proposition.

Proposition 2.3.15. *Let* $\mathbf{x} : [T_1,T_2] \to \mathscr{P}(\mathscr{V})$ *be a local viable solution of the system* (2.17) *such that* $x_v^{eq} \le x_v(t) < x_v^{eq} + \varepsilon$ *for all* $t \in [T_1,T_2]$, *for some* $T_2 > T_1 > 0$, $v \in \mathscr{V}$, *and* $\varepsilon > 0$. *Additionally, assume that* $c_e(0) > 0$ *for some (and hence all)* $e \in \mathscr{E}$ *such that* $S(e) = v$. *Suppose that there exists* $z \in \mathscr{N}^{\text{in}}(v)$ *such that* $\int_{T_1}^{T_2} \lambda_z^+(\tau) k_{(z,v)}^+(x_z(\tau)) x_z(\tau) d\tau + \int_{T_1}^{T_2} \lambda_z^-(\tau) k_{(z,v)}^-(x_z(\tau)) x_z(\tau) d\tau > 2\varepsilon$. *Then there exists a constant* $C_v > 0$, *which depends only on* $v \in \mathscr{V}$, *such that* $\int_{T_1}^{T_2} \lambda_v^+(\tau) k_{(v,w)}^+(x_v(\tau)) x_v(\tau) d\tau + \int_{T_1}^{T_2} \lambda_v^-(\tau) k_{(v,w)}^-(x_v(\tau)) x_v(\tau) d\tau > C_v\varepsilon$ *for all* $w \in \mathscr{N}^{\text{out}}(v)$.

Proof. From the assumed bounds on $x_v(t)$ over the time-interval $[T_1,T_2]$, we can conclude that $\int_{T_1}^{T_2} \dot{x}_v(\tau)d\tau \le \varepsilon$. Hence, it follows that

$$\begin{aligned} &x_v(T_2) - x_v(T_1) = \int_{T_1}^{T_2} \dot{x}_v(\tau)d\tau = \\ &\sum_{p\in\{+,-\}} \sum_{w\in\mathscr{N}^{\text{in}}(v)} \int_{T_1}^{T_2} \lambda_w^p(\tau) k_{(w,v)}^p(x_w(\tau)) x_w(\tau) d\tau - \\ &\sum_{p\in\{+,-\}} \sum_{w\in\mathscr{N}^{\text{out}}(v)} \int_{T_1}^{T_2} \lambda_v^p(\tau) k_{(v,w)}^p(x_v(\tau)) x_v(\tau) d\tau \ < \ \varepsilon. \end{aligned}$$

Since $\int_{T_1}^{T_2} \lambda_z^+(\tau) k_{(z,v)}^+(x_z(\tau)) x_z(\tau) d\tau + \int_{T_1}^{T_2} \lambda_z^-(\tau) k_{(z,v)}^-(x_z(\tau)) x_z(\tau) d\tau > 2\varepsilon$, we can conclude that

$$\sum_{p\in\{+,-\}} \sum_{w\in\mathscr{N}^{\text{out}}(v)} \int_{T_1}^{T_2} \lambda_v^p(\tau) k_{(v,w)}^p(x_v(\tau)) x_v(\tau) d\tau \;>\; \varepsilon.$$

From this, it follows that

$$\max_{w\in\mathscr{N}^{\text{out}}(v)} \sum_{p\in\{+,-\}} \int_{T_1}^{T_2} \lambda_v^p(\tau) k_{(v,w)}^p(x_v(\tau)) x_v(\tau) d\tau \;>\; \frac{\varepsilon}{|\mathscr{N}^{\text{out}}(v)|},$$

where $|\mathscr{N}^{\text{out}}(v)|$ represents the number of outgoing edges from v. Let $c_{\max} = \max\limits_{w\in\mathscr{N}^{\text{out}}(v)} \{k_{(v,w)}^+(1)\}$ and $c_{\min} = \min\limits_{w\in\mathscr{N}^{\text{out}}(v)} \{k_{(v,w)}^+(x_v^{eq})\}$. Then it follows that

$$\sum_{p\in\{+,-\}} \int_{T_1}^{T_2} \lambda_v^p(\tau) k_{(v,w)}^p(x_v(\tau)) x_v(\tau) d\tau \;>\; \frac{c_{\min}}{c_{\max}} \frac{\varepsilon}{|\mathscr{N}^{\text{out}}(v)|}$$

for all $w \in \mathscr{N}^{\text{out}}(v)$. Note that $c_{\min} \neq 0$ due to the assumption that $c_e(0) > 0$ for some (and hence all) $e \in \mathscr{E}$ such that $S(e) = v$. Hence, we have our result. □

The above proposition does not hold true if assumption 5 is not satisfied by all functions c_e. This can happen only when, for a given vertex $v \in \mathscr{V}$, the functions $c_e(y)$ are discontinuous at $y = 0$ for some but not all outgoing edges e from v. In fact, violation of this assumption can create spurious equilibrium solutions of the DI (2.30). This is highlighted in the following counterexample.

Example 2.3.16. *Let $\mathscr{V} = \{1,2,3\}$ and $\mathscr{E} = \{(1,2),(2,1),(2,3),(3,2)\}$. Suppose $\mathbf{x}^{eq} = [0.5\ 0.5\ 0]^T$. Let $c_{(1,2)}$ be an arbitrary function with the appropriate domain and range satisfying assumptions 1-5. The other functions c_e are defined as*

$$c_{(2,1)}(y) = y \ \textit{ for all } \ y \in [0,0.5]$$

$$c_{(2,3)}(y) = 1 \ \textit{ for all } \ y \in [0,0.5]$$

$$c_{(3,2)}(y) = 1 \ \textit{ for all } \ y \in [0,1]$$

Then $\mathbf{x} = [0\ 0.5\ 0.5]^T$ *is an equilibrium solution of the DI* (2.30)*, that is,* $\mathbf{0} \in \mathbf{F}(\mathbf{x})$*. This is true because* $k^+_{(1,2)}(x_1) = k^+_{(2,1)}(x_2) = 0$ *and* $k^+_{(2,3)}(x_2)x_2 - k^+_{(3,2)}(x_3)x_3 = 0$*. Hence,* $\sum_{e\in\mathscr{E}} k^+_e(\mathbf{x})\mathbf{B}_e\mathbf{x} = \mathbf{0}$*. Note that* $\mathbf{x}$ *is not an equilibrium point of the original system* (2.17) *because* $\sum_{e\in\mathscr{E}} k_e(\mathbf{x})\mathbf{B}_e\mathbf{x} \neq \mathbf{0}$*.*

Now, we are ready to prove our main result.

Theorem 2.3.17. *Let* $\mathbf{x}^0, \mathbf{x}^{eq} \in \mathscr{P}(\mathscr{V})$*. Then a global viable solution* $\mathbf{x} : [0,\infty) \to \mathscr{P}(\mathscr{V})$ *of the DI* (2.30) *exists. Moreover, the equilibrium point* $\mathbf{x}^{eq}$ *is asymptotically stable with respect to all global viable solutions of the DI* (2.30)*.*

Proof. The existence of global viable solutions has been already established (Theorem 2.3.11). Lyapunov stability of the equilibrium point $\mathbf{x}^{eq}$ follows from Lemma 2.3.14 and by noting that $\mathbf{x}(t) \in \mathscr{P}(\mathscr{V})$ for all $t \geq 0$. Suppose, for the sake of contradiction, that the limit condition $\lim_{t\to\infty} \|\mathbf{x}(t) - \mathbf{x}^{eq}\| = 0$ is not satisfied by a global viable solution. Then there exists $v_1 \in \mathscr{V}$ such that $\lim_{t\to\infty} x_{v_1}(t) \neq x^{eq}_{v_1}$. Since $\mathbf{x}(t) \in \mathscr{P}(\mathscr{V})$ for all $t \geq 0$, and from the monotonicity property of the components of the solution proved in Lemma 2.3.14, we can assume that the vertex $v_1 \in \mathscr{V}$ is such that $x_{v_1}(t) > x^{eq}_{v_1}$ for all $t \geq T$, for some $T \geq 0$. Then there exists an increasing sequence of positive numbers $(T_n)_{n=1}^{\infty}$ such that $\lim_{n\to\infty} T_n = \infty$ and $x_{v_1}(T_n) > x^{eq}_{v_1} + \varepsilon_0$ for all $n \in \mathbb{Z}_+$ for some $\varepsilon_0 > 0$ independent of n. Note that $|\dot{x}_{v_1}(t)| \leq C u_{\max}$ for a.e. $t \in [0,\infty)$, for some constant $C > 0$. Hence, there exists $\Delta T > 0$ such that $x_{v_1}(t) > x^{eq}_{v_1} + \frac{\varepsilon_0}{2}$ for all $t \in [T_n, T_n + \Delta T]$ and all $n \in \mathbb{Z}_+$. Now we consider a subsequence of $(T_n)_{n=1}^{\infty}$. We use the same notation $(T_n)_{n=1}^{\infty}$ to denote this new subsequence, and choose this subsequence such that $T_{n+1} - T_n > \Delta T$ for all $n \in \mathbb{Z}_+$. Let $\tilde{T}_n = T_n + \Delta T$ for all $n \in \mathbb{Z}_+$. From this and the assumption that c_e is non-decreasing on $[0, 1 - x^{eq}_{v_1}]$, it follows that $\sum_{p\in\{+,-\}} \int_{T_n}^{\tilde{T}_n} \lambda^p_{v_1}(\tau) k^p_e(x_{v_1}(\tau)) x_{v_1}(\tau) d\tau > \varepsilon_1$ for some $\varepsilon_1 > 0$, for all $e \in \mathscr{E}$ such that $S(e) = v_1$, and for all $n \in \mathbb{Z}_+$.

Next, let $\mu = (e_i)_{i=1}^m$ be a directed path from the node $S(e_1) = v_1$ to some node $T(e_m) =$

v_{m+1} such that $\lim_{t\to\infty} x_{v_{m+1}}(t) < x^{eq}_{v_{m+1}}$ and $\lim_{t\to\infty} x_{v_g}(t) = x^{eq}_{v_g}$ with $v_g = S(e_g)$ for all $g \in \{2,...,m\}$. Since the graph $\mathscr{G}$ is strongly connected, and from the result in Lemma 2.3.14, such a path necessarily exists. Now, there are two possibilities. Either there exists some $j \in \{2,...,m\}$ such that $k^+_{e_j}(x^{eq}_{S(e_j)})x^{eq}_{S(e_j)} = 0$ for some $j \in \{2,...,m\}$, or such a j does not exist. We will consider the first possibility and show that such a j cannot exist due to the assumption made on the path μ, and then consider the second possibility. Let j be the smallest element of $\{2,...,m\}$ such that $k^+_{e_j}(x^{eq}_{S(e_j)})x^{eq}_{S(e_j)} = 0$. We know that

$$\sum_{p\in\{+,-\}} \int_{T_n}^{\tilde{T}_n} \lambda^p_{v_1}(\tau) k^p_{e_1}(x_{v_1}(\tau)) x_{v_1}(\tau) d\tau > \varepsilon_1 \tag{2.35}$$

for some $\varepsilon_1 > 0$ and for all $n \geq N$. It follows from Proposition 2.3.15 that if N is large enough, then since $\lim_{t\to\infty} x_{v_2}(t) = x^{eq}_{v_2}$, we have that $\sum_{p\in\{+,-\}} \int_{T_n}^{\tilde{T}_n} \lambda^p_{v_2}(\tau) k^p_{e_2}(x_{v_2}(\tau)) x_{v_2}(\tau) d\tau > \varepsilon_2$ for some $\varepsilon_2 > 0$ depending only on ε_1, for all $n \geq N$. Using the same argument, it follows that if N is large enough, then since $\lim_{t\to\infty} x_{v_g}(t) = x^{eq}_{v_g}$ for all $g = \{3,...,j-1\}$, we have that $\sum_{p\in\{+,-\}} \int_{T_n}^{\tilde{T}_n} \lambda^p_w(\tau) k^p_e(x_{v_g}(\tau)) x_{v_g}(\tau) d\tau > \varepsilon_g$ for some $\varepsilon_g > 0$ depending only on ε_1, for all $n \geq N$ and for all $g = \{2,...,j-1\}$. This implies that if N is large enough,

$$\begin{aligned}
&\int_{T_n}^{\tilde{T}_n} \dot{x}_w(\tau) d\tau \; = \\
&\quad \sum_{p\in\{+,-\}} \sum_{a\in\mathscr{N}^{\text{in}}(w)} \int_{T_n}^{\tilde{T}_n} \lambda^p_a(\tau) k^p_{(a,w)}(x_a(\tau)) x_a(\tau) d\tau \\
&- \sum_{p\in\{+,-\}} \sum_{a\in\mathscr{N}^{\text{out}}(w)} \int_{T_n}^{\tilde{T}_n} \lambda^p_w(\tau) k^p_{(w,a)}(x_w(\tau)) x_w(\tau) d\tau \\
&\quad > \; \varepsilon_{j-1} - \delta_n
\end{aligned} \tag{2.36}$$

for all $n \geq N$, with $w = S(e_j)$. Here, $\delta_n > 0$ is an n-dependent constant, yet to be defined, that satisfies the inequality $\sum_{p\in\{+,-\}} \sum_{a\in\mathscr{N}^{\text{out}}(w)} \int_{T_n}^{\tilde{T}_n} \lambda^p_w(\tau) k^p_{(w,a)}(x_w(\tau)) x_w(\tau) d\tau < \delta_n$ for all $n \in \mathbb{Z}_+$. Since $k^+_{(w,a)}(x^{eq}_w)x^{eq}_w = 0$ for all $a \in \mathscr{N}^{\text{out}}(w)$ and $\lim_{t\to\infty} x_w(t) = x^{eq}_w$, we know that δ_n can be chosen such that $\lim_{n\to\infty} \delta_n = 0$. This last observation and the inequality (2.36) lead to a

contradiction that $x_w(\tilde{T}_n) > \varepsilon_{j-1} - \delta_n > 0$ for all $n \geq N$ if N is large enough. Hence, the second possibility must be true; i.e., that there exists no $j \in \{2,...,m\}$ such that $k^+_{e_j}(x^{eq}_{v_j})x^{eq}_{v_j} = 0$. This implies that k_{e_j} must be discontinuous at $x^{eq}_{S(e_j)}$, with $k^+_{e_j}(x^{eq}_{v_j})x^{eq}_{v_j} > 0$ for each $j \in \{2,...,m\}$. Then Proposition 2.3.15 implies that $\sum_{p\in\{+,-\}} \int_{T_n}^{\tilde{T}_n} \lambda^p_{v_g}(\tau) k^p_{e_g}(x_{v_g}(\tau)) x_{v_g}(\tau) d\tau > \varepsilon_g$ for some $\varepsilon_g > 0$ depending only on ε_1, for all $g \in \{2,...,m\}$, and for all $n \geq N$ if N is large enough. This contradicts the assumption that $\lim_{t\to\infty} x_{v_{m+1}}(t) < x^{eq}_{v_{m+1}}$ for all $t \geq 0$. Hence, it must be true that $\lim_{t\to\infty} x_{v_1}(t) = x^{eq}_{v_1}$. □

Simulations

In this subsection, we numerically verify the effectiveness of the decentralized feedback controllers that are defined in Lemma 2.3.5 (the *linear controller*) and Theorem 2.3.17. The controllers were constructed to redistribute populations of $N = 80$ and $N = 1200$ agents on the six-vertex bidirected graph shown in Fig. 2.4. In all cases, the initial distribution of agents was set to $\mathbf{x}^0 = [0.2\ \ 0.1\ \ 0.2\ \ 0.15\ \ 0.2\ \ 0.15]^T$, and the desired distribution was $\mathbf{x}^d = [0.1\ \ 0.2\ \ 0.05\ \ 0.25\ \ 0.15\ \ 0.25]^T$. For both feedback controllers, the numerical solution of the mean-field model (2.17) was compared to stochastic simulations of the CTMC characterized by expression (2.1). This CTMC was simulated using an approximating DTMC that evolves in discrete time. The probability that an agent i in state (vertex) $S(e)$, $e \in \mathscr{E}$, at time t transitions to state $T(e)$ at time $t + \Delta t$ was set to:

$$\mathbb{P}(X_i(t+\Delta t) = T(e) | X_i(t) = S(e)) = k_e\left(\frac{1}{N}\mathbf{N}^p(t)\right)\Delta t.$$

Here, $\{k_e\}_{e\in\mathscr{E}}$ is the set of feedback laws and $\mathbf{N}^p(t) = [N^p_1(t)\ N^p_2(t)\ ...\ N^p_M(t)]^T$, where $N^p_v(t)$ is the number of agents in state $v \in \mathscr{V}$ at time t. We assume that each agent can measure the agent populations in its current state and in adjacent states.

In Figs. 2.5 and 2.6, we compare simulations of the closed-loop system (2.17) with the feedback controllers to simulations of the open-loop system (2.2). The controller for the

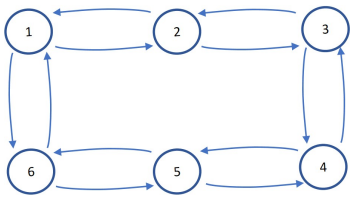
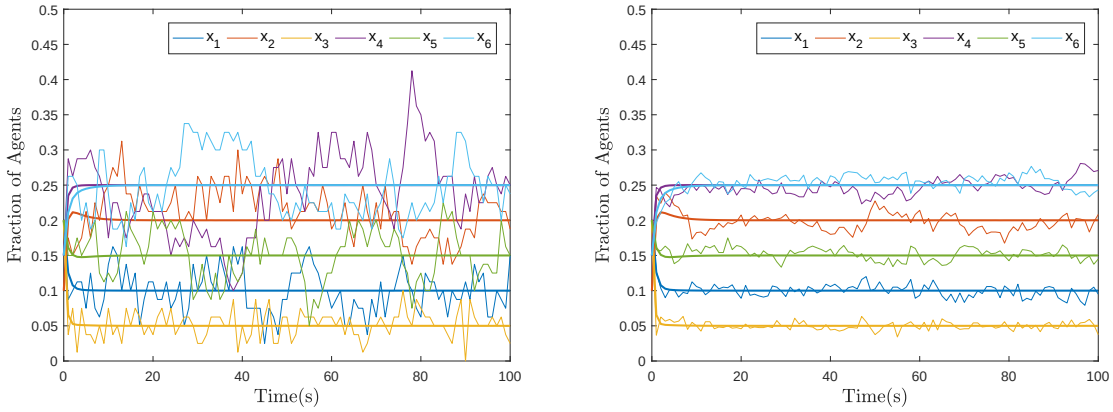


Figure 2.4: Six-vertex Bidirected Graph.

open-loop system was constructed by setting the right-hand side of system (2.2) equal to $\mathbf{Gx} = -\mathbf{L}(\mathscr{G})\mathbf{Dx}$, where $\mathbf{L}(\mathscr{G})$ is the Laplacian matrix of the graph $\mathscr{G}$ and $\mathbf{D}$ is a diagonal matrix with entries $D^{ij} = 1/x_i^d$ if $i = j$, $D^{ij} = 0$ otherwise. This makes the desired distribution $\mathbf{x}^d$ invariant for the corresponding CTMC. The transition rates (control inputs) for this controller were defined as $u_e(t) = G^{T(e)S(e)}$ for all $t \in [0,\infty)$, $e \in \mathscr{E}$. Fig. 2.5 shows that the open-loop controller produces large variances in the agent populations at steady-state. As an expected consequence of the law of large numbers, these variances are smaller for $N = 1200$ agents than for $N = 80$ agents. In comparison, the variances are much smaller when the feedback controllers are used, as shown in Fig. 2.6. This is due to the property of the feedback controllers that as the agent densities approach their desired equilibrium values, the transition rates tend to zero. This property reduces the number of unnecessary agent state transitions at equilibrium. Using open-loop control, the agents' states keep switching and never reach steady-state values. In contrast, using the feedback controllers, the agents' states remain constant after a certain time.

As discussed in beginning of this section, the underlying assumption of using the mean-field models (2.2) and (2.17) is that the swarm behaves like a continuum. That is, the ODEs (2.2) and (2.17) are valid as number of agents $N \to \infty$. Hence, it is imperative to check the performance of the feedback controllers for different agent populations. We observe in Fig. 2.6b that in the case of $N = 1200$ agents, the stochastic simulation follows the mean-field model solution quite closely for both feedback controllers. In addition, in all simulations, the numbers of agents in each state remain constant after some time; in the case of $N = 80$



(a) Open-loop system, $N = 80$

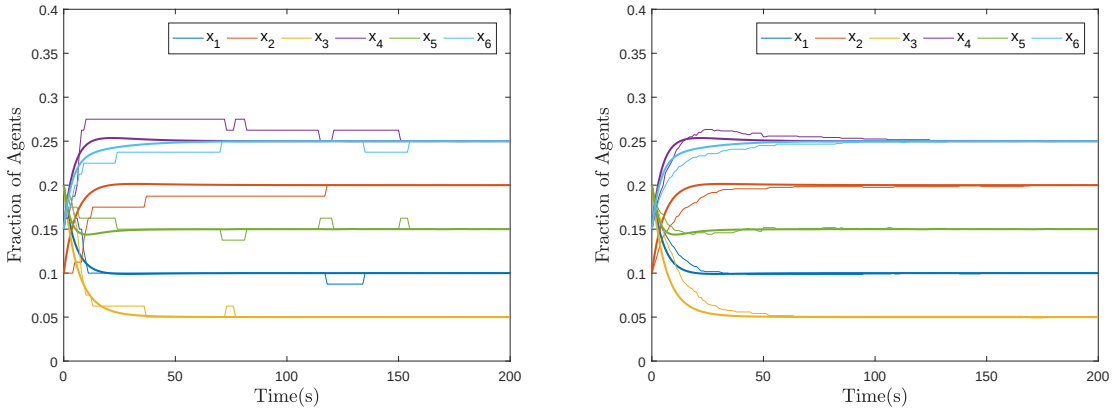
(b) Open-loop system, $N = 1200$

Figure 2.5: Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*.

agents, the fluctuations stop earlier than in the case of $N = 1200$ agents.

Next, we numerically verify the effectiveness of the decentralized feedback strategy presented in Section 2.3.3 in two scenarios with different graph topologies and agent population sizes. In the first scenario, we redistribute $N = 60$ agents over a directed 6-vertex cycle graph with $\mathscr{V} = \{1,...,5\}$, $\mathscr{E} = \{(v, v+1) : v \in \mathscr{V}\} \cup \{(6,1)\}$. The initial distribution of agents was set to $\mathbf{x}^0 = [0.2 \;\; 0.1 \;\; 0.2 \;\; 0.15 \;\; 0.2 \;\; 0.15]^T$, and the target distribution was $\mathbf{x}^{eq} = [0.25 \;\; 0 \;\; 0.25 \;\; 0 \;\; 0.25 \;\; 0.25]^T$. Note that the target fractions of agents are zero for two states. Figs. 2.7a and 2.7b compare the solution of the mean-field model (2.17) to a stochastic simulation of the CTMC characterized by expression (2.1) for two different control laws that we design according to equation (2.25).

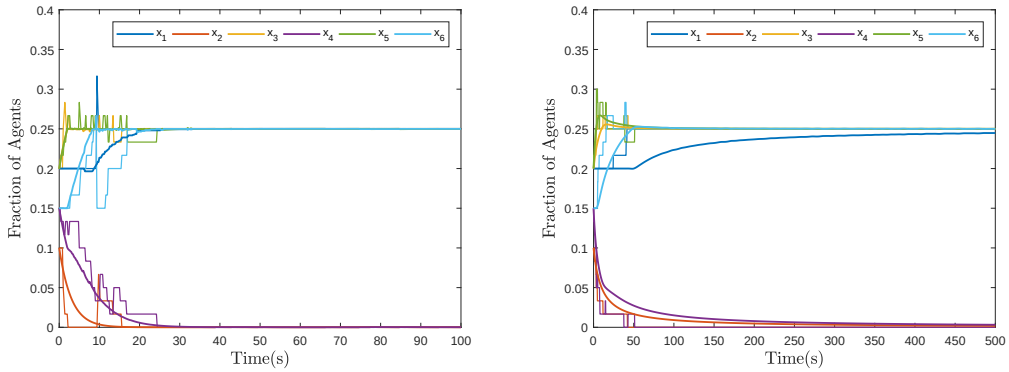
In Fig. 2.7a, we have used a discontinuous control law $\{k_e(\cdot)\}$ by setting $c_e(y) = 1/S(e)$ for all $y \in [0, 1 - x^{eq}_{S(e)}]$. We call this control law *controller 1*. As shown in the figure, the transitions exhibit chattering behavior that is typical of discontinuous control laws. Also, as a consequence of the transition rates not tending to zero near the equilibrium, the agents can transition between states with a high probability even near equilibrium. On the other



(a) Closed-loop system with linear controller, $N = 80$

(b) Closed-loop system with linear controller, $N = 1200$

Figure 2.6: Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*.



(a) Closed-loop system with *controller 1*, $N = 60$

(b) Closed-loop system with *controller 2*, $N = 60$

Figure 2.7: Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*.

hand, in Fig. 2.7b, we have used a Lipschitz continuous law $\{k_e(\cdot)\}$ by setting $c_e(y) = y$ for all $y \in [0, 1 - x^{eq}_{S(e)}]$. We call this control law *controller 2*. The fractions of agents in each state exhibit fewer fluctuations. The figures show that the stochastic simulation follows the mean-field model solution fairly closely for both feedback controllers. In addition, the fractions of agents in each state remain constant after some time.

2.4 Controllability and Stabilization of a Model for Herding a Swarm using a Leader

In this section, we will consider the controllability and stabilization problem for herding a swarm of agents using a single leader agent. The leader agent performs a sequence of deterministic transitions from one vertex to another. The leader's location at time t is denoted by $\ell(t) \in \mathscr{V}$.

The transition rates $u_e(t)$ are constrained in this section by the leader's location $\ell(t)$. In particular, for each $e \in \mathscr{E}$ and each $t \geq 0$, we set

$$\begin{aligned} u_e(t) = \quad & 1 + u^0_e(\mathbf{x}(t)) \text{ if } S(e) = \ell(t), \\ & u^0_e(\mathbf{x}(t)) \text{ otherwise} \end{aligned}$$

for a set of Lipschitz functions $u^0_e : \mathscr{P}(\mathscr{V}) \to \mathbb{R}_{\geq 0}$, which model inter-follower interactions. For example, the followers could have an attractive effect on each other, in which case the interaction could be modeled as $u^0_e(\mathbf{x}) = x_{T(e)}$. Alternatively, $u^0_e(\cdot)$ could model *congestion affects* by setting $u^0_e(\mathbf{x}) = 0$ whenever $x_{T(e)}$ exceeds some threshold value.

Then for a given leader trajectory $\ell : \mathbb{R}_{\geq 0} \to \mathscr{V}$, the system (2.2) can be rewritten as

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \sum_{e \in \mathscr{E}} u^0_e(\mathbf{x}(t)) \mathbf{B}_e \mathbf{x}(t) + D_{\ell(t)} \mathbf{x}(t), \quad t \in [0, \infty), \\ \mathbf{x}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{2.37}$$

where, for each $v \in \mathscr{V}$, the matrix $\mathbf{D}_v \in \mathbb{R}^{N \times N}$ is given by

$$\mathbf{D}_v = \sum_{e \in \mathscr{E}, S(e) = v} \mathbf{B}_e. \tag{2.38}$$

We make the following assumptions about the agents' capabilities for the case of non-interacting agents (i.e., $u_e^0 = 0$ for all $e \in \mathscr{E}$):

1. The leader can perfectly localize itself in $\mathscr{V}$; i.e., it knows its location $l(t) \in \mathscr{V}$ at each time t.

2. The leader can measure the density of follower agents $x_{l(t)}(t)$ that are at its current location $l(t)$ at time t.

3. Each follower can sense whether the leader is present at the follower's current location.

We can now state the control problems that we address in this section. The first problem relates to the controllability of system (2.37).

Problem 2.4.1. *Given a target probability distribution $\mathbf{x}^{eq} \in \mathscr{P}(\mathscr{V})$ among the states in $\mathscr{V}$, and a time $T > 0$, construct a trajectory $\ell : [0,T] \to \mathscr{V}$ of the leader agent such that $\mathbf{x}(T) = \mathbf{x}^{eq}$.*

After addressing the controllability problem, we will construct solutions for the following stabilization problem.

Problem 2.4.2. *Given a target probability distribution $\mathbf{x}^{eq} \in \mathscr{P}(\mathscr{V})$ among the states in $\mathscr{V}$, design the leader agent's trajectory $\ell : \mathbb{R}_{\geq 0} \to \mathscr{V}$ so that $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^{eq}$.*

2.4.1 Controllability

In this subsection, we will address Problem 2.4.1. It is a standard approach in control theory literature (Cheng, 2005; Sun *et al.*, 2002) to study controllability properties of switched systems of the form (2.37) using controllability properties of a related *relaxed system*. The controllability results in (Cheng, 2005; Sun *et al.*, 2002) are restricted to bilinear

systems. Since system (2.37) is not bilinear in general, we will perform our controllability analysis using the concept of *relaxed controls* (Young, 1980; Fattorini, 1999). The approach of using relaxed controls to study controllability properties of herding models was first performed in (Colombo and Pogodaev, 2012), where the authors studied the reachability properties of the differential inclusion based herding model that was initially presented in (Bressan and Zhang, 2012). In contrast to the models used in (Bressan and Zhang, 2012; Colombo and Pogodaev, 2012), where the swarm of followers was represented using a set, in this work the swarm is represented as a probability distribution. Following this approach, we first prove the controllability of the following relaxed system,

$$\begin{aligned} \dot{\mathbf{y}}(t) &= \sum_{e\in\mathscr{E}} u_e^0(\mathbf{y}(t))\mathbf{B}_e\mathbf{y}(t) + \sum_{v\in\mathscr{V}} \alpha_v(t)\mathbf{D}_v\mathbf{y}(t) \\ &\qquad\qquad t\in[0,\infty), \qquad (2.39) \\ \mathbf{y}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned}$$

where $\alpha_v(t)$ is a non-negative function for each $v\in\mathscr{V}$.

If system (2.39) is controllable with $\alpha(t) = [\alpha_1(t) \ldots \alpha_M(t)]^T$ as the control inputs, then it can concluded that system (2.37) is controllable. In order to establish controllability of the relaxed system (2.39), we will show that the span of the set $\cup_{v\in\mathscr{V}}\{\mathbf{D}_v\mathbf{x}\}$ is equal to $M-1$ for all $\mathbf{x}\in\mathscr{P}(\mathscr{V})$. To conclude this, we will use some spectral properties of $\mathbf{Q} := \sum_{v\in\mathscr{V}}\mathbf{D}_v = \sum_{e\in\mathscr{E}}\mathbf{B}_e$ that can be established using the *Perron-Frobenius theorem* (Berman and Plemmons, 1994) for positive matrices. These properties are stated in Lemma 2.4.3 below. Since the proof of this Lemma is standard in the literature (see for example (Berman *et al.*, 2009)[Theorem 1]), we omit it here. Here and in the following sections, we define int $\mathscr{P}(\mathscr{V}) = \{\mathbf{x}\in\mathscr{P}(\mathscr{V}); x_v > 0\ \forall v\in\mathscr{V}\}$.

Lemma 2.4.3. *The matrix* $\mathbf{Q}$ *has rank* $M-1$ *with* 0 *as its principal eigenvalue. Moreover, there exists* $\beta\in$ int $\mathscr{P}(\mathscr{V})$ *such that* $\mathbf{Q}\beta = \mathbf{0}$.

Lemma 2.4.4. *Let* $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$ *be an equilibrium point of the system* (2.39) *with steady-state control input* $\alpha^{ss} = [\alpha_1^{ss} \ ... \ \alpha_M^{ss}]^T \in \text{int } \mathscr{P}(\mathscr{V})$. *Let* $T > 0$ *be given. Then there exists a neigborhood* U *of* $\mathscr{P}(\mathscr{V})$, *such that for each* $\mathbf{x}^0 \in U$, *there exists a set of measurable functions* $\tilde{\alpha}_v : [0,T] \to [0,1]$ *such that* $\sum_{v\in\mathscr{V}} \tilde{\alpha}_v(t) = 1$ *for almost every* $t \in [0,1]$ *and the solution* $\mathbf{x}(t)$ *of the system* (2.39) *satisfies* $\mathbf{x}(T) = \mathbf{x}^{eq}$, *with* $\alpha_v(t) = \tilde{\alpha}_v(t) + \alpha_v^{ss}$ *for all* $v \in \mathscr{V}$ *and almost every* $t \in [0,T]$.

Proof. Fix $\mathbf{x} \in \text{int } \mathscr{P}(\mathscr{V})$. We will show that the set $A^{\mathbf{x}} = \{\sum_{v\in\mathscr{V}} \gamma_v \mathbf{D}_v \mathbf{x}; [\gamma_1 \ ... \ \gamma_M]^T \in \mathbb{R}^M, \ \sum_{v\in\mathscr{V}} \gamma_v = 0\}$ is an $(M-1)$-dimensional subspace of $\mathbb{R}^M$. This would imply that (2.39) is *locally controllable* at $\mathbf{x}$ on $\mathscr{P}(\mathscr{V})$, i.e., there is a neighborhood U of $\mathbf{x}$ in $\mathscr{P}(\mathscr{V})$ in which system (2.39) is controllable to $\mathbf{x}^{eq}$.

According to Lemma 2.4.3, the matrix $\mathbf{Q}$ has rank $M-1$. Moreover, there exists $\beta = [\beta_1 \ ... \ \beta_M] \in \text{int } \mathscr{P}(\mathscr{V})$, such that $\mathbf{Q}\beta = \mathbf{0}$. Note that $\mathbf{D}_v \mathbf{x} = x_v (D_v)^v$, where $(D_v)^v$ denotes the v^{th} column of $\mathbf{D}_v$. Therefore, we can conclude that $A_r^{\mathbf{x}} = \{\sum_{v\in\mathscr{V}} \gamma_v \mathbf{D}_v \mathbf{x}; [\gamma_1 \ ... \ \gamma_M]^T \in \mathbb{R}^M\}$ has dimension $M-1$. Let $\mathbf{y} = \sum_{v\in\mathscr{V}} \gamma_v \mathbf{D}_v \mathbf{x}$ be an element of $A_r^{\mathbf{x}}$ for some $[\gamma_1 \ ... \ \gamma_M] \in \mathbb{R}^M$. Suppose that $\sum_{v\in\mathscr{V}} \gamma_v = c$. Then setting $\eta_v = \gamma_v - \frac{c\beta_v}{x_v \sum_{w\in\mathscr{V}} \beta_w}$ for each $v \in \mathscr{V}$, we note that $\mathbf{y} = \eta_v \mathbf{D}_v \mathbf{x}$ and $\sum_{w\in\mathscr{V}} \eta_w = 0$. This implies that $A^{\mathbf{x}} = A_r^{\mathbf{x}}$ and hence the set A^r is an $(M-1)$-dimensional subspace of $\mathbb{R}^M$. This implies that there are sufficient number of *control directions* for system (2.39) on the $(M-1)$-dimensional submanifold $\mathscr{P}(\mathscr{V})$ in a neighborhood of $\mathbf{x}$. This concludes the proof. □

In order to prove the next result, we will need some new defintions and terminologies from measure theory (Bogachev, 2007). Let $C(\mathscr{V})$ denote the space of continuous functions on $\mathscr{V}$, with the standard discrete topology on $\mathscr{V}$. The space $L^1(0,T;C(\mathscr{V}))$ is defined by

$$L^1(0,T;C(\mathscr{V})) = \{f : (0,T) \to C(\mathscr{V}) \ \text{ is a measurable function}; \int_0^T \|f(t)\|_\infty dt < \infty\}$$

where $\|f(t)\|_\infty$ denotes the maximum of the function $f(t) \in C(\mathscr{V})$ attained over $\mathscr{V}$. We will also need the space $R(0,T;\mathscr{V})$, which will be used to denote the set of *relaxed controls*, i.e.,

the set of elements e for which $e(t)$ is probability measure on C for almost every $t \in (0,T)$. Since the set $\mathscr{V}$ has finite cardinality, we can identify the set of probability measures on $\mathscr{V}$ with $\mathscr{P}(\mathscr{V})$. Thus, if μ is a relaxed control, there exists a time-dependent vector-valued function $\alpha(t) = [\alpha_1(t)...\alpha_v(t)]^T$ such that $\mu(t,\mathscr{U}) = \sum_{v\in\mathscr{U}} \mu(t,v) = \sum_{v\in\mathscr{U}} \alpha_v(t)$ for almost every $t \in (0,T)$ and all $\mathscr{U} \subset \mathscr{V}$. Then the solution of the system (2.39) coincides with the solution of the system

$$\begin{aligned} \dot{\mathbf{z}}(t) &= \sum_{e\in\mathscr{E}} u_e^0(\mathbf{z}(t))\mathbf{B}_e\mathbf{z}(t) + \int_{\mathscr{V}} \mathbf{D}_v\mathbf{y}(t)\mu(t,dv) \quad t \in [0,\infty), \\ \mathbf{z}(0) &= \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \end{aligned} \tag{2.40}$$

This implies that we can identify $R(0,T;\mathscr{V})$ with $L^\infty(0,1;\mathscr{P}(\mathscr{V}))$. The duality map $<\cdot,\cdot>$ from $L^1(0,T;C(\mathscr{V}))\times R(0,T;\mathscr{V})$ to $\mathbb{R}$ will be defined by $<\mu,f>=\int_0^T\int_{\mathscr{V}} f(t)d\mu(t,dv)dt$ for all $f \in L^1(0,T;C(\mathscr{V}))$ and all $\mu \in R(0,T;\mathscr{V})$. A sequence in μ_n in $R(0,T;\mathscr{V})$ is said to *weakly converge* to $\mu \in R(0,T;\mathscr{V})$ if

$$\lim_{n\to\infty} <\mu_n, f> = <\mu, f> \tag{2.41}$$

for all $f \in L^1(0,T;C(\mathscr{V}))$. Let $PC(0,T;D)$ denote the elements of $R(0,T;\mathscr{V})$ that are piecewise constant, and for each $t \in [0,T]$ the measure is a Dirac mass, that is, for each $t \in [0,T]$ there exists a $v \in \mathscr{V}$ such that the measure of v is equal to 1. With these definitions, we can state and prove our next result.

Proposition 2.4.5. *Given $T > 0$, let $y(t)$ be the solution of the system* (2.39) *for a set of controls $\alpha_v : [0,T] \to [0,1]$ such that $\sum_{v\in\mathscr{V}} \alpha_v(t) = 1$ for all $t \in [0,T]$. Then, for each $\varepsilon > 0$ there exists a control $\ell : [0,T] \to \mathscr{V}$ such that the solution $\mathbf{x}(t)$ of the system* (2.37) *satisfies $\|\mathbf{x}(T) - \mathbf{y}(T)\|_2 \leq \varepsilon$.*

Proof. Let $\alpha \in L^\infty(0,1;\mathscr{P}(\mathscr{V}))$ and let $\mu \in R(0,T;\mathscr{V})$ be the corresponding relaxed control. Then from (Fattorini, 1999)[Theorem 12.6.7], there exists a sequence $(\mu_n)_{n=1}^\infty \in$

$PC(0,T;D)$ that weakly converges to μ. Let $(\ell_n)_{n=1}^{\infty}$ be the sequence of piecewise constant functions from $[0,T]$ to $\mathscr{V}$ constructed by setting, for each $t \in [0,\infty]$ and each $v \in \mathscr{V}$, $\ell_n(t) = v$ if $\mu_n(t,v) = 1$. From (Fattorini, 1999), we know that solutions $\mathbf{z}_n(t)$ of the system (2.40) with relaxed control μ_n converge to the solution $\mathbf{z}$ of the system (2.40) with relaxed control μ, uniformly over the time interval $[0,T]$. This concludes the proof. □

Lemma 2.4.4 states that trajectories of system (2.39) can be approximated arbitrarily well using trajectories of system (2.37). Combining Lemma 2.4.4 and Proposition 2.4.5, we obtain the following main theorem on *approximate controllability* of system (2.37), which gives an affirmative answer to a weaker form of Problem 2.4.1 for which the state at final time is only required to be within distance ε of the target final state.

Theorem 2.4.6. *Let $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$ be an equilibrium point of the system* (2.39) *with steady-state control input $\alpha^{ss} = [\alpha_1^{ss} \ \ldots \ \alpha_M^{ss}]^T \in \text{int } \mathscr{P}(\mathscr{V})$. Additionally, let $T > 0$. Then there exists a neigborhood U of $\mathscr{P}(\mathscr{V})$, such that for each $\mathbf{x}^0 \in U$ and each $\varepsilon > 0$, there exists a control $\ell : [0,T] \to \mathscr{V}$ such that the solution $\mathbf{x}(t)$ of the system* (2.37) *satisfies $\|\mathbf{x}(T) - \mathbf{x}^{eq}\|_2 \leq \varepsilon$.*

Remark 2.4.7. ***(Lack of Global Controllability)*** *While Theorem 2.4.6 states that system* (2.37) *is locally approximately controllable about an equilibrium point $\mathbf{x}^{eq}$, in general, we cannot expect global controllability of the system for any $T > 0$. For example, take the two node graph $\mathscr{G}$, with $\mathscr{V} = \{1,2\}$. Then, for a given positive parameter c, we set $u_{(1,2)} = cy_2^2$ and $u_{(2,1)} = cy_1^2$, for all $y_1, y_2 \in [0,1]$. If $x_1^0 < 0.5$ and $c > 0$ is large enough, then $\lim_{t \to 0^+} \dot{x}_1(t) < 0$ for any choice of piecewise constant $\ell(t)$. This implies that system* (2.37) *is not controllable to the equilibrium point $\mathbf{x}^{eq} = [0.5 \ 0.5]^T$ from $\mathbf{x}^0$ for any final time $T > 0$.*

Remark 2.4.8. ***(Unbounded Speed of the Leader)*** *It is important to note that in order to prove controllability of the system* (2.37)*, we have implicitly assumed that the leader can*

switch between states arbitrarily fast. This implies that the leader can move at arbitrarily large speeds in space, which might not be a realistic assumption in practice. In practice, the leader's speed will have an upper bound, which implies a lower bound on the switching times. This would, in turn, impose a lower bound on the parameter ε in Theorem 2.4.6 so that the approximate controllability result remains true. However, it is difficult to analytically quantify such a lower bound on ε as a function of a lower bound on the switching times.

2.4.2 Stabilization

From here on, we will assume that the followers are not interacting with one another; that is, $u_e^o(\mathbf{x}) = 0$ for all $\mathbf{x} \in \mathscr{P}(\mathscr{V})$ and all $e \in \mathscr{E}$.

To address Problem 2.4.2, we will construct two control laws that govern the leader's state transitions. Toward this end, we introduce some new definitions. A *complete walk*, denoted by $\mathscr{W} = (e_i)_{i=1}^{w}$, is a sequence of size $w \in \mathbb{Z}_{>0}$ in $\mathscr{E}$ such that $S(e_1) = T(e_w)$, $T(e_i) = S(e_{i+1})$ for each $i \in \{1, ..., w-1\}$, and for each $v \in \mathscr{V}$ there exists $j \in \{1, ..., w\}$ such that $T(e_j) = v$. We will extend a given complete walk $\mathscr{W}$ to an *extended complete walk (ECW)*, $\mathscr{W}^\infty = (e_i)_{i=1}^{\infty}$, by defining

$$e_{nw+j} = e_j \text{ for } n \in \mathbb{Z}_{>0},\ j \in \{1, ..., w\}. \tag{2.42}$$

The sequence $\mathscr{W}^\infty$ denotes the path along which the leader can transition from one state to another.

Open-Loop Controller

We first construct an open-loop control strategy for the leader agent. An advantage of this control law over the feedback control law presented in the next subsection is that the leader is not required to measure the density of follower agents.

Let $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$, $\varepsilon > 0$, and $t_0^\varepsilon = 0$, and define $R_v = \{k \in \{1,...,w\};\ S(e_k) = v\}$. We define switching times $(t_j^\varepsilon)_{j=1}^\infty$ as

$$t_j^\varepsilon = t_{j-1}^\varepsilon + \frac{\varepsilon}{|R_{S(e_j)}| x_{S(e_j)}^{eq}} \quad \text{for } j \in \mathbb{Z}_{>0}, \tag{2.43}$$

where $|R_v|$ denotes the cardinality of the set R_v for each $v \in \mathscr{V}$. We also define $\ell^\varepsilon : [0,\infty) \to \mathscr{V}$ as

$$\ell^\varepsilon(t) = S(e_j) \quad \text{for } t \in [t_{j-1}^\varepsilon, t_j^\varepsilon),\ j \in \mathbb{Z}_{>0}. \tag{2.44}$$

Let $P = \sum_{k=1}^w t_k^1$ and

$$\mathbf{A}_{\text{av}} = \frac{1}{P}\int_0^P D_{\ell^1(t)} dt. \tag{2.45}$$

Then, setting $\tilde{\mathbf{A}} = \frac{1}{P}\sum_{v\in\mathscr{V}} \mathbf{D}_v$ and $\mathbf{D} = \text{diag}\,[x_1^{eq}\ ...\ x_M^{eq}]^T$, we have that $\mathbf{A}_{\text{av}} = \tilde{\mathbf{A}}\mathbf{D}^{-1}$.

Lemma 2.4.9. *Let $\ell(t) = \ell^\varepsilon(t)$ in* (2.37)*. There exists $\varepsilon_0 > 0$ and a time-varying matrix $\mathbf{A} : [0,\infty) \to \mathbb{R}^{M\times M}$ such that if $\varepsilon \in (0,\varepsilon_0]$, then the solution $\mathbf{x}(t)$ of* (2.37) *can be approximated using the solution $\mathbf{y}(t)$ of the equation*

$$\dot{\mathbf{y}}(t) = \mathbf{A}_{\text{av}}\mathbf{y}(t) + \varepsilon\mathbf{A}(\frac{t}{\varepsilon})\mathbf{y}(t), \quad \mathbf{y}(0) = \mathbf{x}^0. \tag{2.46}$$

In particular, $\|\mathbf{x}(t) - \mathbf{y}(t)\| = O(\varepsilon)$. Moreover, the map $t \mapsto \mathbf{A}(t)$ is such that the induced 2-norm $\|\mathbf{A}(t)\|$ is globally bounded over $t \in \mathbb{R}_{\geq 0}$ and $\mathbf{A}(t+P) = \mathbf{A}(t)$ for all $t \in [0,\infty)$.

Proof. Consider the change of variables $\tau = \frac{t}{\varepsilon}$. Then (2.37) becomes

$$\dot{\mathbf{x}}(\tau) = \varepsilon\mathbf{D}_{\ell(t)}\mathbf{x}(\tau) \tag{2.47}$$

Let $\mathbf{H}(\tau) = \mathbf{D}_{\ell(\tau)} - \mathbf{A}_{\text{av}}$ for each $\tau \in [0,\infty)$. Set $\mathbf{U}(\tau) = \int_0^\tau \mathbf{H}(s)ds$. Consider the change of variables

$$\mathbf{x}(\tau) = \mathbf{y}(\tau) + \varepsilon\mathbf{U}(\tau)\mathbf{y}(\tau) \tag{2.48}$$

Then we see that

$$\dot{\mathbf{x}}(\tau) = \dot{\mathbf{y}}(\tau) + \varepsilon\mathbf{U}(\tau)\dot{\mathbf{y}}(\tau) + \varepsilon\dot{\mathbf{U}}(\tau)\mathbf{y}(\tau) \tag{2.49}$$

For all ε small enough, $\mathbf{I}+\varepsilon\mathbf{U}(\tau)$ is invertible for all $\tau \in [0,\infty)$ and can be represented by the power series expression

$$(\mathbf{I}+\varepsilon\mathbf{U}(\tau))^{-1} = \sum_{i=0}^{\infty}(-\varepsilon)^i\mathbf{U}^i(\tau) \tag{2.50}$$

From (2.47), (2.50), and the fact that $\dot{\mathbf{U}}(\tau) = \mathbf{H}(\tau)$ and $\mathbf{D}_{\ell(\tau)} - \mathbf{H}(\tau) = \mathbf{A}_{\text{av}}$ for all $\tau \in [0,\infty)$, equation (2.49) can be used to solve for $\dot{\mathbf{y}}(\tau)$:

$$\dot{\mathbf{y}}(\tau) = \varepsilon\mathbf{A}_{\text{av}}\mathbf{y}(t) + \varepsilon^2\mathbf{A}(\tau)\mathbf{y}(\tau), \tag{2.51}$$

where the time-varying matrix $\mathbf{A}(\tau)$ is globally norm-bounded in time. From equation (2.48), and noting again that $\mathbf{I}+\varepsilon\mathbf{U}(\tau)$ is invertible for all $\tau \in [0,\infty)$ for small enough ε, we conclude that $\|\mathbf{x}(t)-\mathbf{y}(t)\| = O(\varepsilon)$ for all $t \geq 0$. The periodicity of $\mathbf{A}(\tau)$ follows from the fact that both $\mathbf{H}(\tau)$ and $\mathbf{U}(\tau)$ are periodic. □

Using Lemma 2.4.9, we can now establish the stability properties of system (2.37) with the control input $\ell(t) = \ell^{\varepsilon}(t)$ defined in (2.44). The following theorem uses the fact that solutions of system (2.46) can be used to approximate solutions of (2.37). The theorem applies an argument based on averaging theory (Sanders *et al.*, 2007) to prove *practical stability* of system (2.37).

Theorem 2.4.10. *Suppose the graph $\mathcal{G}$ is bidirected, $\mathscr{W}^{\infty} = (e_i)_{i=1}^{\infty}$ is an ECW, and $\mathbf{x}^{eq} \in$ int $\mathscr{P}(\mathscr{V})$. Let $\ell(t) = \ell^{\varepsilon}(t)$. There exists $\varepsilon_0 >$ such that for each $\varepsilon \in (0,\varepsilon_0]$, there exists $C_\varepsilon \geq 0$ with $\lim_{\varepsilon\to 0} C_\varepsilon = 0$ and $T_\varepsilon^{0,eq} > 0$, which depends on $\mathbf{x}^0$, $\mathbf{x}^{eq}$, and ε, such that $\|\mathbf{x}(t)-\mathbf{x}^{eq}\| < C_\varepsilon$ for all $t \geq T_\varepsilon^{0,eq}$.*

Proof. Let $\mathbf{A} : [0,\infty) \to \mathbb{R}^{M\times M}$ be the time-varying matrix from Lemma 2.4.9. Then consider the linear equation (2.46). Define a Lyapunov function $V : \mathscr{P}(\mathscr{V}) \to \mathbb{R}_{\geq 0}$ given by

$$V(\mathbf{z}) = (\mathbf{z}-\mathbf{x}^{eq})^T\mathbf{D}(\mathbf{z}-\mathbf{x}^{eq}) \tag{2.52}$$

for all $\mathbf{z} \in \mathscr{P}(\mathscr{V})$. Since the graph $\mathscr{G}$ is bidirected and strongly connected, we compute that $\frac{\partial V}{\partial \mathbf{y}}^T \mathbf{A}_{\text{av}}(\mathbf{y}(t) - \mathbf{x}^{eq}) = -\sum_{e \in \mathscr{E}} \frac{1}{2}(y_{S(e)}(t) - x^{eq}_{S(e)} - y_{T(e)}(t) + x^{eq}_{T(e)})^2 < 0$ for all $t \geq 0$ such that $\mathbf{y}(t) \in \mathscr{P}(\mathscr{V}) \setminus \{\mathbf{x}^{eq}\}$. Then we have that

$$\dot{V}(\mathbf{y}(t)) = \frac{\partial V(\mathbf{y}(t))}{\partial \mathbf{y}}^T \mathbf{A}_{\text{av}}(\mathbf{y}(t) - \mathbf{x}^{eq}) + \varepsilon \frac{\partial V(\mathbf{y}(t))}{\partial \mathbf{z}}^T \mathbf{A}(t)\mathbf{y}(t) \tag{2.53}$$

since $\mathbf{A}_{av}\mathbf{x}^{eq} = 0$.

It follows from the computations in the proof of Lemma 2.4.9 that all off-diagonal elements of $\mathbf{A}_{av} + \varepsilon \mathbf{A}(\frac{t}{\varepsilon})$ are non-negative and that $\mathbf{1}^T(\mathbf{A}_{av} + \varepsilon \mathbf{A}(\frac{t}{\varepsilon})) = \mathbf{0}$ for all $t \in [0, \infty)$ and for all $\varepsilon > 0$ small enough. Then, from Lemma 2.4.9, we can conclude that $\mathscr{P}(\mathscr{V})$ is invariant for the solution $\mathbf{y}(t)$. This result implies that the term $\frac{\partial V(\mathbf{y}(t))}{\partial \mathbf{z}}^T \mathbf{A}(t)\mathbf{y}(t)$ is uniformly bounded. Thus, the second term in the right-hand side of equation (2.53) is bounded by a parameter C'_ε for each $\varepsilon > 0$ with $\lim_{\varepsilon \to 0} C'_\varepsilon = 0$. This implies that for all $\varepsilon > 0$ small enough, $\dot{V}(\mathbf{y}(t)) < 0$ for all $t \in [0, \infty)$ such that $\|\mathbf{y}(t) - \mathbf{x}^{eq}\| > C_\varepsilon$, where $C_\varepsilon \to 0$ as $\varepsilon \to 0$. □

Remark 2.4.11. *The assumption that the graph $\mathscr{G}$ is bidirected has been made for the sake of simplicity. Theorem 2.4.10 can be generalized to strongly connected graphs that are not necessarily bidirected by replacing* $\frac{\varepsilon}{|R_{S(e_j)}| x^{eq}_{S(e_j)}}$ *with* $\frac{\varepsilon x^d_{S(e_j)}}{|R_{S(e_j)}| x^{eq}_{S(e_j)}}$ *in* (2.43) *for each* $j \in \mathbb{Z}_{>0}$*, where, from the Perron-Frobenius theorem (Berman and Plemmons, 1994),* $\mathbf{x}^d$ *is the unique vector in* $\mathscr{P}(\mathscr{V})$ *such that* $\sum_{e \in \mathscr{E}} \mathbf{B}_e \mathbf{x}^d = \mathbf{0}$.

Closed-Loop Controller

In contrast to the open-loop control law presented in the previous section, the control law that we present in this section is a function of the density of the followers at the leader's current state. We show through numerical simulations in Section 2.4.2 that this closed-loop controller ensures faster convergence of the followers to the target distribution than the open-loop controller.

Given $\mathbf{x}^{eq} \in \mathbb{R}^M$, we define the set $\mathscr{Q} \subset \mathbb{R}^M \times \mathbb{Z}_{>0}$ as:

$$\mathscr{Q} = \{(\mathbf{x}, k) \in \mathbb{R}^M \times \mathbb{Z}_{>0};\ x_{S(e_k)} \leq x^{eq}_{S(e_k)}\}. \tag{2.54}$$

The set $\mathscr{Q}$ is used as follows to define the feedback control law according to which the leader transitions from one state to another. If the leader is in state $S(e_k)$ and the density of follower agents in that state, $x_{S(e_k)}$, is less than or equal to the target value $x^{eq}_{S(e_k)}$, then the leader transitions to the next state $T(e_k)$ in $\mathscr{W}^\infty$. While the path that the leader takes is predetermined by the specification of $\mathscr{W}^\infty$, the times at which it switches from one state to another is a function of the follower density that it measures at its current state, according to the following equations:

$$\begin{aligned} k(t^+) &= k(t^-) + 1, \\ \ell(t^+) &= T(e_{k(t^-)}),\ \ (\mathbf{x}(t^-), k(t^-)) \in \mathscr{Q}, \end{aligned} \tag{2.55}$$

where $k(t^+)$ and $\ell(t^+)$ denote the right-sided limits of the functions $k(t)$ and $\ell(t)$, respectively, at time t, and $k(t^-)$ and $\ell(t^-)$ denote the left-sided limits of $k(t)$ and $\ell(t)$ at t. This control law for the leader, combined with the ODE model (2.37) that governs the follower agent densities, results in a *hybrid dynamical system* (Goebel *et al.*, 2012) in which the continuous-time dynamics are given by:

$$\begin{aligned} \dot{\mathbf{x}}(t) &= \mathbf{D}_{\ell(t)}\mathbf{x}(t), \\ \dot{k}(t) &= 0, \\ \dot{\ell}(t) &= 0, \quad t \in [0, \infty), \end{aligned} \tag{2.56}$$

the discrete-time dynamics are given by equations (2.55), and the initial conditions are defined as:

$$\mathbf{x}(0) = \mathbf{x}^0 \in \mathscr{P}(\mathscr{V}), \quad k(0) = 0, \quad \ell(0) = S(e_1). \tag{2.57}$$

Since the closed-loop system (2.56)-(2.57) is a hybrid system, we need an appropriate notion of a solution to this type of system in order to establish our stability result in

Theorem 2.4.16. Hence, we provide the following definition that will be sufficient for the purposes of this section.

Definition 2.4.12. *Suppose that $\mathscr{W}^\infty = (e_i)_{i=1}^\infty$ is a given ECW. By a solution of the system* (2.56)-(2.57)*, we mean that there exists a time $t_f \geq 0$ (possibly equal to ∞), an absolutely continuous function $\mathbf{x} : [0,t_f) \to \mathscr{P}(\mathscr{V})$, piecewise constant functions $k : [0,t_f) \to \mathbb{Z}_{>0}$ and $v_l : [0,t_f) \to \mathscr{V}$, and a sequence of non-decreasing **switching times** $(t_i)_{i=1}^\infty$ such that $\lim_{j\to\infty} t_j = t_f$ and, for each $i \in \mathbb{Z}_{>0}$, we have that*

$$\mathbf{x}(t) = \mathbf{x}_0 + \int_0^{\min\{t,t_i\}} \mathbf{D}_{\ell(t)}\mathbf{x}(\tau)d\tau \tag{2.58}$$

and

$$\begin{aligned} k(t) &= i, \quad t \in [t_{i-1},t_i),\ (\mathbf{x}(t_i),k(t_i)) \in \mathscr{Q}, \\ \ell(t) &= T(e_{i-1}),\ t \in [t_{i-1},t_i), \end{aligned} \tag{2.59}$$

where $t_0 = 0$ and $[t_{i-1},t_i) := \emptyset$ is the null set if $t_{i-1} = t_i$.

Given this definition, we prove the following result on the existence and uniqueness of solutions of the system (2.56)-(2.57). In the following theorem and henceforth, int $\mathscr{P}(\mathscr{V})$ will denote the interior of the set $\mathscr{P}(\mathscr{V})$ in the subspace topology of $\mathscr{P}(\mathscr{V})$ as a subset of $\mathbb{R}^M$.

Theorem 2.4.13. *Suppose that $\mathscr{W}^\infty = (e_i)_{i=1}^\infty$ is an ECW and $\mathbf{x}^{eq} \in$ int $\mathscr{P}(\mathscr{V})$. Then there exists a unique solution to the system* (2.56)-(2.57) *with switching times $(t_i)_{i=1}^\infty$.*

Proof. First, we show that there at least exists a unique *local solution* of system (2.56)-(2.57). Specifically, we show that there exists $j \in \mathbb{Z}_{>0}$ and a sequence of switching times $(t_i)_{i=1}^j$ such that $\mathbf{x} : [0,t_j] \to \mathscr{P}(\mathscr{V})$ is absolutely continuous and equations (2.58)-(2.59) hold for each $i \in \{1,...,j\}$. Let $i_1 = \min\{m \in \mathbb{Z}_{>0};\ x_{S(e_m)}(0) > x^{eq}_{S(e_m)}\}$. If i_1 does not exist, we set $t_i = 0$ for all $i \in \mathbb{Z}_{>0}$, and the existence of a unique local solution is trivial.

Alternatively, suppose that i_1 is finite. Let $t_i = 0$ for all $i \in \{\tilde{i} \in \mathbb{Z}_{>0};\ 0 < \tilde{i} < i_1\}$. Set $v = S(e_{i_1})$ and $t_{i_1} = \frac{1}{|\mathscr{N}(S(v))|}\ln\frac{x_v(0)}{x_v^{eq}}$. Since $\mathbf{x}^{eq}$ lies in the interior of $\mathscr{P}(\mathscr{V})$, the quantity $\ln\frac{x_v(0)}{x_v^{eq}}$ is well-defined. It follows that $\dot{x}_v(s) = -|\mathscr{N}(v)|x_v(s)$ for all $s \in [0, t_{i_1})$. This implies that $x_v(s) = e^{-|\mathscr{N}(v)|s}x_v(0)$ for all $s \in [0, t_{i_1})$, and hence that $\lim_{s\to t_{i_1}} x_v(s) = x_v^{eq}$. Then we set $k(t) = i_1$ and $\ell(t) = v$ for all $t \in [0, t_{i_1})$. Thus, we have established that at least one local solution of system (2.56)-(2.57) exists. This constructed local solution can be non-unique only if there is an alternative possible choice of switching times, $(\tilde{t}_i)_{i=1}^{\tilde{j}}$. This alternative set of switching times is valid only if the first non-zero switching time is chosen to have an index greater than i_1. However, this would violate the requirement in constraint (2.59) that $(\mathbf{x}(t_i), k(t_i)) \in \mathscr{Q}$. Hence, the constructed local solution is unique.

Next, we will show that any local solution can be extended to a unique *global solution* that is defined over a countably infinite sequence of switching times. Suppose there exists a unique local solution of system (2.56)-(2.57). That is, there exists $p \in \mathbb{Z}_{>0}$, possibly larger than i_1, and a sequence of switching times such that $\mathbf{x} : [0, t_p] \to \mathscr{P}(\mathscr{V})$ is absolutely continuous and equations (2.58)-(2.59) hold for each $i \in \{1, ..., p\}$. Let $q_1 = \min\{m \in \mathbb{Z}_{>0};\ m > p\ \&\ \lim_{t\to t_p} x_{S(e_m)}(t) > x_{S(e_m)}^{eq}\}$. If q_1 does not exist, then we set $t_i = t_p$ for all $i \in \mathbb{Z}_+$ such that $i \geq p$, and the existence of a unique global solution is trivial. Alternatively, suppose that q_1 is finite. Let $t_i = t_p$ for all $i \in \{\tilde{i} \in \mathbb{Z}_{>0};\ p < \tilde{i} < q_1\}$. Set $v = S(e_{q_1})$ and $t_{q_1} = t_p + \frac{1}{|\mathscr{N}(S(v))|}\ln\frac{x_v(t_p)}{x_v^{eq}}$. Then we can see that $\dot{x}_v(s) = -|\mathscr{N}(v)|x_v(s)$ for all $s \in [t_{q_1-1}, t_{q_1})$. This implies that $x_v(s) = e^{-|\mathscr{N}(v)|(s-t_{q_1-1})}x_v(t_{q_1-1})$ for all $s \in [t_{q_1-1}, t_{q_1})$, and hence that $\lim_{s\to t_{q_1}} x_v(s) = x_v^{eq}$. Then we set $k(t) = q_1$ and $\ell(t) = v$ for all $t \in [t_{q_1-1}, t_{q_1})$. Therefore, any local solution can be extended uniquely over a longer time interval. Since we have already constructed one such local solution, this implies that we can iteratively construct a unique global solution $\mathbf{x}$ to system (2.56)-(2.57) by extending each local solution to another local solution over successively longer time intervals. Because this solution is both continuous and piecewise continuously differentiable, and therefore Lipschitz, it is

absolutely continuous. This concludes the proof. □

In the next lemma, we derive an estimate of the solutions of system (2.56)-(2.57) that will be used to prove Theorem 2.4.16.

Lemma 2.4.14. *Suppose that* $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$. *Let there exist some* $j \in \mathbb{Z}_{>0}$ *and* $\varepsilon > 0$ *such that the solution of system* (2.56)-(2.57) *satisfies* $x_v(t_{j-1}) > x_v^{eq} + \varepsilon$ *for* $v = S(e_j)$. *Then* $x_w(t_j) > x_w(t_{j-1}) + \frac{\varepsilon}{|\mathscr{N}(v)|}$ *for all* $w \in \mathscr{N}(v)$.

Proof. By assumption, we have that $x_v(t_{j-1}) > x_v^{eq} + \varepsilon$ for $v = S(e_j)$. Then $t_j = t_{j-1} + \frac{1}{|\mathscr{N}(v)|} \ln \frac{x_v(t_{j-1})}{x_v^{eq}}$. This implies that $\dot{x}_v(t) = |\mathscr{N}(v)| x_v(t)$ and $\dot{x}_w(t) = -x_v(t)$ for all $t \in [t_{j-1}, t_j)$ and all $w \in \mathscr{N}(v)$. Therefore, $x_w(t_j) = x_w(t_{j-1}) + \frac{x_v(t_{j-1}) - x_v^{eq}}{|\mathscr{N}(v)|}$ for all $w \in \mathscr{N}(v)$. □

The following proposition establishes an important monotonicity property of solutions of system (2.56)-(2.57) that is used in the proof of Theorem 2.4.16.

Proposition 2.4.15. *Suppose there exist times* $\tau_1 > 0$, $\tau_2 > \tau_1$ *and state* $v \in \mathscr{V}$ *such that the solution* $\mathbf{x}(t)$ *of system* (2.56)-(2.57) *satisfies* $x_v(t) \leq x_v^{eq}$ *for all* $t \in [\tau_1, \tau_2]$. *Then* $x_v(t)$ *is non-decreasing over the interval* $t \in [\tau_1, \tau_2]$. *Hence, if there exist* $\tilde{\tau} \geq 0$ *and* $w \in \mathscr{V}$ *such that the solution* $\mathbf{x}(t)$ *satisfies* $x_w(\tilde{\tau}) \geq x_v^{eq}$, *then* $x_w(t) \geq x_v^{eq}$ *for all* $t \in [\tilde{\tau}, t_f)$.

Proof. We are given that $x_v(t) \leq x_v^{eq}$ for all $t \in [\tau_1, \tau_2]$. Hence, $t_k - t_{k-1} = 0$ for all $k \in \mathbb{Z}_+$ such that $v = S(e_k)$ and $t_k \in [\tau_1, \tau_2]$. Moreover, $\dot{x}_v(t) \geq 0$ for all $t \in [t_{k-1}, t_k)$, since $g_{e_k}(\ell(t)) = 0$ whenever $v \neq S(e_k)$. This implies that $\dot{x}_v(t) \geq 0$ for $t \in (\tau_1, \tau_2)$, and therefore $\int_{\tau_1}^{t} \dot{x}_v(s) ds$ is non-decreasing for $t \in [\tau_1, \tau_2]$. Since the solution $\mathbf{x}$ is absolutely continuous, we have that $x_v(t) - x_v(\tau_1) = \int_{\tau_1}^{t} \dot{x}_v(s) ds$ for all $t \in [\tau_1, \tau_2]$. This concludes the proof. □

The result proved in Proposition 2.4.15 can be used to demonstrate that the target probability distribution $\mathbf{x}^{eq}$ is stable in the sense of Lyapunov. In the following theorem, we establish that this distribution is also globally attractive.

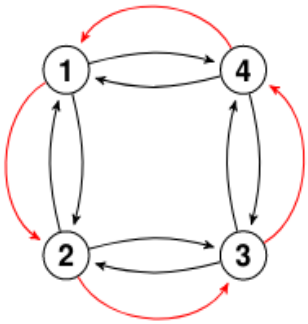
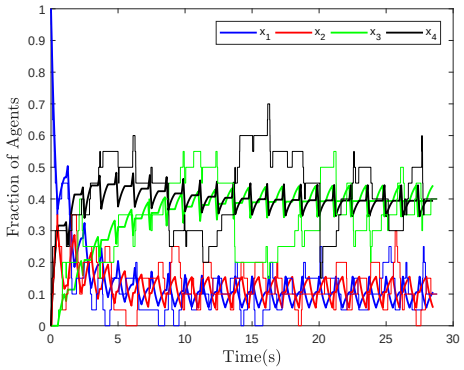


Figure 2.8: Bidirected graph with 4 vertices, representing agent states. Red edges define the leader's sequence of state transitions; black edges define followers' possible state transitions.

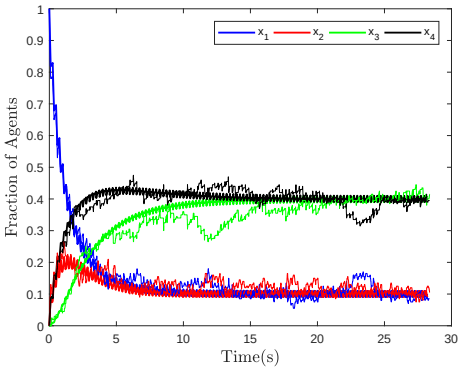
Theorem 2.4.16. *Suppose that* $\mathscr{W}^{\infty} = (e_i)_{i=1}^{\infty}$ *is an ECW and* $\mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$. *Then the unique solution of system* (2.56)-(2.57) *satisfies* $\lim_{t \to t_f} \mathbf{x}(t) = \mathbf{x}^{eq}$.

Proof. For the sake of contradiction, suppose that $\lim_{i \to \infty} \mathbf{x}(t_i) \neq \mathbf{x}^{eq}$. Then there must be a $v \in \mathscr{V}$ and $\varepsilon > 0$ such that for each $N \in \mathbb{Z}_{>0}$, there exists $p_N \geq N$ for which $x_v(t_{p_N - 1}) > x_v^{eq} + \varepsilon$. From Lemma 2.4.14, this implies that for every $w \in \mathscr{N}(v)$, $x_w(t_{p_N}) > x_w(t_{p_N - 1}) + \frac{\varepsilon}{|\mathscr{N}(v)|}$. Because $\varepsilon > 0$, it follows that there exists an $M \in \mathbb{Z}_{>0}$ that satisfies $\frac{M\varepsilon}{|\mathscr{N}(v)|} \geq x_v^{eq}$, and hence Proposition 2.4.15 implies that $x_w(t) \geq x_w^{eq} + \frac{\varepsilon}{|\mathscr{N}(v)|}$ for $t = t_{p_{M+1}}$ for all $w \in \mathscr{N}(v)$. Since the graph $\mathscr{G}$ is assumed to be strongly connected, Lemma 2.4.14 also implies that for each $r \in \mathscr{V}$, there exists $\tilde{\varepsilon}_r > 0$ and $q_N^r \in \mathbb{Z}_+$ corresponding to each $N \in \mathbb{Z}_{>0}$ such that $x_r(t) \geq x_r^{eq} + \tilde{\varepsilon}_r$ for $t = t_z$ with $z = q_N^r$. This leads to a contradiction with the monotonicity result in Proposition 2.4.15 and the fact that the set $\mathscr{P}(\mathscr{V})$ is invariant for the solution $\mathbf{x}(t)$. Hence, it must be true that $\lim_{i \to \infty} \mathbf{x}(t_i) = \mathbf{x}^{eq}$. From the monotonicity property of solutions proved in Proposition 2.4.15, it follows that $\lim_{t \to t_f} \mathbf{x}(t) = \mathbf{x}^{eq}$. □

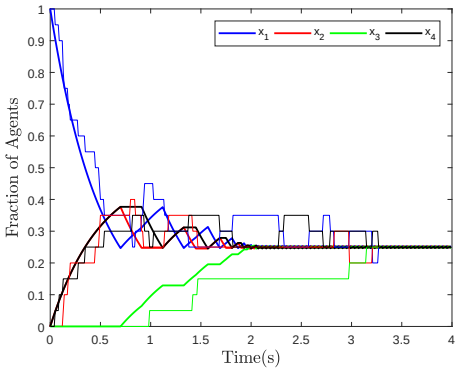
Remark 2.4.17. ***(Zeno Behavior)*** *Note that it is possible that* $\lim_{i \to \infty} t_i = t_f < \infty$. *In fact, this is trivially true when* $\mathbf{x}(0) = \mathbf{x}^{eq} \in \text{int } \mathscr{P}(\mathscr{V})$.



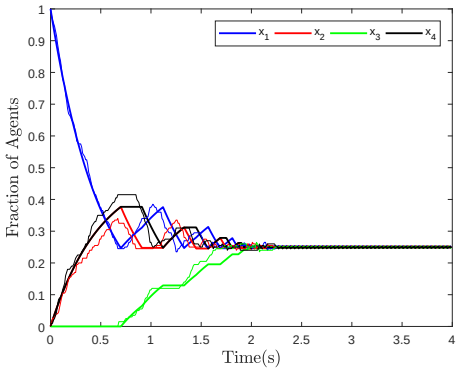
(a) Open-loop control with $N = 20$ follower agents ($\varepsilon = 0.05$)



(b) Open-loop control with $N = 200$ follower agents ($\varepsilon = 0.01$)



(c) Closed-loop control with $N = 20$ follower agents



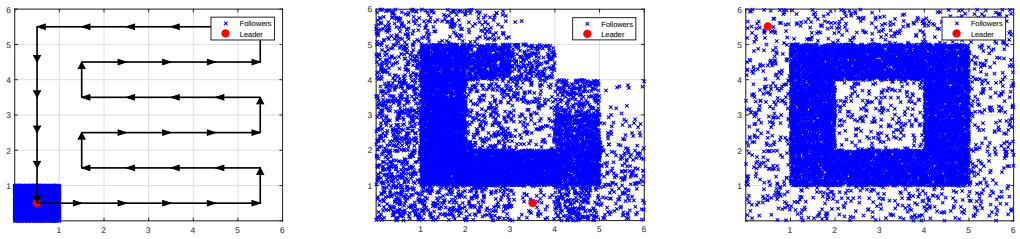
(d) Closed-loop control with $N = 200$ follower agents

Figure 2.9: Trajectories of the Mean-Field Model *(Thick Lines)* and the Corresponding Stochastic Simulations *(Thin Lines)*.

Simulations

In this subsection, we verify the effectiveness of our control strategies with numerical simulations of three scenarios with different controllers, graph topologies, and follower agent population sizes. In the first scenario, the leader agent must herd the follower agents to a target distribution over an undirected 4-vertex grid graph with the topology shown in Figure 2.8. The leader moves along the path $\mathscr{W}^{\infty} = ((1,2),(2,3),(3,4),(4,1),(1,2),...)$. The initial distribution of followers was set to $\mathbf{x}^0 = [1 \;\; 0 \;\; 0 \;\; 0]^T$, and the target distribution was defined as $\mathbf{x}^{eq} = [0.1 \;\; 0.1 \;\; 0.4 \;\; 0.4]^T$. Figures 2.9a and 2.9b compare the solution of the system (2.37) to the stochastic simulation of the CTMC characterized by expression (2.1) with the open-loop control (2.44) for two different follower population sizes, $N = 20$ and $N = 200$, with the corresponding switching parameter value set to $\varepsilon = 0.05$ and $\varepsilon = 0.01$, respectively. As expected, the plots show that the stochastic simulation for the $N = 200$ case follows the mean-field model solution more closely than for the $N = 20$ case. In both cases, the average follower populations converge to the target distribution within 27.5 s. For the $N = 20$ case, in which $\varepsilon = 0.05$, the solution of the mean-field model shows larger fluctuations about the target distribution than for the $N = 200$ case, in which $\varepsilon = 0.01$. This is consistent with the result in Theorem 2.4.10 that decreasing the value of ε produces smaller fluctuations of the solution of the mean-field model about the target distribution as $t \to \infty$.

In the second scenario, the graph topology and the path of the leader are the same as in the first scenario. The initial distribution of followers was set to $\mathbf{x}^0 = [1 \;\; 0 \;\; 0 \;\; 0]^T$, and the target distribution was defined as $\mathbf{x}^{eq} = [0.25 \;\; 0.25 \;\; 0.25 \;\; 0.25]^T$. Figures 2.9c and 2.9d compare the solution of system (2.37) to a stochastic simulation of the CTMC characterized by expression (2.1) with the feedback controller (2.55) for two different follower population sizes, $N = 20$ and $N = 200$. As expected, the plots show that the stochastic simulation for



(a) $t = 0$ s (b) $t = 100$ s (c) $t = 10^3$ s

Figure 2.10: Snapshots at three times t of $N = 10^4$ follower agents redistributing over a 36-vertex graph during a stochastic simulation of the closed-loop system. The black arrows define the sequence of state transitions by the leader agent.

the $N = 200$ case follows the mean-field model solution more closely than for the $N = 20$ case. In both cases, the average follower populations converge to the target distribution within 3.5 s. Compared to the open-loop controller used in the first scenario, we observe that the closed-loop controller achieves much faster convergence of the swarm to the target distribution.

To demonstrate the scalability of our control approach, we also considered a scenario in which the leader must herd $N = 10^4$ follower agents to a target distribution over a bidirected 36-vertex graph with a two-dimensional grid structure. All the follower agents start in a single state (the bottom left grid cell). One-tenth of the follower agents are required to distribute equally among the boundary cells and four cells at the center, while nine-tenths of the population is required to distribute equally among the remaining cells to form the letter 'O'. Figure 2.10 shows snapshots at times $t = 0$ s, $t = 100$ s, and $t = 10^3$ s of the distribution of follower agents and location of the leader agent in a stochastic simulation of the CTMC characterized by expression (2.1) with the feedback controller (2.55). Let $N_v(t)$ denote the number of follower agents in state $v \in \mathscr{V}$ at time t in the stochastic simulation, and define $\mathbf{x}^s(t) = \frac{1}{N}[N_1(t) \ \ldots \ N_{36}(t)]^T$ as the vector of followers' population fractions

in different states at time t. Measuring the difference between the simulated and target distributions of follower agents at time t as $E(t) = ||\mathbf{x}^s(t) - \mathbf{x}^{eq}||_2$, we compute $E(0) = 5.83$, $E(100) = 0.63$, and $E(10^3) = 3 \times 10^{-3}$ for the times of the snapshots in Figure 2.10. The decreasing value of $E(t)$ over time indicates that the follower agent distribution converges to the target distribution as desired, which can also be confirmed qualitatively from the snapshots.

Chapter 3

CONTROLLABILITY AND STABILIZATION OF PARTIAL DIFFERENTIAL EQUATION TYPE FORWARD EQUATIONS

In this chapter, we consider a controllability problem for a robotic swarm that is described by a mean-field model in the form of an *advection-diffusion* partial differential equation (PDE). Similar controllability problems have been addressed previously in the literature. For example, motivated by problems arising from quantum physics, Blaquiere (Blaquiere, 1992) used techniques from stochastic control to study a controllability problem in which a stochastic process evolves on a n-dimensional Euclidean space $\mathbb{R}^n$. A similar controllability result was proved in (Dai Pra, 1991). In (Porretta, 2014), Porretta addressed a controllability problem for a Fokker-Planck equation evolving on the n-dimensional torus, along with an associated *mean-field game* problem (Bensoussan *et al.*, 2013). This work applied observability inequalities that are typically used in PDE controllability problems. The results in (Blaquiere, 1992; Dai Pra, 1991) were extended to a more general setting in which the stochastic process is a linear control system perturbed by a diffusion process (Chen *et al.*, 2017). Controllability problems for systems with a similar structure have also been considered in work on *multiplicative control* of PDEs (Khapalov, 2010).

In contrast to these works, in this chapter, the stochastic process that models the agents' motion is confined to a bounded subset of a Euclidean space. Boundedness of the domain is a common constraint in many problems in swarm robotics, e.g. in (**Elamvazhuthi** *et al.*, 2018c; Milutinovic and Lima, 2007), where optimal control techniques were used to optimize swarm behavior. Additionally, the results in previous controllability studies were proven with control parameters that are square-integrable. However, in bilinear optimal control of PDEs associated with stochastic processes, the boundedness of the vector

fields is a common requirement (**Elamvazhuthi** *et al.*, 2018c; Finotti *et al.*, 2012; Fleig and Guglielmi, 2017). Toward this end, we establish controllability with control inputs that are (essentially) bounded in space and time.

Another contribution of this chapter is our analysis of a controllability problem for the forward equation of a class of hybrid switching diffusion processes (HSDPs) (Yin and Zhu, 2010). These processes can be used as models for robots that switch between multiple behavioral states, e.g. (Milutinovic and Lima, 2006; **Elamvazhuthi** *et al.*, 2018c). Our result is based on a controllability result for the forward equation of a related class of continuous-time Markov chains (CTMCs). A nontrivial issue in the problem of controlling the forward equation of CTMCs is the fact that the control parameters, which correspond to the transition rates of the Markov chain, are constrained to be positive. Hence, classical results on controllability of nonlinear control systems governed by ordinary differential equations (ODEs) do not apply. In spite of this issue, we prove controllability of these forward equations using piecewise constant control inputs. This controllability property can be attributed to the strong connectivity of the associated graphs.

We also consider the problem of stabilizing HSDPs to desired stationary distributions using time-independent and spatially-dependent controls or state feedback laws. A similar problem was considered in (Mesquita and Hespanha, 2012) for general controllable systems on unbounded domains with a single discrete state. As a final contribution, we consider the problem of using mean-field feedback laws to stabilize HSDPs that model a swarm of agents with nonholonomic dynamics to desired stationary distributions with disconnected supports.

3.1 Notation

We denote the n-dimensional Euclidean space by $\mathbb{R}^n$. $\mathbb{R}^{n\times m}$ refers to the space of $n \times m$ matrices, and $\mathbb{R}_+$ refers to the set of non-negative real numbers. Given a vector $\mathbf{x} \in \mathbb{R}^n$,

x_i denotes the i^{th} coordinate value of $\mathbf{x}$. For a matrix $\mathbf{A} \in \mathbb{R}^{n\times m}$, A^{ij} refers to the element in the i^{th} row and j^{th} column of $\mathbf{A}$. For a subset $B \subset \mathbb{R}^M$, int(B) refers to the interior of the set B. $\mathbb{C}$, $\mathbb{C}_-$, and $\bar{\mathbb{C}}_-$ denote the set of complex numbers, the set of complex numbers with negative real parts, and the set of complex numbers with non-positive real parts, respectively. $\mathbb{Z}_+$ refers to the set of positive integers. We denote by Ω an open, bounded, and connected subset of a Euclidean domain $\mathbb{R}^n$. The boundary of Ω is denoted by $\partial\Omega$.

Definition 3.1.1. *We will say that Ω is a $C^{1,1}$ **domain** if each point $\mathbf{x} \in \partial\Omega$ has a neighborhood $\mathscr{N}$ such that $\Omega \cap \mathscr{N}$ is represented by the inequality $x_n < \gamma(x_1,...,x_{n-1})$ in some Cartesian coordinate system for some function $\gamma : \mathbb{R}^{n-1} \to \mathbb{R}$ that is at least once differentiable and has derivatives of order 1 that are Lipschitz continuous.*

For each $1 \leq p < \infty$, we define $L^p(\Omega)$ as the Banach space of complex-valued measurable functions over the set Ω whose absolute value raised to p^{th} power has finite integral. We define $L^\infty(\Omega)$ as the space of essentially bounded measurable functions on Ω. The space $L^\infty(\Omega)$ is equipped with the norm $\|z\|_\infty = \text{ ess sup}_{\mathbf{x}\in\Omega}|z(\mathbf{x})|$, where ess $\sup_{\mathbf{x}\in\Omega}(\cdot)$ denotes the *essential supremum* attained by its argument over the interval Ω. The space $L^2(\Omega)$ is a Hilbert space when equipped with the standard inner product, $\langle\cdot,\cdot\rangle_2 : L^2(\Omega)\times L^2(\Omega) \to \mathbb{C}$, given by $\langle f,g\rangle_2 = \int_\Omega f(\mathbf{x})\bar{g}(\mathbf{x})d\mathbf{x}$ for each $f,g \in L^2(\Omega)$, where $\bar{g}$ is the complex conjugate of the function g. The norm $\|\cdot\|_2$ on the space $L^2(\Omega)$ is defined as $\|f\|_2 = \langle f,f\rangle_2^{1/2}$ for each $f \in L^2(\Omega)$. For a function $f \in L^2(\Omega)$ and a given constant c, we write $f \geq c$ to imply that f is real-valued and $f(\mathbf{x}) \geq c$ for almost every (a.e.) $\mathbf{x} \in \Omega$.

Let f_{x_i} denote the first-order (weak) partial derivative of the function f with respect to the coordinate x_i. Similarly, $f_{x_ix_i}$ will denote the second-order partial derivative of the function f with respect to the coordinate x_i. We define the Sobolev space $H^1(\Omega) = \left\{f \in L^2(\Omega) : f_{x_i} \in L^2(\Omega) \text{ for } 1 \leq i \leq N\right\}$. We equip this space with the usual Sobolev norm

$\|\cdot\|_{H^1}$, given by $\|f\|_{H^1} = \left(\|f\|_2^2 + \sum_{i=1}^n \|f_{x_i}\|_2^2\right)^{1/2}$ for each $f \in H^1(\Omega)$. The weak gradient of a function $f \in H^1(\Omega)$ will be denoted by $\nabla f = [f_{x_1} \ ... \ f_{x_n}]^T$.

Definition 3.1.2. *We will call Ω an **extension domain** if there exists a linear bounded operator $E : H^1(\Omega) \to H^1(\mathbb{R}^n)$ such that $(Ef)(\mathbf{x}) = f(\mathbf{x})$ for a.e. $\mathbf{x} \in \Omega$.*

An example of an extension domain is a domain with Lipschitz boundary (Agranovich, 2015)[Theorem 10.4.1]. Unless otherwise stated, the **default assumption** in this section will be that Ω **is an extension domain**. The exponential stability results will only require this default assumption. However, to prove the controllability result, we will need the stronger assumption that the domain Ω is $C^{1,1}$ or convex. An additional assumption about the domain Ω will be needed to prove the controllability result, which motivates the following definition.

Definition 3.1.3. *The domain Ω will be said to satisfy the **chain condition** if there exists a constant $C > 0$ such that for every $\mathbf{x}, \bar{\mathbf{x}} \in \Omega$ and every positive $n \in \mathbb{Z}_+$, there exists a sequence of points $\mathbf{x}_i \in \Omega$, $0 \leq i \leq n$, such that $\mathbf{x}_0 = \mathbf{x}$, $\mathbf{x}_n = \bar{\mathbf{x}}$, and $|\mathbf{x}_i - \mathbf{x}_{i+1}| \leq \frac{C}{n}|\mathbf{x} - \bar{\mathbf{x}}|$ for all $i = 0, ..., n-1$. Here $|\cdot|$ denotes the standard Euclidean norm.*

Note that every convex domain satisfies the chain condition. For a given real-valued function $a \in L^\infty(\Omega)$, $L_a^2(\Omega)$ refers to the set of all functions f such that $\int_0^1 |f(\mathbf{x})|^2 a(\mathbf{x}) d\mathbf{x} < \infty$. We will always assume that the associated function a is uniformly bounded from below by a positive constant, in which case the space $L_a^2(\Omega)$ is a Hilbert space with respect to the weighted inner product $\langle \cdot, \cdot \rangle_a : L_a^2(\Omega) \times L_a^2(\Omega) \to \mathbb{R}$, given by $\langle f, g \rangle_a = \int_\Omega f(\mathbf{x}) \bar{g}(\mathbf{x}) a(\mathbf{x}) d\mathbf{x}$ for each $f, g \in L_a^2(\Omega)$. We will also need the space $H_a^1(\Omega) = \left\{ z \in L_a^2(\Omega) : (az)_{x_i} \in L^2(\Omega) \text{ for } 1 \leq i \leq N \right\}$, equipped with the norm $\|f\|_{H_a^1} = \left(\|f\|_a^2 + \sum_{i=1}^n \|(af)_{x_i}\|_2^2\right)^{1/2}$. When $a = \mathbf{1}$, where $\mathbf{1}$ is the function that takes the value 1 a.e. on Ω, the spaces $L^1(\Omega)$ and $H^1(\Omega)$ coincide with the spaces $L_a^1(\Omega)$ and $H_a^1(\Omega)$, respectively. We will also need the spaces

$W^{1,\infty}(\Omega) = \left\{z \in L^\infty(\Omega) : z_{x_i} \in L^\infty(\Omega) \text{ for } 1 \leq i \leq N\right\}$ and $W^{2,\infty}(\Omega) = \left\{z \in W^{1,\infty}(\Omega) : z_{x_i x_i} \in L^\infty(\Omega) \text{ for } 1 \leq i \leq N\right\}$. Let X be a Hilbert space with the norm $\|\cdot\|_X$. The space $C([0,T];X)$ consists of all continuous functions $u : [0,T] \to X$ for which $\|u\|_{C([0,T];X)} := \max_{0\leq t\leq T} \|u(t)\|_X < \infty$. If Y is a Hilbert space, then $\mathscr{L}(X,Y)$ will denote the space of linear bounded operators from X to Y.

We will need an appropriate notion of a solution of the PDE (3.5). Toward this end, let A be a closed linear operator that is densely defined on a subset $\mathscr{D}(A)$, the domain of the operator, of a Hilbert space X. We will define spec(A) as the set $\{\lambda \in \mathbb{C} : (\lambda \mathbb{I} - A)^{-1}$ does not exist$\}$, where $\mathbb{I}$ is the identity map on X. If A is a bounded operator, then $\|A\|_{op}$ will denote the operator norm induced by the norm defined on X. From (Engel and Nagel, 2000), we have the following definition.

Definition 3.1.4. *For a given time $T > 0$, a **mild solution** of the ODE*

$$\dot{u}(t) = Au(t); \;\; u(0) = u_0 \in X \tag{3.1}$$

is a function $u \in C([0,T];X)$ such that $u(t) = u_0 + A\int_0^t u(s)ds$ for each $t \in [0,T]$.

Under appropriate conditions satisfied by A, the mild solution is given by a *strongly continuous semigroup* of linear operators, $(\mathscr{T}(t))_{t\geq 0}$, that are *generated* by the operator A (Engel and Nagel, 2000).

The differential equations that we analyze in this chapter will be non-autonomous in general. Hence, we must adapt the notion of a mild solution to these types of equations.

Definition 3.1.5. *Let A_i be a closed linear operator with domain $\mathscr{D}(A_i)$ for each $i \in \mathbb{Z}_+$. Suppose that for a certain time interval $[0,T]$, a piecewise constant family of operators is given by a map, $t \mapsto A(t)$, for which there exists a partition $[0,T] = \cup_{i\in\mathbb{Z}_+}[a_i, a_{i+1})$ such that $a_i \leq a_{i+1}$ for each $i \in \mathbb{Z}_+$ and $A(t) = A_i$ for each $t \in [a_i, a_{i+1})$. Then a mild solution of the ODE*

$$\dot{u}(t) = A(t)u(t); \;\; u(0) = u_0 \in X \tag{3.2}$$

is a function $u \in C([0,T];X)$ *such that*

$$u(t) = u_0 + \sum_{i \in \mathbb{Z}_+} A_i \int_{\min\{t,a_i\}}^{\min\{t,a_{i+1}\}} u(s)ds \tag{3.3}$$

for each $t \in [0,T]$.

There is in fact a more general notion of mild solutions that arises from two-parameter semigroups of operators generated by time-varying linear operators. However, the definition (3.3) will be sufficient for our purposes, since one can construct solutions of the ODE (3.2) by treating it as an autonomous system in each time interval $[a_i, a_{i+1})$ and concatenating these solutions together to obtain the solution u. Note that the mild solution is defined with respect to an operator A or collection of operators $A(t)$; when we refer to such a solution, the associated operator(s) will be clear from the context. We will also need the notion of a positive semigroup, which is defined as follows.

Definition 3.1.6. *A strongly continuous semigroup of linear operators* $(\mathscr{T}(t))_{t\geq 0}$ *on a Hilbert space* X *is called* ***positive*** *if* $u \in X$ *such that* $u \geq 0$ *implies that* $\mathscr{T}(t)u \geq 0$ *for all* $t \geq 0$.

We introduce some additional notation from graph theory which will be used in the coming sections. We denote by $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ a directed graph with a set of M vertices, $\mathscr{V} = \{1, 2, ..., M\}$, and a set of $N_{\mathscr{E}}$ edges, $\mathscr{E} \subset \mathscr{V} \times \mathscr{V}$. An edge from vertex $i \in \mathscr{V}$ to vertex $j \in \mathscr{V}$ is denoted by $e = (i, j) \in \mathscr{E}$. We define a source map $S : \mathscr{E} \to \mathscr{V}$ and a target map $T : \mathscr{E} \to \mathscr{V}$ for which $S(e) = i$ and $T(e) = j$ whenever $e = (i, j) \in \mathscr{E}$. There is a *directed path* of length s from vertex $i \in \mathscr{V}$ to vertex $j \in \mathscr{V}$ if there exists a sequence of edges $\{e_i\}_{i=1}^{s}$ in $\mathscr{E}$ such that $S(e_1) = i$, $T(e_s) = j$, and $S(e_k) = T(e_{k-1})$ for all $2 \leq k \leq s$. A directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ is called *strongly connected* if for every pair of distinct vertices $v_0, v_T \in \mathscr{V}$, there exists a *directed path* of edges in $\mathscr{E}$ connecting v_0 to v_T. We assume that $(i, i) \notin \mathscr{E}$ for all $i \in \mathscr{V}$. The graph $\mathscr{G}$ is said to be *bidirected* if $e \in \mathscr{E}$ implies that $\tilde{e} = (T(e), S(e))$ also lies in $\mathscr{E}$.

3.2 Controllability of an Advection-Diffusion Equation

Consider a swarm of N_p agents that are deployed on the n-dimensional domain Ω. The position of each agent, indexed by $i \in \{1,2,...,N_p\}$, evolves according to a stochastic process $\mathbf{Z}_i(t) \in \Omega$, where t denotes time. We assume that the agents are non-interacting. Therefore, the random variables that correspond to the dynamics of each agent are independent and identically distributed, and we can drop the subscript i and define the problem in terms of a single stochastic process $\mathbf{Z}(t) \in \Omega$. The deterministic motion of each agent is defined by a velocity vector field $\mathbf{v}(\mathbf{x},t) \in \mathbb{R}^n$, where $\mathbf{x} \in \Omega$. This motion is perturbed by a n-dimensional *Wiener process* $\mathbf{W}(t)$, which models noise. This process can be a model for stochasticity arising from inherent sensor and actuator noise. Alternatively, noise could be actively programmed into the agents' motion to implement more exploratory agent behaviors and to take advantage of the smoothening effect of the process on the agents' probability densities. Given the parameter $\mathbf{v}(\mathbf{x},t)$, each agent evolves according to a *reflected diffusion process* $\mathbf{Z}(t)$ that satisfies the following SDE (Pilipenko, 2014):

$$\begin{aligned} d\mathbf{Z}(t) &= \mathbf{v}(\mathbf{Z}(t),t)dt + \sqrt{2D}d\mathbf{W}(t) + \mathbf{n}(\mathbf{Z}(t))d\psi(t), \\ \mathbf{Z}(0) &= \mathbf{Z}_0, \end{aligned} \tag{3.4}$$

where $\psi(t) \in \mathbb{R}$ is called the *reflecting function* or *local time* (Bass and Hsu, 1991; Pilipenko, 2014), a stochastic process that constrains $\mathbf{Z}(t)$ to the domain Ω; $\mathbf{n}(\mathbf{x})$ is the normal to the boundary at $\mathbf{x} \in \partial\Omega$; and $D > 0$ is the diffusion constant. Without loss of generality, we assume that $D = 1$.

We now pose the problem of determining the existence of the robot control law, defined as the velocity field $\mathbf{v}(\cdot,t)$, that drives the swarm to a target spatial distribution over the domain.

Problem 3.2.1. *Given a time $t_f > 0$ and a target probability density $f : \Omega \to \mathbb{R}_+$ such that*

$\int_\Omega f(\mathbf{x})d\mathbf{x} = 1$, *determine if there exists a feedback control law* $\mathbf{v} : \Omega \times [0,t_f] \to \mathbb{R}^n$ *such that the process* (3.4) *satisfies* $\mathbb{P}(\mathbf{Z}(T) \in \Gamma) = \int_\Gamma f(\mathbf{x})d\mathbf{x}$ *for each measurable subset* $\Gamma \subset \Omega$.

The *Kolmogorov forward equation* corresponding to the SDE (3.4) is given by:

$$\begin{aligned} y_t &= \Delta y - \nabla \cdot (\mathbf{v}(\mathbf{x},t)y) \quad \textit{in } \ \Omega \times [0,T] \\ y(\cdot,0) &= y^0 \qquad\qquad\quad \textit{in } \ \Omega \\ \mathbf{n} \cdot (\nabla y - \mathbf{v}(\mathbf{x},t)y) &= 0 \quad \textit{in } \ \partial\Omega \times [0,T]. \end{aligned} \tag{3.5}$$

The solution $y(\mathbf{x},t)$ of this equation represents the probability density of a single agent occupying position $\mathbf{x} \in \Omega$ at time t, or alternatively, the density of a population of agents at this position and time. The PDE (3.5) is related to the SDE (3.4) through the relation $\mathbb{P}(\mathbf{Z}(t) \in \Gamma) = \int_\Gamma y(\mathbf{x},t)d\mathbf{x}$ for all $t \in [0,t_f]$ and all measurable $\Gamma \subset \Omega$. Therefore, the solution $y(\mathbf{x},t)$ captures the *mean-field* behavior of the population. In particular, as the number of agents tends to infinity, the *empirical measures* (Bensoussan *et al.*, 2013) converge to the measure for which this PDE's solution is the density $y(\mathbf{x},t)$. See (Zhang *et al.*, 2018) for such a convergence analysis. Problem 3.2.1 can be reframed in terms of equation (3.5) as a PDE controllability problem as follows:

Problem 3.2.2. *Given* $t_f > 0$, $y^0 : \Omega \to \bar{\mathbb{R}}_+$, *and* $f : \Omega \to \mathbb{R}_+$ *such that* $\int_\Omega y^0(\mathbf{x})d\mathbf{x} = \int_\Omega f(\mathbf{x})d\mathbf{x} = 1$, *determine whether there exists a space- and time-dependent parameter* $\mathbf{v} : \Omega \times [0,t_f] \to \mathbb{R}^n$ *such that the solution* y *of the PDE* (3.5) *satisfies* $y(\cdot,t_f) = f$.

Now, we prove one of the main theorems of this chapter. Specifically, we show that the PDE system (3.5) is controllable to a large class of sufficiently regular target probability densities. We first provide some new definitions that will be used in the subsequent analysis.

Given $a \in L^\infty(\Omega)$ such that $a \geq c$ for some positive constant c, and $\mathscr{D}(\omega_a) = H^1_a(\Omega)$,

we define the *sesquilinear form* $\omega_a : \mathscr{D}(\omega_a) \times \mathscr{D}(\omega_a) \to \mathbb{C}$ as

$$\omega_a(u,v) = \int_\Omega \nabla(a(\mathbf{x})u(\mathbf{x})) \cdot \nabla(a(\mathbf{x})\bar{v}(\mathbf{x}))d\mathbf{x} \tag{3.6}$$

for each $u \in \mathscr{D}(\omega_a)$. We associate with the form ω_a an operator $A_a : \mathscr{D}(A_a) \to L^2_a(\Omega)$, defined as $A_a u = v$ if $\omega_a(u,\phi) = \langle v,\phi\rangle_a$ for all $\phi \in \mathscr{D}(\omega_a)$ and for all $u \in \mathscr{D}(A_a) = \{g \in \mathscr{D}(\omega_a) : \exists f \in L^2_a(\Omega) \text{ s.t. } \omega_a(g,\phi) = \langle f,\phi\rangle_a \ \forall \phi \in \mathscr{D}(\omega_a)\}$.

Similarly, given $a \in L^\infty(\Omega)$ such that $a \geq c$ for some positive constant c and $\mathscr{D}(\sigma_a) = H^1_a(\Omega)$, we define the *sesquilinear form* $\sigma_a : \mathscr{D}(\sigma_a) \times \mathscr{D}(\sigma_a) \to \mathbb{C}$ as

$$\sigma_a(u,v) = \int_\Omega \frac{1}{a(\mathbf{x})}\nabla(a(\mathbf{x})u(\mathbf{x})) \cdot \nabla(a(\mathbf{x})\bar{v}(\mathbf{x}))d\mathbf{x} \tag{3.7}$$

for each $u \in \mathscr{D}(\sigma_a)$. As for the form ω_a, we associate an operator $B_a : \mathscr{D}(B_a) \to L^2_a(\Omega)$ with the form σ_a. We define this operator as $B_a u = v$ if $\sigma_a(u,\phi) = \langle v,\phi\rangle_a$ for all $\phi \in \mathscr{D}(\sigma_a)$ and for all $u \in \mathscr{D}(B_a) = \{g \in \mathscr{D}(\sigma_a) : \exists f \in L^2_a(\Omega) \text{ s.t. } \sigma_a(g,\phi) = \langle f,\phi\rangle_a \ \forall \phi \in \mathscr{D}(\sigma_a)\}$.

Note that, formally, $A_{\mathbf{1}} = B_{\mathbf{1}}$ is the Laplacian operator $-\Delta(\cdot)$ with *Neumann boundary condition* ($\mathbf{n} \cdot (\nabla \cdot\ ) = 0$ in $\partial\Omega$). For general extension domains Ω, the normal derivative might not make sense since it might not be true that $\mathscr{D}(A_{\mathbf{1}})$ is a subset of $H^2(\Omega)$ (Jerison and Kenig, 1989). Then, the Neumann boundary condition has to be interpreted in a *weak sense*.

Using the above definitions, we derive some preliminary results on the unbounded operators $-A_a$ and $-B_a$. The semigroups generated by these operators will each play an important role in the proof of controllability of system (3.5).

Lemma 3.2.3. *The operator $A_a : \mathscr{D}(A_a) \to L^2_a(\Omega)$ is closed, densely-defined, and self-adjoint. Moreover, the operator has a purely discrete spectrum.*

Proof. Consider the associated form ω_a. This form is *closed*, i.e., the space $\mathscr{D}(\omega_a)$ equipped with the norm $\|\cdot\|_{\omega_a}$, given by $\|u\|_{\omega_a} = (\|u\|^2_a + \omega_a(u,u))^{1/2}$ for each $u \in \mathscr{D}(\omega_a)$, is complete. This is true due to the fact that the multiplication map $u \mapsto a \cdot u$ is an isomorphism

from $H^1_a(\Omega)$ to $H^1(\Omega)$ and $H^1(\Omega)$ is a Banach space. Moreover, the space $H^1_a(\Omega)$ is dense in $L^2_a(\Omega)$. This follows from the inequality $\|au - av\|_2 \leq \|a\|_\infty \|u - v\|_2$ for each $u, v \in L^2(\Omega)$, the fact that the spaces $L^2_{\mathbf{1}}(\Omega)$ and $L^2_a(\Omega)$ are isomorphic, and the fact that the $H^1(\Omega)$ is dense in $L^2(\Omega)$. In addition, it follows from the definition of the form ω_a that ω_a is *symmetric*, meaning that $\omega_a(u,v) = \overline{\omega_a(v,u)}$ for each $u, v \in \mathscr{D}(\omega_a)$. The form ω_a is also *semibounded*, i.e., there exists $m \in \mathbb{R}$ such that $\omega_a(u,u) \geq m\|u\|^2_a$ for each $u \in \mathscr{D}(\omega_a)$. Hence, it follows from (Schmüdgen, 2012)[Theorem 10.7] that the operator A_a is self-adjoint. To establish the discreteness of the spectrum of A_a, we note that $H^1(\Omega)$ is compactly embedded in $L^2(\Omega)$ whenever Ω is an extension domain (Definition 3.1.2). This implies that when $H^1_a(\Omega) = \mathscr{D}(\omega_a)$ is equipped with the norm $\|\cdot\|_{\omega_a}$, then it is also compactly embedded in $L^2_a(\Omega)$. From (Schmüdgen, 2012)[Proposition 10.6], it follows that A_a has a purely discrete spectrum. □

Lemma 3.2.4. *The operator $B_a : \mathscr{D}(B_a) \to L^2_a(\Omega)$ is closed, densely-defined, and self-adjoint. Moreover, the operator has a purely discrete spectrum.*

Proof. We only check that the form σ_a is closed. The rest of the proof follows exactly the same arguments as the proof of Lemma 3.2.3. To prove that the form σ_a is closed, we need to prove that the space $\mathscr{D}(\sigma_a)$ equipped with the norm $\|\cdot\|_{\sigma_a}$, given by $\|u\|_{\sigma_a} = (\|u\|^2_a + \sigma_a(u,u))^{1/2}$ for each $u \in \mathscr{D}(\sigma_a)$, is complete. Note that due to the lower bound c on a, there exist constants $k_1, k_2 > 0$ such that

$$\begin{aligned} k_1 \textstyle\int_\Omega \nabla(a(\mathbf{x})u(\mathbf{x})) \cdot \nabla(a(\mathbf{x})\bar{u}(\mathbf{x}))d\mathbf{x} & \\ \leq \textstyle\int_\Omega \frac{1}{a(\mathbf{x})} \nabla(a(\mathbf{x})u(\mathbf{x})) \cdot \nabla(a(\mathbf{x})\bar{u}(\mathbf{x}))d\mathbf{x} & \\ \leq k_2 \textstyle\int_\Omega \nabla(a(\mathbf{x})u(\mathbf{x})) \cdot \nabla(a(\mathbf{x})\bar{u}(\mathbf{x}))d\mathbf{x} & \end{aligned} \tag{3.8}$$

for all $u \in H^1_a(\Omega)$. From these inequalities, it follows that $k_1\|u\|_{H^1_a} \leq \|u\|_{\sigma_a} \leq k_2\|u\|_{H^1_a}$ for all $u \in \mathscr{D}(\sigma_a) = H^1_a(\Omega)$. Hence, the form σ_a is closed. Due to the symmetry and

semiboundedness of this form, it follows from (Schmüdgen, 2012)[Theorem 10.7] that the operator is self-adjoint. Since the norm $\|\cdot\|_{\sigma_a}$ is equivalent to the norm $\|\cdot\|_{H^1_a}$, the discreteness of the spectrum of B_a again follows from (Schmüdgen, 2012)[Proposition 10.6] due to the compact embedding of $H^1_a(\Omega)$ in $L^2_a(\Omega)$. □

Corollary 3.2.5. *Consider the PDE*

$$\begin{aligned} y_t &= \Delta(a(\mathbf{x})y) \qquad \text{in } \Omega \times [0,T] \\ y(\cdot,0) &= y^0 \qquad \text{in } \Omega \\ \mathbf{n}\cdot\nabla(a(\mathbf{x})y) &= 0 \qquad \text{in } \partial\Omega \times [0,T]. \end{aligned} \tag{3.9}$$

Let $y^0 \in L^2_a(\Omega)$. Then $-A_a$ generates a semigroup of operators $(\mathscr{T}^A_a(t))_{t\geq 0}$ such that the unique mild solution $y \in C([0,T];L^2_a(\Omega))$ of the above PDE exists and is given by $y(\cdot,t) = \mathscr{T}^A_a(t)y^0$ for all $t \geq 0$. Additionally, the semigroup $(\mathscr{T}^A_a(t))_{t\geq 0}$ is positive. Finally, if $\|\mathscr{M}_a^{-1}y_0\|_\infty \leq 1$, then $\|\mathscr{M}_a^{-1}\mathscr{T}^A_a(t)y^0\|_\infty \leq 1$ for all $t \geq 0$.

Proof. First, we note that the operator $-A_a$ is *dissipative*, i.e., $\|(\lambda + A_a)u\|_a \geq \lambda\|u\|_a$ for all $\lambda > 0$ and all $u \in \mathscr{D}(A_a)$ since $\omega_a(u,u) \geq 0$ for all $u \in \mathscr{D}(\omega_a)$. Next, we note that $-A_a$ is self-adjoint, and hence the adjoint operator $-A^*_a$ is dissipative as well. It follows from a corollary of the *Lumer-Phillips theorem* (Engel and Nagel, 2000)[Corollary II.3.17] that $-A_a$ generates a semigroup of operators $(\mathscr{T}^A_a(t))_{t\geq 0}$ that solves the PDE (3.9) in the mild sense.

Second, we establish the positivity of the semigroup. Toward this end, we note that the absolute value function $|\cdot| : \mathbb{R} \to \mathbb{R}$ is Lipschitz. Hence, it follows from (Ziemer, 2012)[Theorem 2.1.11] that $v \in H^1(\Omega)$ implies that $|v| \in H^1(\Omega)$ whenever v is only real-valued. This implies that if $u \in \mathscr{D}(\omega_a)$, then $|\mathrm{Re}(u)| \in \mathscr{D}(\omega_a)$, where $\mathrm{Re}(\cdot)$ denotes the real component of its argument. Then the positivity of the semigroup follows from (Ouhabaz, 2009)[Theorem 2.7].

For the last statement, consider convex set $C = \{u \in L^2(\Omega); \mathrm{Re}(u) = u,\ u(\mathbf{x}) \leq 1/a(\mathbf{x})$ *a.e.* $\mathbf{x} \in \Omega\}$. This set is also closed in $L^2(\Omega)$. We will show that this set is invariant under the semigroup $\mathscr{T}_a^A(t)$. The projection of a function $u \in L_a^2(\Omega)$ on to the set C can be represented by the (nonlinear) operator P given by $Pu = \mathrm{Re}(u) \wedge 1/a = \frac{1}{2}(\mathrm{Re}(u + 1/a) + \frac{1}{2}|\mathrm{Re}(u) - 1/a|$. According to (Ouhabaz, 2009)[Theorem 2.3], the set C is invariant under the semigroup $\mathscr{T}_a^A(t)$, if for each $u \in \mathscr{D}(\omega_a)$, $Pu \in \mathscr{D}(\omega_a)$ and $\omega_a(Pu, Pu) \leq \omega(u,u)$. This is straightforward to verify. If $u \in \mathscr{D}(\omega_a)$, then it follows from follows from the chain rule (Ziemer, 2012)[Theorem 2.1.11] that $Pu = \frac{1}{2}(\mathrm{Re}(u) + 1/a) + \frac{1}{2}|\mathrm{Re}(u) - 1/a| \in \mathscr{D}(\omega_a)$ and $\nabla(aPu) = \frac{1}{2}\mathrm{sign}(\mathrm{Re}(au) - 1)\nabla(\mathrm{Re}(au)) + \frac{1}{2}\nabla(\mathrm{Re}(au)$. Hence, it follows that $\omega_a(Pu, Pu) \leq \omega_a(u,u)$ for all $u \in \mathscr{D}(\Omega_a)$. This implies that the set C is invariant under the semigroup $(\mathscr{T}_a^A(t))_{t\geq 0}$ and therefore, if $\mathscr{M}_a^{-1} y_0 \leq 1$, then $\mathscr{M}_a^{-1}\mathscr{T}_a^A(t)y^0 \leq 1$ for all $t \geq 0$ from (Ouhabaz, 2009)[Theorem 2.3]. Since the semigroup is also positive, we can conclude that, if $\|\mathscr{M}_a^{-1} y_0\|_\infty \leq 1$, then $\|\mathscr{M}_a^{-1}\mathscr{T}_a^A(t)y^0\|_\infty \leq 1$ for all $t \geq 0$. □

Using the same arguments as in the proof of Corollary 3.2.5, we have the following result.

Corollary 3.2.6. *The operator $-B_a$ generates a semigroup of operators $(\mathscr{T}_a^B(t))_{t\geq 0}$ on $L_a^2(\Omega)$. If additionally $a \in W^{1,\infty}(\Omega)$ and $y^0 \in L_a^2(\Omega)$, then $y(\cdot,t) = \mathscr{T}_a^B(t)y^0$ is a mild solution of the PDE*

$$\begin{aligned} y_t = \Delta y - \nabla \cdot (\tfrac{\nabla f(\mathbf{x})}{f(\mathbf{x})}\, y) \qquad & \text{in } \ \Omega \times [0,T] \\ y(\cdot,0) = y^0 \qquad & \text{in } \ \Omega \\ \mathbf{n} \cdot (\nabla y - \tfrac{\nabla f(\mathbf{x})}{f(\mathbf{x})} y) = 0 \qquad & \text{in } \ \partial\Omega \times [0,T], \end{aligned} \tag{3.10}$$

with $f = 1/a \in W^{1,\infty}(\Omega)$. Moreover, the semigroup $(\mathscr{T}_a^B(t))_{t\geq 0}$ is positive.

When $f \in C^\infty(\bar{\Omega})$, the representation of the operator $\Delta(\cdot) - \nabla \cdot (\frac{\nabla f(\mathbf{x})}{f(\mathbf{x})}\ \cdot\ )$ in the form $\nabla \cdot (f\nabla(\frac{1}{f}\ \cdot\ ))$ is a well-known technique in the literature on Fokker-Planck equations

for SDEs with drifts generated by potential functions (Stroock, 1993). In Corollary 3.2.6, however, since a is only once weakly differentiable and $\mathscr{D}(\sigma_a) = H_a^1(\Omega)$ (or equivalently, $H^1(\Omega)$), the operation Δy is not admissible unless a has additional regularity. Hence, the mild solution y should be interpreted as the weak solution of the PDE (3.10) when $a, f \in W^{1,\infty}(\Omega)$; i.e., it can be shown that y satisfies

$$\langle y_t, \phi \rangle_{V^*,V} = -\sigma_a(u, \phi/a) = -\int_\Omega \nabla y(\mathbf{x},t) \cdot \nabla \phi(\mathbf{x}) d\mathbf{x} + \int_\Omega \frac{\nabla f(\mathbf{x})}{f(\mathbf{x})} y(\mathbf{x},t) \cdot \nabla \phi(\mathbf{x}) d\mathbf{x} \quad (3.11)$$

for all $\phi \in V$ and a.e. $t \in [0,T]$, where $V = H^1(\Omega)$ and V^* is the dual space of V. Here, the second equality follows from the product rule (Theorem 3.2.9) and the fact that $a, f \in W^{1,\infty}(\Omega)$. Note that in the weak formulation (3.11), the second-order differentiability of f or y is not required. That the mild solution of a linear PDE is also a weak solution can be shown using standard energy estimates and weak topology arguments.

Next, we establish that the semigroups $(\mathscr{T}_a^A(t))_{t\geq 0}$ and $(\mathscr{T}_a^B(t))_{t\geq 0}$ are *analytic* (Lunardi, 2012). Additionally, we will show some mass-conserving properties and long-term stability properties of these semigroups.

Lemma 3.2.7. *The semigroups $(\mathscr{T}_a^A(t))_{t\geq 0}$ and $(\mathscr{T}_a^B(t))_{t\geq 0}$ that are generated by the operators $-A_a$ and $-B_a$, respectively, are analytic. Additionally, these semigroups have the following mass conservation property: if $y^0 \geq 0$ and $\int_\Omega y^0(\mathbf{x}) d\mathbf{x} = 1$, then $\int_\Omega (\mathscr{T}_a^A(t) y^0)(\mathbf{x}) d\mathbf{x} = \int_\Omega (\mathscr{T}_a^B(t) y^0)(\mathbf{x}) d\mathbf{x} = 1$ for all $t \geq 0$. Moreover, 0 is a simple eigenvalue of the operators $-A_a$ and $-B_a$ with the corresponding eigenvector $f = 1/a$. Hence, if $y^0 \geq 0$ and $\int_\Omega y^0(\mathbf{x}) d\mathbf{x} = \int_\Omega f(\mathbf{x}) d\mathbf{x} = 1$, then the following estimates hold:*

$$\|\mathscr{T}_a^A(t) y^0 - f\|_a \leq M_0 e^{-\lambda t} \|y^0 - f\|_a, \quad (3.12)$$

$$\|\mathscr{T}_a^B(t) y^0 - f\|_a \leq \tilde{M}_0 e^{-\tilde{\lambda} t} \|y^0 - f\|_a \quad (3.13)$$

for some positive constants $M_0, \tilde{M}_0, \lambda, \tilde{\lambda}$ and all $t \geq 0$.

Proof. The operators A_a and B_a are self-adjoint and positive semi-definite. Hence, their spectra lie in $[0,\infty)$. From this, it follows that the corresponding semigroups generated by $-A_a$ and $-B_a$ are analytic. Let $\int_\Omega y^0(\mathbf{x})d\mathbf{x} = 1$ such that $y^0 \in L^2_a(\Omega)$. Then $\int_\Omega (y(\mathbf{x},t) - y^0(\mathbf{x}))d\mathbf{x} = -\int_\Omega A_a(\int_0^t y(\mathbf{x},s)ds)d\mathbf{x} = -\omega_a(\int_0^t y(\mathbf{x},s)ds, 1/a) = 0$ for all $t \geq 0$. Hence, the integral preserving property of the semigroups holds. For the exponential stability estimates (3.12) and (3.13), we note that since the domain Ω is a connected bounded extension domain, it follows from Poincaré's inequality (Leoni, 2009)[Theorem 12.23] that there exists a constant $C > 0$ such that for all $u \in H^1(\Omega)$,

$$\int_\Omega |u(\mathbf{x}) - u_\Omega|^2 d\mathbf{x} \leq C \int_\Omega |\nabla u(\mathbf{x})|^2 d\mathbf{x}, \tag{3.14}$$

where $u_\Omega = \frac{1}{\mu(\Omega)} \int_\Omega u(\mathbf{x})d\mathbf{x}$. This implies that 0 is a simple eigenvalue of the Neumann Laplacian operator $A_{\mathbf{1}}$. Since the operator A_a can be written as a composition of operators $A_{\mathbf{1}}\mathscr{M}_a$, where $\mathscr{M}_a$ is the multiplication map $u \mapsto au$ from $H^1_a(\Omega)$ to $H^1(\Omega)$, it follows that 0 is also a simple eigenvalue of A_a with the corresponding eigenvector $f = 1/a$. Additionally, for a given $u \in H^1_a(\Omega)$, $\omega_a(u,u) = 0$ iff $\sigma_a(u,u) = 0$ due to the assumed positive lower bound on a. Hence, it also holds that 0 is a simple eigenvalue of the operator B_a. Then the result follows from (Engel and Nagel, 2000)[Corollary V.3.3]. □

The above result implies that if $\mathbf{v}(\cdot,t) = \nabla f/f$, then the solution of system (3.5) exponentially converges to f if f is in $W^{1,\infty}(\Omega)$ and is bounded from below by a positive constant. Hence, this choice of $\mathbf{v}(\cdot,t)$ is a possible control law for achieving exponential stabilization of desired probability densities. In the next few results, we verify some regularizing properties of the semigroups considered above, which will be critical to our controllability analysis.

Lemma 3.2.8. *Let $a \in L^\infty(\Omega)$ be real-valued and uniformly bounded from below by a positive constant c_1. Moreover, let $y^0 \in L^2_a(\Omega)$ such that $y^0 \geq c_2$ for some positive constant*

c_2. *If* $(\mathscr{T}_a^F(t))_{t\geq 0}$ *is the semigroup generated by the operator* $-A_a$ or $-B_a$, *then* $\mathscr{T}_a^F(t)y^0 \geq \frac{c_1c_2}{\|a\|_\infty}$ *for all* $t \geq 0$.

Proof. Let $k = c_1c_2$. Then we know that $a \cdot y^0 \geq k$. Hence, we can decompose the initial condition as $y^0 = kf + (y^0 - kf)$, where $f = 1/a$. Note that $y^0 - kf$ is positive and $\mathscr{T}_a^F(t)f = f$ for all $t \geq 0$. Therefore, it follows from the positivity preserving property of the semigroup (Corollary 3.2.6) that $\mathscr{T}_a^B(t)y^0 \geq k/\|a\|_\infty$ for all $t \geq 0$. □

We note the following well-known result on the product rule of differentiation for Sobolev functions, which will be used to prove Proposition 3.2.10 and other results later in this section.

Theorem 3.2.9. *(Product Rule) Let* $\Omega \subset \mathbb{R}^n$ *be an open bounded set. Suppose that* $u \in H^1(\Omega)$ *and* $v \in W^{1,\infty}(\Omega)$. *Then* $u \cdot v \in H^1(\Omega)$ *and the weak derivatives of the product* $u \cdot v$ *are given by*

$$(uv)_{x_i} = u_{x_i}v + v_{x_i}u \tag{3.15}$$

for each $i \in \{1, ..., n\}$.

Proposition 3.2.10. *Let* $a \in W^{1,\infty}(\Omega)$. *Then* $\mathscr{D}(A_a) = \mathscr{D}(B_a)$.

Proof. Let $u \in \mathscr{D}(B_a)$. Then using the product rule (Theorem 3.2.9), we have that

$$\omega_a(u, \frac{\phi}{a}) = \left\langle B_a u, \phi \right\rangle_a - \left\langle \frac{1}{a^2}\nabla a \cdot \nabla(au), \phi \right\rangle_a \tag{3.16}$$

for all $\phi \in H_a^1(\Omega)$. Since a is in $W^{1,\infty}(\Omega)$ and is bounded from below by a positive constant, $H^1(\Omega) = H_a^1(\Omega)$. Hence, $\phi \in H_a^1(\Omega)$ implies that $a \cdot u$, $\frac{\phi}{a} \in H^1(\Omega)$ due to the product rule (Theorem 3.2.9). Therefore, we can conclude that

$$\omega_a(u, \phi) = \left\langle a \cdot B_a u, \phi \right\rangle_a - \left\langle \frac{1}{a}\nabla a \cdot \nabla(au), \phi \right\rangle_a \tag{3.17}$$

for all $\phi \in H^1_a(\Omega)$. Hence, $u \in \mathscr{D}(B_a)$ implies $u \in \mathscr{D}(A_a)$. To establish that $u \in \mathscr{D}(A_a)$ implies $u \in \mathscr{D}(B_a)$, we can use a similar argument: if $u \in \mathscr{D}(A_a)$, then

$$\sigma_a(u, a\phi) = \left\langle A_a u, \phi \right\rangle_a + \left\langle \frac{1}{a} \nabla a \cdot \nabla(au), \phi \right\rangle_a \tag{3.18}$$

for all $\phi \in H^1_a(\Omega)$. □

In the following lemma, we will consider the space $(\mathscr{D}(A^m_a), \|\cdot\|_{a|m})$, where $\|\cdot\|_{a|m}$ is the norm given by $\|z\|_{a|m} = \|(\mathbb{I} + A_a)^m z\|_a$ for each $z \in \mathscr{D}(A^m_a)$ and $\mathbb{I}$ is the identity map $u \mapsto u$. This lemma will play an important role in the theorem on controllability, Theorem 3.2.16. It will be used to conclude that solutions of the parabolic systems (3.9) and (3.10) have bounded gradients for each $t > 0$, provided that the boundary of the domain Ω is regular enough. This will enable us to prove later on that the control inputs constructed to prove controllability are bounded.

Lemma 3.2.11. *Let $a \in W^{1,\infty}(\Omega)$. Let Ω be a domain that is either $C^{1,1}$ or convex. Then there exists $m \in \mathbb{Z}_+$ large enough such that, for some $C_m > 0$, $\|\nabla(a(\mathbf{x})u)\|_\infty \leq C_m(\|(\mathbb{I} + A_a)^m u\|_a)$ for all $u \in \mathscr{D}(A^m_a)$. Similarly, there exists $m' \in \mathbb{Z}_+$ large enough such that, for some $C_{m'} > 0$, $\|\nabla(a(\mathbf{x})u)\|_\infty \leq C_m(\|(\mathbb{I} + B_a)^{m'} u\|_a)$ for all $u \in \mathscr{D}(B^{m'}_a)$.*

Proof. First, we consider the case where Ω is a $C^{1,1}$ domain. Let $W^{2,p}(\Omega)$ be the set of elements in $L^p(\Omega)$ with second-order weak derivatives in $L^p(\Omega)$. Then we know that for the Neumann problem with $a = \mathbf{1}$,

$$\begin{aligned} -\Delta u + a_0 u &= f \ \text{ in } \Omega \\ \mathbf{n} \cdot \nabla u &= 0 \ \text{ in } \partial\Omega \end{aligned} \tag{3.19}$$

has solutions $u \in W^{2,p}(\Omega)$ if $f \in L^p(\Omega)$ whenever $1 < p < \infty$, $a_0 \in L^\infty(\Omega)$ (Grisvard, 2011)[Theorem 2.4.2.7] and $a_0 \geq \beta$ for some $\beta > 0$. These solutions have bounds

$$\|u\|_{W^{2,p}} \leq C_p \|f\|_{L^p} \tag{3.20}$$

for some constant $C_p > 0$. Let $\mathscr{M}_a : L^2(\Omega) \to L^2(\Omega)$ be the multiplication operator defined as $(\mathscr{M}_a u)(\mathbf{x}) = a(\mathbf{x})u(\mathbf{x})$ for a.e. $\mathbf{x} \in \Omega$ and each $u \in L^2(\Omega)$. Note that $u \in L^p(\Omega)$ implies that $\mathscr{M}_a u \in L^p(\Omega)$ for each $2 \leq p \leq \infty$. Suppose that $n > 2$, where we recall that n is the dimension of the Euclidean space $\mathbb{R}^n$ of which Ω is a subset. Note that Ω is an extension domain (Definition 3.1.2). Then from the $W^{2,p}$ *regularity estimate* (3.20) of equation (3.19) and from the embedding theorem (Leoni, 2009)[Corollary 11.9], it follows that $f \in L^2(\Omega)$ implies $(A_a + \mathbb{I})^{-m} f = (A_{\mathbf{1}} \mathscr{M}_a + \mathbb{I})^{-m} f = \left((A_{\mathbf{1}} + \mathscr{M}_a^{-1})\mathscr{M}_a\right)^{-m} f \in L^q(\Omega)$ for some $q \geq n$ for $m \in \mathbb{Z}_+$ large enough. For the general case $n \geq 1$, it follows from the embedding theorem (Leoni, 2009)[Theorem 11.23] that $f \in L^2(\Omega)$ implies $(A_a + \mathbb{I})^{-m} f \in L^q(\Omega)$ for any desired $n \leq q < \infty$, provided $m \in \mathbb{Z}_+$ for m large enough. Since $a \in W^{1,\infty}(\Omega)$, it follows from the $W^{2,p}$ regularity estimate (3.20) of equation (3.19) and *Morrey's inequality* (Leoni, 2009)[Theorem 11.34] that if $f \in L^2(\Omega)$, $(A_a + \mathbb{I})^{-m} f = u \in L^q(\Omega)$, and $m \in \mathbb{Z}_+$ is large enough, then $\|\nabla u\|_\infty \leq C_\infty \|f\|_2$, where $C_\infty > 0$ is independent of f.

A similar argument can be used when Ω is convex. However, it is not clear if the $W^{2,p}$ regularity estimate (3.20) holds for general convex domains. On the other hand, it can be established that the L^p *regularity estimate* of the PDE (3.19) holds for such domains. In particular, it follows from (Bakry *et al.*, 2013)[Corollary 6.3.3] and the embedding theorems (Leoni, 2009)[Corollary 11.9, Theorem 11.23] that for any $1 < p \leq \infty$, there exists $m \in \mathbb{Z}_+$ large enough such that $(A_a + \mathbb{I})^{-m}$ is a bounded operator from $L^2(\Omega)$ to $L^p(\Omega)$. This last statement uses only the extension property of the domain Ω, which is not required to be convex for the statement to hold true; the convexity of Ω is required mainly to derive the bounds on the gradient of the solution u. For this derivation, we use a result from (Maz'ya, 2009). Since $a \in W^{1,\infty}(\Omega)$, it follows from the theorem (Maz'ya, 2009)[Theorem] that there exists a constant $C'_\infty > 0$ such that if $f \in L^2(\Omega)$, $(A_a + \mathbb{I})^{-m} f = u$, and $m \in \mathbb{Z}_+$ is large enough, then $\|\nabla u\|_\infty \leq C'_\infty \|f\|_2$, where C'_∞ is independent of f. In this last statement, the theorem (Maz'ya, 2009)[Theorem] can be applied to derive the gradient bounds due to the

fact that $(A_a+\mathbb{I})^{-m+1}f \in L^q(\Omega)$ for some $q > n$ for $m \in \mathbb{Z}_+$ large enough.

For the operator B_a, the inequalities can be derived using exactly the same approach as done for A_a. Hence, we only point out the key results needed. Particularly, for a $C^{1,1}$ domain, the $W^{2,p}$ regularity estimate also holds for the equation $\nabla\cdot(\frac{1}{a(\mathbf{x})}\nabla u)+a_0u=f$ from (Grisvard, 2011)[Theorem 2.4.2.7]. For Ω being convex, the $W^{1,p}$ regularity estimate has been proved in (Geng, 2018)[Theorem 1.3] for solutions of elliptic operators in divergence form on convex domain. Since $a \in W^{1,\infty}(\Omega)$, using the product rule (Theorem 3.2.9), the gradient bounds for the Neumann Laplacian (Maz'ya, 2009)[Theorem] also give the desired gradient bounds of the operator $\nabla\cdot(\frac{1}{a(\mathbf{x})}\nabla\cdot)$ □

Lemma 3.2.12. *Let Ω be a domain that is either $C^{1,1}$ or convex. Let $y^0 \in \mathscr{D}(A_a^m)$ for some $m \in \mathbb{Z}_+$. Then the mild solution, $y \in C([0,\infty);L^2_a(\Omega))$, of the PDE* (3.9) *satisfies $y(\cdot,t) \in \mathscr{D}(A_a^m)$ for each $t \in [0,\infty)$. Moreover, the following estimates hold for some positive constants M_m and λ:*

$$\|y(\cdot,t)-f\|_{a|m} \;\leq\; M_m e^{-\lambda t}. \tag{3.21}$$

Proof. We are given that $y^0 \in \mathscr{D}(A_a^m)$. Since the semigroup $(\mathscr{T}_a^A(t))_{t\geq 0}$ and its generator $-A_a$ commute, we know that $\|y(\cdot,t)-f\|_{a|m} = \|(\mathbb{I}+A_a)^m(\mathscr{T}_a^A(t))(y^0-f)\|_a = \|\mathscr{T}_a^A(t)(\mathbb{I}+A_a)^m(y^0-f)\|_a \leq M_0e^{-\lambda t}\|y^0-f\|_{a|m}$ for some positive constants M_0 and λ. □

Since controllability will first be proved in Lemma 3.2.14 under the assumption that the initial condition is bounded from below by a positive constant, the following lemma will be used to relax this assumption in Theorem 3.2.16.

Lemma 3.2.13. *Let Ω be a domain that is either $C^{1,1}$ or convex and that satisfies the chain condition (see Definition 3.1.3). Let $y^0 \in L^2(\Omega)$ be such that $y^0 \geq 0$. Let $y \in C([0,T];L^2(\Omega))$ be the unique mild solution of the PDE* (3.9)*. Then for all $t \in (0,\infty)$, there exists a positive constant, $c_t > 0$, such that $y(\cdot,t) \geq c_t$.*

Proof. Consider the heat equation with Neumann boundary condition, that is, the PDE (3.5) with $\mathbf{v} \equiv \mathbf{0}$. The solution y of this PDE can be represented using the *Neumann heat kernel* K. That is, there exists a measurable map $K : (0,\infty) \times \Omega^2 \to [0,\infty)$ such that the mild solution y can be constructed using the relation $y(\mathbf{x},t) = \int_\Omega K(t,\mathbf{x},\mathbf{z}) y^0(\mathbf{z}) d\mathbf{z}$ for each $t \in (0,\infty)$ and almost every $\mathbf{x} \in \Omega$. From (Choulli and Kayser, 2015)[Theorem 3.1] (for $C^{1,1}$ domains) and (Li and Yau, 1986)[Corollary 2.1] (for convex domains), for some $C > 0$, we know that $K(t,\mathbf{x},\mathbf{z}) \geq \frac{C}{(4\pi t)^{1/2}} \exp(\frac{-|\mathbf{x}-\mathbf{z}|^2}{4t})$ for each $t > 0$ and almost every $\mathbf{x},\mathbf{z} \in \Omega$. From this, the lower bound on $y(\cdot,t)$ follows. □

Lemma 3.2.14. *Let $y^0 \in \mathscr{D}(A_a^m)$ for some $m \in \mathbb{Z}_+$ such that $y^0 \geq c_1$ for some positive constant c_1. Suppose that $f \in W^{1,\infty}(\Omega)$ such that $f \geq c_2$ for some positive constant c_2 and $\int_\Omega f(\mathbf{x}) d\mathbf{x} = \int_\Omega y^0(\mathbf{x}) d\mathbf{x}$. Let $t_f = \sum_{k=1}^\infty \frac{1}{k^2}$ be the final time. Define the vector field $\mathbf{v}$ in the PDE* (3.5) *by*

$$\mathbf{v}(\cdot,t) = \frac{\nabla y}{y} - \alpha j \frac{\nabla(ay)}{y} \tag{3.22}$$

for some $\alpha > 0$, $j \in \mathbb{Z}_+$, where $a = 1/f$ whenever $t \in [\sum_{k=1}^{j-1} \frac{1}{k^2}, \sum_{k=1}^{j} \frac{1}{k^2})$. Here, we define $\sum_{k=1}^{j} \frac{1}{k^2} = 0$ if $j = 0$.

If Ω is a domain that is $C^{1,1}$ or convex, then there exists $m \in \mathbb{Z}_+ \setminus \{0\}$ large enough and $\alpha > 0$ such that $\mathbf{v} \in L^\infty([0,t_f]; L^\infty(\Omega)^n)$ and $y(\cdot,t_f) = f$.

Proof. Substituting $\mathbf{v}(\cdot,t) = \frac{\nabla y}{y} - \alpha j \frac{\nabla(ay)}{y}$ whenever $t \in [\sum_{k=1}^{j-1} \frac{1}{k^2}, \sum_{k=1}^{j} \frac{1}{k^2})$ in the PDE (3.5), it can be seen that if the solution of this PDE exists, then it can be constructed from mild solutions of the *closed-loop* PDE

$$\begin{aligned} \tilde{y}_t &= \alpha j \Delta(a(\mathbf{x})\tilde{y}) && \text{in } \Omega \times [0, \frac{1}{j^2}) \\ \tilde{y}(\cdot,0) = \tilde{y}^0 &= y(\cdot, \textstyle\sum_{k=1}^{j-1} \frac{1}{k^2}) && \text{in } \Omega \\ \mathbf{n} \cdot \nabla(a\tilde{y}) &= 0 && \text{in } [0, \frac{1}{j^2}), \end{aligned} \tag{3.23}$$

and we obtain the relation $y(\cdot, \sum_{k=1}^{j-1} \frac{1}{k^2} + i) = \tilde{y}(\cdot, i)$ for each $i \in [0, \frac{1}{j^2})$ and each $m \in \mathbb{Z}_+$. Since y^0 is uniformly bounded from below by a positive constant, $\tilde{y}$ is also uniformly

bounded from below according to Lemma 3.2.13. Moreover, since $a \in W^{1,\infty}(\Omega)$, it follows that $\mathscr{D}(A_a) \subset H^1(\Omega)$. Hence, the velocity field $\mathbf{v}$ is well-defined for all t in the half-open interval $[0,t_f)$.

It follows from Lemma 3.2.7 that $\|y(\cdot,\sum_{k=1}^{j}\frac{1}{k^2}) - f\|_a \leq M_0 e^{-\alpha\lambda\sum_{k=1}^{j}\frac{k}{k^2}} = M_0 e^{-\alpha\lambda\sum_{k=1}^{j}\frac{1}{k}}$ for each $j \in \mathbb{Z}_+$, for some positive constants M_0 and λ independent of j. Since the summation $\sum_{k=1}^{j}\frac{1}{k}$ is diverging, we have that $y(\cdot,t_f) = f$ if the solution is defined over the interval $[0,t_f]$. Since y is continuous on $[0,t_f)$ and uniformly bounded, it follows that y is in $C([0,t_f);L^2_a(\Omega))$ and can be extended to a unique mild solution $y \in C([0,t_f];L^2_a(\Omega))$ defined over the time interval $[0,t_f]$.

It is additionally required to prove the existence of $m \in \mathbb{Z}_+$ and $\alpha > 0$ such that $\mathbf{v} \in L^\infty([0,t_f];L^\infty(\Omega)^n)$. First, we derive bounds on the term $1/y(\cdot,t)$. Due to the lower bound on the initial condition y^0, and noting that $y(\cdot,t) = \mathscr{T}^A_a(\tilde{t})y^0$ for some $\tilde{t} \in [0,\infty)$ depending on $t \in [0,t_f)$, it follows from Lemma 3.2.8 that there exists a positive constant d such that

$$y(\cdot,t) \geq d \tag{3.24}$$

for all $t \in [0,t_f)$. This gives us the uniform upper bound $1/d$ on the term $1/y(\cdot,t)$. Next, we consider the term $\alpha\nabla(a(\mathbf{x})y(\cdot,t))$. We note that $y_0 \in \mathscr{D}(A^j_a)$. Hence, we can apply the estimate in Lemma 3.2.12 to obtain

$$\|y(\cdot,\sum_{k=1}^{m}\frac{1}{k^2}) - f\|_{a|m} \;\leq\; \tilde{M}e^{-\alpha\lambda\sum_{k=1}^{j}\frac{1}{k}}$$

for some positive constants $\tilde{M}$ and λ. From Lemma 3.2.11, it follows that when Ω is a domain that is $C^{1,1}$ or convex, there exists $C > 0$ depending only on a such that

$$\|\alpha j\nabla(a(\mathbf{x})y)(\cdot,\sum_{k=1}^{j}\frac{1}{k^2})\|_\infty \;\leq\; C\alpha j\tilde{M}e^{-\alpha\lambda\sum_{k=1}^{j}\frac{1}{k}} \tag{3.25}$$

for some positive constants $\tilde{M}$ and λ. The right-hand side of the estimate (3.25) is not uniformly bounded for arbitrary $\alpha > 0$ due to its dependence on j. However, we note

that $\lim_{j\to\infty} -\ln\, j + \sum_{k=1}^{j} \frac{1}{k} = \gamma$, where $\gamma > 0$ is the *Euler-Mascheroni* constant (Finch, 2003)[Section 1.5]. Therefore, by setting $\alpha \geq 1/\lambda$, the right-hand side becomes uniformly bounded for all $j \in \mathbb{Z}_+$. Since $a \in W^{1,\infty}(\Omega)$, it follows from the product rule and the estimate (3.25) that

$$\|\nabla y(\cdot,t)\|_\infty \;\leq\; C_2 \tag{3.26}$$

for some positive constant C_2 and for all $t \in [0,t_f)$.

From the estimates (3.24)-(3.26), it follows that if $\alpha > 0$ is large enough, then $\mathbf{v} \in L^\infty([0,t_f];L^\infty(\Omega)^n)$ and $y(\cdot,t_f) = f$. This concludes the proof for the case when the domain Ω is $C^{1,1}$ or convex. □

Note that any control law of the form $\mathbf{v}(\cdot,t) = \frac{\nabla y}{y} - \alpha m^\beta \frac{\nabla(ay)}{y}$ for numerous other values of β and α will also achieve the desired controllability objective, due to the fact that an exponential function of a variable grows faster than a polynomial function as the variable tends to infinity. Additionally, we could replace the parameter m with a continuous function $m(t)$ such that $\int_0^T m(\tau)d\tau = \infty$.

The following corollary follows from Lemma 3.2.14 using a straightforward scaling argument.

Corollary 3.2.15. *Let $y^0 \in \mathscr{D}(A_a^m)$ be such that $y^0 \geq c_1$ for some positive constant c_1 and $m \in \mathbb{Z}_+$. Let Ω be a domain that is either $C^{1,1}$ or convex. Suppose that $f \in W^{1,\infty}(\Omega)$ such that $f \geq c_2$ for some positive constant c_2, $\int_\Omega f(\mathbf{x})d\mathbf{x} = \int_\Omega y^0(\mathbf{x})d\mathbf{x}$, and $a = 1/f$. Let $t_f > 0$ be the final time. Then there exists $\mathbf{v} \in L^\infty([0,t_f];L^\infty(\Omega)^n)$ such that the mild solution y of the PDE* (3.5) *satisfies $y(\cdot,t_f) = f$.*

Now, we are ready to state and prove our main theorem, where we relax the assumptions on the initial condition y^0 made in Corollary 3.2.15. However, we will need to impose the additional constraint that Ω should satisfy the chain condition.

Theorem 3.2.16. *Let Ω be a domain that is either $C^{1,1}$ or convex and that satisfies the chain condition. Let $y^0 \in L^2(\Omega)$ be such that $y^0 \geq 0$ and $\int_\Omega y^0(\mathbf{x})d\mathbf{x} = 1$. Suppose that $f \in W^{1,\infty}(\Omega)$ such that $f \geq c$ for some positive constant c, $\int_\Omega f(\mathbf{x})d\mathbf{x} = 1$. Let $t_f > 0$ be the final time. Then there exists $\mathbf{v} \in L^\infty([0,t_f];L^\infty(\Omega)^n)$ such that the unique mild solution y of the PDE* (3.5) *satisfies $y(\cdot,t_f) = f$.*

Proof. Set $\mathbf{v}(\cdot,t) = \mathbf{0}$ in the PDE (3.5) for each $t \in [0,\varepsilon/3]$, where $\varepsilon \in (0,t_f)$ is small enough. Then this PDE is the heat equation with Neumann boundary condition. From Lemma 3.2.13, it follows that the solution y satisfies $y(\cdot,\varepsilon/2) \geq c$ for some positive constant c. For each $t \in (\varepsilon/3,2\varepsilon/3]$, let $\mathbf{v}(\cdot,t) = \frac{\nabla f}{f}$. Then the mild solution of the PDE is given by the semigroup $(\mathscr{T}_a^B(t))_{t\geq 0}$, where $a = 1/f$. From Lemma 3.2.7, the semigroup $(\mathscr{T}_a^B(t))_{t\geq 0}$ is analytic. Hence, from regularizing properties of analytic semigroups (Lunardi, 2012)[Theorem 2.1.1], it follows that $y(\cdot,\varepsilon) \in \mathscr{D}(B_a^j)$ for each $j \in \mathbb{Z}_+$. From Lemma 3.2.11, this implies that $\|B_a y(\cdot,2\varepsilon/3)\|_\infty < \infty$. Due to the product rule (Theorem 3.2.9), Proposition 3.2.10, and Lemma 3.2.11, this inequality implies that $\|A_a y(\cdot,2\varepsilon/3)\|_\infty < c$ for some $c > 0$. Let $\mathbf{v}(\cdot,t) = \frac{\nabla y}{y} - \frac{\nabla(ay)}{y}$ for $t \in [2\varepsilon/3,\varepsilon]$. Then from the last statement of Corollary 3.2.5 and the fact that the operator $-A_a$ commutes with the semigroup it generates, it follows that $\|\mathscr{M}_a A_a y(\cdot,t)\|_\infty = \|\mathscr{M}_a \mathscr{T}_a^A(t-2\varepsilon/3)_a A_a y(\cdot,2\varepsilon/3)\|_\infty < c'$ for some $c' > 0$ and for all $t \in (2\varepsilon/3,\varepsilon]$. Since $a \in W^{1,\infty}(\Omega)$, we can apply the result in Proposition 3.2.10 and the gradient estimates of the Neumann Laplacian in (Grisvard, 2011)[Theorem 2.4.2.7] and (Maz'ya, 2009)[Theorem]. Taken together, all of these observations imply that $\|\nabla y(\cdot,t)\|_\infty < k$ for some $k > 0$ and all $t \in (2\varepsilon/3,\varepsilon]$. From Lemma 3.2.8, it also follows that y is uniformly bounded from below, and hence $\mathbf{v}(\cdot,t)$ is essentially bounded for all $t \in [0,\varepsilon]$. Lastly, due to the analyticity of the semigroup $(\mathscr{T}_a^A(t))_{t\geq 0}$, $y(\cdot,\varepsilon) \in \mathscr{D}(A_a^j)$ for each $j \in \mathbb{Z}_+$. Then the result follows from Corollary 3.2.15. □

In the following theorem, we note that system (3.5) has stronger controllability proper-

ties than Theorem 3.2.16 describes: this system is *path controllable* if the path is confined to a subset of $L^2(\Omega)$ that is regular enough. This should not be surprising due to the large dimensionality of the system's control inputs as compared to the choice of controls in classical PDE control problems, where control inputs are typically localized on a small subset of the interior or boundary of the domain. We restrict the path to the space $W^{2,\infty}(\Omega)$ for simplicity.

Theorem 3.2.17. *Let Ω be a domain that is either $C^{1,1}$ or convex. Suppose that $\gamma \in C^1([0,1];W^{2,\infty}(\Omega))$ such that $\gamma(t) \geq c$ for some positive constant c and for all $t \in [0,1]$. Additionally, suppose that $\int_\Omega \gamma(\mathbf{x},t)d\mathbf{x} = 1$ for all $t \in [0,1]$. Then there exists $\mathbf{v} \in L^\infty([0,1];L^\infty(\Omega)^n)$ such that a solution of the PDE (3.5) satisfies $y(t) = \gamma(t)$ for all $t \in [0,1]$.*

Proof. Fix $t \in [0,1]$. Consider the solution $\phi(t) \in L^2(\Omega)$ of the Poisson equation in weak form,

$$\omega_{\mathbf{1}}(\phi(t),\mu) = \left\langle \frac{\partial \gamma}{\partial t}(t), \mu \right\rangle \text{ for all } \mu \in H^1(\Omega), \tag{3.27}$$

where $\mathbf{1}$ is the function taking value 1 everywhere on Ω. Note that since $\int_\Omega \gamma(\mathbf{x},t)d\mathbf{x} = 1$ for all $t \in [0,1]$, we have that $\int_\Omega \frac{\partial \gamma}{\partial t}(\mathbf{x},t)d\mathbf{x} = 0$ for each $t \in [0,1]$, and therefore the Poisson equation has a unique solution for each $t \in [0,1]$. Then it follows from (Grisvard, 2011)[Theorem 2.4.2.7] and Morrey's inequality (Leoni, 2009)[Theorem 11.34] (when Ω is a $C^{1,1}$ domain) and (Maz'ya, 2009)[Theorem] (when Ω is convex) that there exists a constant C such that $\|\nabla\phi(t)\|_\infty \leq C\|\nabla(\partial\gamma(t)/\partial t)\|_2 \leq C\|\nabla(\partial\gamma(t)/\partial t)\|_\infty$ for each $t \in [0,1]$. Then setting $\mathbf{v}(\cdot,t) = \frac{\nabla\gamma(t)}{\gamma(t)} - \frac{\nabla\phi(t)}{\gamma(t)}$ for each $t \in [0,1]$ gives us the desired controllability result. □

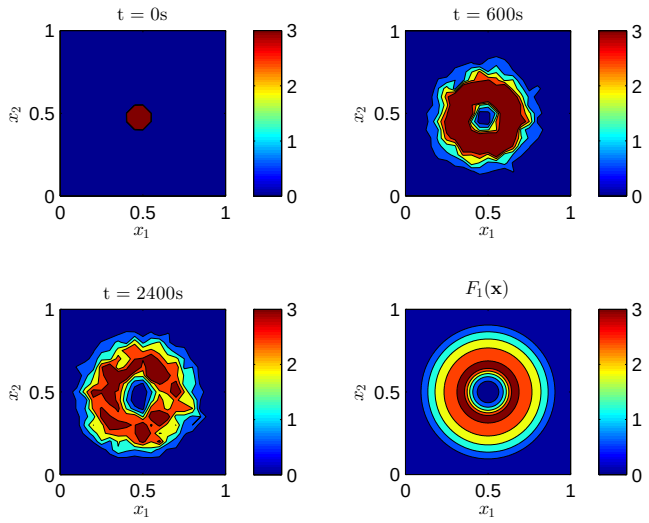


Figure 3.1: Simulated Agent Densities at Three Times t and the Underlying Scalar Field.

3.2.1 Simulation

In this subsection, we validate the stability result presented in this section. This result was presented in (**Elamvazhuthi** *et al.*, 2016) to verify the stabilization of a simulated swarm to a target probability density. The results have also been experimentally validated in (Li *et al.*, 2017) with a single robot over multiple trials.

We validate our coverage approach in a simulated scenario. The scalar field is defined as $F_1(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}) + \varepsilon$ for all $\mathbf{x} \in \Omega$, where f_n, $n = 1,2$, are given by

$$f_n(\mathbf{x}) = \exp\left(\frac{-1}{1 - \|a_n\mathbf{x} - b_n\|^2}\right) \qquad \text{for } \|a_n\mathbf{x} - b_n\|^2 < 1,$$
$$= 0 \quad \text{otherwise.}$$

We set $a_1 = 2$, $a_2 = 6$, $b_1 = 1$, $b_2 = 2$, and $\varepsilon = 0.01$. The field $F_1(\mathbf{x})$ is shown in the lower right plot of Fig. 3.1.

The diffusion-based feedback control law was chosen to be $D_n(\mathbf{x}) = 10^{-5}/F_n(\mathbf{x})^{1/2}$, $n = 1,2$. Since D_n is in $C^\infty(\bar{\Omega})$ and is uniformly bounded from below away from zero, it

is globally Lipschitz on Ω. For the stochastic simulation, 3000 agents were simulated on a domain $\Omega = (0,1) \times (0,1)$. The agents were initially distributed as a Gaussian centered at $(0.5, 0.5)$. The stochastic motion of each agent was approximated in discrete time using the *Euler-Maruyama scheme* (Talay, 1994):

$$\mathbf{X}(t+\Delta t) - \mathbf{X}(t) = (2D_n^2(\mathbf{X})\Delta t)^{1/2}\, \mathbf{Z}(t), \tag{3.28}$$

where $\mathbf{Z} \in \mathbb{R}^2$ is a vector of independent, standard normal random variables. When an agent encounters the boundary, it performs a specular reflection. As shown in Figs. 3.1, the steady-state swarm density closely matches the underlying scalar field.

3.3 Controllability of a System of Advection-Diffusion-Reaction Equations

In models of robotic swarms, we consider the hybrid variants of the SDE (3.4) to account for the fact that each robot, in addition to a continuous spatial state $\mathbf{Z}(t)$, can be associated with a discrete state $Y(t) \in \mathscr{V} = \{1, ..., N\}$ at each time t (Milutinovic and Lima, 2006; **Elamvazhuthi** *et al.*, 2018c). This model was introduced in Section 1.2.2. In this case, the state of each agent is denoted by the pair $(\mathbf{Z}(t), Y(t)) \in \Omega \times \mathscr{V}$. Suppose that the variable $Y(t)$ evolves according to a CTMC. We define a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ in which the vertex set $\mathscr{V}$ is the set of discrete states, and the edge set $\mathscr{E}$ defines the possible agent transitions between the discrete states in $\mathscr{V}$. The agents' transition rules are determined by the control parameters $u_e : [0, \infty) \to U$ for each $e \in \mathscr{E}$, also known as the *transition rates* of the associated CTMC. Here $U \subset \mathbb{R}_+$ is the set of *admissible transition rates*.

The variable $Y(t)$ evolves on the state space $\mathscr{V}$ according to the conditional probabilities

$$\mathbb{P}(Y(t+h) = T(e) \mid Y(t) = S(e)) = u_e(t)h + o(h) \tag{3.29}$$

for each $e \in \mathscr{E}$. Let $\mathscr{P}(\mathscr{V}) = \{\mathbf{y} \in \mathbb{R}_+^N : \sum_v y_v = 1\}$ be the simplex of probability densities on $\mathscr{V}$. Corresponding to the CTMC is a set of ODEs that determine the time evolution of

the probability densities $\mathbb{P}(Y(t)=v)=\mu_v(t)\in\mathbb{R}_+$. The forward equation is given by a system of linear ODEs,

$$\begin{aligned}\dot{\mu}(t) &= \sum_{e\in\mathscr{E}} u_e(t)\mathbf{Q}_e\mu(t), \quad t\in[0,\infty), \\ \mu(0) &= \mu^0\in\mathscr{P}(\mathscr{V}),\end{aligned} \tag{3.30}$$

where $\mathbf{Q}_e$ are control matrices whose entries are given by

$$Q_e^{ij}=\begin{cases}-1 & \text{if } i=j=S(e),\\ 1 & \text{if } i=T(e),\ j=S(e),\\ 0 & \text{otherwise.}\end{cases}$$

Given these definitions, we can define a hybrid switching diffusion process $(\mathbf{Z}(t),Y(t))$ as a system of SDEs of the form

$$\begin{aligned}d\mathbf{Z}(t) &= \mathbf{v}(Y(t),\mathbf{Z}(t),t)dt+\sqrt{2\mathbf{D}}d\mathbf{W}(t)+\mathbf{n}(\mathbf{Z}(t))d\psi(t),\\ \mathbf{Z}(0) &= \mathbf{Z}_0,\end{aligned} \tag{3.31}$$

where $\mathbf{v}:\mathscr{V}\times\Omega\times[0,t_f]\to\mathbb{R}^n$ is the state- and time-dependent velocity vector field, and $\mathbf{D}\in\mathbb{R}_+^N$ is a vector of positive elements. Here, D_k is the diffusion parameter associated with each discrete state $k\in\mathscr{V}$. Let $\mathbf{v}_k$ denote the velocity field associated with discrete state $k\in\mathscr{V}$. Then the forward equation for this system of SDEs is given by the system of PDEs

$$\begin{aligned}(y_k)_t = D_k\Delta y_k-\nabla\cdot(\mathbf{v}_k(\mathbf{x},t)y_k)+\mathscr{F}_k \quad & in\ \ \Omega\times[0,t_f]\\ y_k(\cdot,0)=y_k^0 \quad & in\ \ \Omega\\ \mathbf{n}\cdot(\nabla y_k-\mathbf{v}_k(\mathbf{x},t)y_k)=0 \quad & in\ \ \partial\Omega\times[0,t_f],\end{aligned} \tag{3.32}$$

where $k\in\mathscr{V}$ and $\mathscr{F}_k=\sum_{e\in\mathscr{E}}\sum_{j\in\mathscr{V}}u_e(t)Q_e^{kj}y_j$.

We can pose a problem for the system of SDEs (3.31) that is similar to the one defined in Problem 3.2.2, with a target spatial distribution assigned to each discrete state:

Problem 3.3.1. *Given $t_f > 0$, $\mathbf{y}^0 : \Omega^N \to \mathbb{R}_+$, and $\mathbf{f} : \Omega^N \to \mathbb{R}_+$ such that $\sum_{i\in\mathscr{V}} \int_\Omega y_i^0(\mathbf{x})d\mathbf{x} = \sum_{i\in\mathscr{V}} \int_\Omega f_i(\mathbf{x})d\mathbf{x} = 1$, determine whether there exists a set of space- and time-dependent parameters $\mathbf{v}_k : \Omega \times [0,t_f] \to \mathbb{R}^n$ and time-dependent parameters $u_e : [0,t_f] \to \mathbb{R}_+$ such that the solution $\mathbf{y}$ of the system of PDEs* (3.32) *satisfies $\mathbf{y}(\cdot,t_f) = \mathbf{f}$.*

In order to prove the results in this section, we note the following result (**Elamvazhuthi** *et al.*, 2019) on the controllability of system (2.2) using piecewise-constant controls.

Theorem 3.3.2. ***(Elamvazhuthi*** et al.*, 2019) Let $T > 0$. If the graph $\mathscr{G} = (\mathscr{V},\mathscr{E})$ is strongly connected, then the system* (2.2) *is globally controllable within time T from every point in the interior of the simplex $\mathscr{P}(\mathscr{V})$, using piecewise-constant control inputs.*

We define some new notation that will be needed in this section and the following one. These definitions will be used to construct solutions of the system of PDEs (3.32) and hence enable the controllability and stability analysis.

Let $\mathbf{a} = [a_1\ a_2\ ...\ a_N]^T \in \mathbf{L}^\infty(\Omega) = Z_1 \times ... \times Z_N$, where $a_i \in L^\infty(\Omega)$ and $Z_i = L^\infty(\Omega)$ for each $i \in \mathscr{V}$. If $c > 0$, then we write $\mathbf{a} \geq c$ to denote that $a_i \geq c$ for each $i \in \mathscr{V}$. We will assume throughout that this condition is satisfied by $\mathbf{a}$ for some positive constant c. We consider the operator $\mathscr{B}_\mathbf{a} : \mathscr{D}(\mathscr{B}_\mathbf{a}) \to \mathbf{L}^2_\mathbf{a}(\Omega)$, where $\mathbf{L}^2_\mathbf{a}(\Omega) = L^2_{a_1}(\Omega) \times ... \times L^2_{a_N}(\Omega)$ is equipped with the norm $\|\cdot\|_\mathbf{a}$, defined as $\|\mathbf{u}\|_\mathbf{a} = (\sum_{i=1}^N \|u_i\|_a^2)^{1/2}$ for each $\mathbf{u} = [u_1\ ...\ u_N]^T \in \mathbf{L}^2_\mathbf{a}(\Omega)$, and $\mathscr{D}(\mathscr{B}_\mathbf{a}) = \mathscr{D}(B_{a_1}) \times \mathscr{D}(B_{a_2}) \times ... \times \mathscr{D}(B_{a_N})$. The operator $\mathscr{B}_\mathbf{a}$ is defined by $\mathscr{B}_\mathbf{a}\mathbf{v} = [B_{a_1}v_1\ B_{a_2}v_2\ ...\ B_{a_N}v_N]^T$ for each $\mathbf{v} = [v_1\ ...\ v_N]^T \in \mathscr{D}(\mathscr{B}_\mathbf{a})$. Recall that, formally, B_a is the operator $\Delta(a\ \cdot\ )$ for a given positive function $a \in L^\infty(\Omega)$. Corresponding to each matrix $\mathbf{Q}_e$, we associate a bounded operator $\mathscr{Q}_e$ on the space $\mathbf{L}^2_\mathbf{a}(\Omega)$ given by $(\mathscr{Q}_e\mathbf{y})(\mathbf{x}) = \mathbf{Q}_e\mathbf{y}(\mathbf{x})$ for each $\mathbf{y} = [y_1\ ...\ y_N]^T \in \mathbf{L}^2_\mathbf{a}(\Omega)$ and a.e. $\mathbf{x} \in \Omega$. Let $b \in \mathbf{L}^\infty(\Omega)$. $\mathscr{M}_\mathbf{b}$ will denote the multiplication operator defined by $\mathscr{M}_\mathbf{b}\mathbf{v} = [\mathscr{M}_{b_1}v_1\ \mathscr{M}_{b_2}v_2\ ...\ \mathscr{M}_{b_N}v_N]^T$ for each

$\mathbf{v} \in \mathbf{L}^2(\Omega) = \mathbf{L}^2_{\mathbf{a}}(\Omega)$. For a function $K_e \in L^\infty(\Omega)$, $K_e\mathscr{Q}_e$ will denote the product operator $\mathscr{M}_{\mathbf{b}}\mathscr{Q}_e$, where $\mathscr{M}_{\mathbf{b}}$ is the multiplication operator corresponding to the function $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ defined by setting $b_i = K_e$ for each $i \in \mathscr{V}$.

Lemma 3.3.3. *Let $\{K_e\}_{e\in\mathscr{E}}$ be a set of non-negative functions in $L^\infty(\Omega)$. Suppose $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ such that $b_i = D_i\mathbf{1}$ is a positive constant function for each $i \in \mathscr{V}$. Then the operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e$ generates a semigroup of operators $(\mathscr{S}(t))_{t\geq 0}$ on $\mathbf{L}^2_{\mathbf{a}}(\Omega)$. Moreover, the semigroup is positive and mass-conserving, i.e., if $\mathbf{y}^0 \in \mathbf{L}^2_{\mathbf{a}}(\Omega)$ is real-valued, then $\sum_{i\in\mathscr{V}} \int_\Omega (\mathscr{S}(t)\mathbf{y}^0)_i(\mathbf{x})d\mathbf{x} = \sum_{i\in\mathscr{V}} \int_\Omega y_i(\mathbf{x})d\mathbf{x}$ for all $t \geq 0$.*

Additionally, if $a_i \in W^{1,\infty}(\Omega)$, then $\mathscr{S}(t)\mathbf{y}^0$ is the unique mild solution of the system (3.32) with $f_i = 1/a_i$, $\mathbf{v}_i(\cdot,t) = D_i\nabla f_i/f_i$, and $u_e(t) = K_e$ for all $i \in \mathscr{V}$, all $e \in \mathscr{E}$, and all $t \in [0,t_f]$.

Proof. The generation of the semigroup $(\mathscr{S}(t))_{t\geq 0}$ follows from the fact that $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e$ is a bounded perturbation of the operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}}$. The positivity preserving property of the semigroup can be demonstrated as follows using the *Lie-Trotter product formula* (Engel and Nagel, 2000)[Corollary III.5.8]. Let $(\mathscr{U}(t))_{t\geq 0}$ be the semigroup generated by the operator $\sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e$. In fact, the semigroup can be explicitly represented as $\mathscr{U}(t) = e^{\sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e t}$ for each $t \geq 0$. Moreover, $(e^{\sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e t}\mathbf{y}^0)(\mathbf{x}) = e^{\sum_{e\in\mathscr{E}} K_e(\mathbf{x})\mathbf{Q}_e t}\mathbf{y}^0(\mathbf{x})$ for each $\mathbf{y}^0 \in \mathbf{L}^2_{\mathbf{a}}(\Omega)$ and a.e. $\mathbf{x} \in \Omega$. The semigroup $(\mathscr{U}(t))_{t\geq 0}$ is positivity preserving since each matrix $\mathbf{Q}_e$ has positive off-diagonal entries. From Corollary 3.2.5, we also note that the semigroup $(\mathscr{V}(t))_{t\geq 0}$ generated by the operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}}$ is positivity preserving. Moreover, since $\sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e$ is a bounded operator, there exists $w \in \mathbb{R}$ such that $\|\mathscr{U}(t)\|_{op} \leq Me^{wt}$ for some positive constant M for all $t \geq 0$. The semigroup $(\mathscr{V}(t))_{t\geq 0}$ is *contractive*, i.e., $\|\mathscr{V}(t)\|_{op} \leq 1$ for all $t \geq 0$. Hence, it follows from the Lie-Trotter product formula that $\mathscr{S}(t)\mathbf{y}^0 = \lim_{n\to\infty}[\mathscr{V}(t/n)\mathscr{U}(t/n)]^n\mathbf{y}^0$ for all $t \geq 0$. Therefore, the semigroup $(\mathscr{S}(t))_{t\geq 0}$ is positivity preserving. Through another application of the Lie-Trotter product

formula, it follows that $\sum_{i\in\mathscr{V}}\int_\Omega(\mathscr{S}(t)\mathbf{y}^0)_i(\mathbf{x})d\mathbf{x}=\sum_{i\in\mathscr{V}}\int_\Omega y_i^0(\mathbf{x})d\mathbf{x}$ for all $t\geq 0$. □

In the following lemma, we identify a relation between solutions of the system of PDEs (3.32) and solutions of the ODE (2.2).

Lemma 3.3.4. *Let $\{q_e\}_{e\in\mathscr{E}}$ be a set of non-negative constants. Suppose $\mathbf{b}\in\mathbf{L}^\infty(\Omega)$ such that $b_i=D_i\mathbf{1}$ is a positive constant function for each $i\in\mathscr{V}$. Let $(\mathscr{S}(t))_{t\geq 0}$ be the semigroup generated by the operator $-\mathscr{M}_\mathbf{b}\mathscr{B}_\mathbf{a}+\sum_{e\in\mathscr{E}}q_e\mathscr{Q}_e$. Additionally, assume that $\mathbf{y}^0\in\mathscr{D}(-\mathscr{M}_\mathbf{b}\mathscr{B}_\mathbf{a})$ is real-valued and that $\mu_i^0=\int_\Omega y_i^0(\mathbf{x})d\mathbf{x}$ for each $i\in\mathscr{V}$. Then $\mu_i(t)$, given by $\mu_i(t)=\int_\Omega(\mathscr{S}(t)\mathbf{y}^0)_i(\mathbf{x})d\mathbf{x}$ for each $t\geq 0$ and each $i\in\mathscr{V}$, is a solution of the system* (2.2).

Proof. Let $\mathbf{y}(\cdot,t)=\mathscr{S}(t)\mathbf{y}^0$ for each $t\geq 0$. Then the result follows by noting that

$$\begin{aligned}
&\frac{d}{dt}\int_\Omega y_i(\mathbf{x},t)d\mathbf{x}\\
&\quad=\int_\Omega D_iB_{a_i}y_i(\mathbf{x},t)d\mathbf{x}+\sum_{j=1}^N\sum_{e\in\mathscr{E}}\int_\Omega q_eQ_e^{ij}y_i(\mathbf{x},t)d\mathbf{x}\\
&\quad=D_i\sigma_{a_i}(y_i,1/a_i)+\sum_{j=1}^N\sum_{e\in\mathscr{E}}q_eQ_e^{ij}\int_\Omega y_i(\mathbf{x},t)d\mathbf{x}\\
&\quad=\sum_{j=1}^N\sum_{e\in\mathscr{E}}q_eQ_e^{ij}\int_\Omega y_i(\mathbf{x},t)d\mathbf{x}
\end{aligned}$$

for all $t\geq 0$. □

The lemma above allows us to apply the results of Theorems 3.2.16 and 3.3.2 to prove the following controllability result, which addresses Problem 3.3.1.

Theorem 3.3.5. *Let Ω be a domain that is $C^{1,1}$ or convex and that satisfies the chain condition. Let $t_f>0$. Let $f_i\in W^{1,\infty}(\Omega)$ for each $i\in\mathscr{V}$ such that $f_i\geq c$ for some positive constant c. Suppose $\mathbf{y}^0\in\mathbf{L}^2(\Omega)$ such that $\mathbf{y}^0\geq 0$ and $\sum_i\int_\Omega f_i(\mathbf{x})d\mathbf{x}=\sum_i\int_\Omega y_i^0(\mathbf{x})d\mathbf{x}$. Then there exist control parameters $\{v_i\}_{i\in V}$ in $L^\infty([0,t_f];L^\infty(\Omega)^n)$ and $\{u_e\}_{e\in\mathscr{E}}$ in $L^\infty([0,t_f])$, where each u_e is non-negative, such that the unique mild solution of the system* (3.32) *satisfies $y_i(\cdot,t_f)=f_i$ for each $i\in\mathscr{V}$.*

Proof. Let $\mathbf{v}_i(\cdot,t) = \mathbf{0}$ for each $t \in [0,t_f/2]$ and for each $i \in \mathscr{V}$. Then from Theorem 3.3.2 and Lemmas 3.3.3 and 3.3.4, it follows that there exist piecewise constant parameters $u_e : [0,t_f/2] \to \mathbb{R}_+$ such that the mild solution of the PDE (3.32) satisfies $\int_\Omega y_i(\mathbf{x},t_f/2)d\mathbf{x} = \int_\Omega f_i(\mathbf{x})d\mathbf{x}$ for each $i \in \mathscr{V}$. Then the result follows by extending the function u_e to the domain $[0,t_f]$ by defining $u_e(t) = 0$ for $t \in (t_f/2,t_f]$ and by defining $\mathbf{v}_i(\cdot,t)$ for $t \in (t_f/2,t_f]$ as in the proof of Theorem 3.2.16. □

3.4 Stabilization of a System of Advection-Diffusion-Reaction Equations to Target Probability Densities

In this section we will consider the following problem of stabilizing a target stationary distribution $\mathbf{f}^{eq}$ of the process (3.31) using time-independent control laws, which are more practical for implementation than time-dependent control laws.

Problem 3.4.1. *Given $\mathbf{f}^{eq} : \Omega^N \to \mathbb{R}_+$, determine whether there exist time-independent and possibly spatially-dependent parameters $\mathbf{v}_k : \Omega \to \mathbb{R}^n$, $K_e : \Omega \to \mathbb{R}_+$ such that the solution of the system*

$$\begin{aligned}
(y_k)_t &= D_k \Delta y_k - \nabla \cdot (\mathbf{v}_k(\mathbf{x}) y_k) + \mathscr{F}_k \quad && in \;\; \Omega \times [0,\infty) \\
y_k(\cdot,0) &= y_k^0 && in \;\; \Omega \\
\mathbf{n} \cdot (\nabla y_k - \mathbf{v}_k(\mathbf{x}) y_k) &= 0 && in \;\; \partial\Omega \times [0,\infty),
\end{aligned} \tag{3.33}$$

where $k \in \mathscr{V}$ and $\mathscr{F}_k = \sum_{e \in \mathscr{E}} \sum_{j \in \mathscr{V}} K_e(\mathbf{x}) Q_e^{kj} y_j$, satisfies $\lim_{t\to\infty} y_k(\cdot,t) \to f_k$ for each $k \in \mathscr{V}$.

Before addressing this problem, we first briefly review the notion of *irreducibility* of a *positive operator* (Meyer-Nieberg, 2012), which will be used extensively in the theorems in this section. Let $\mathscr{P}$ be a positive operator on the Hilbert space $X = L_a^2(\Omega)$ (or $\mathbf{L}_\mathbf{a}^2(\Omega)$)

for some $a \in L^\infty(\Omega)$ (or $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$), i.e., a linear bounded operator that maps real-valued non-negative elements of X to real-valued non-negative elements of X. Let $\tilde{\Omega} \subset \Omega$ (or $\tilde{\Omega} \subset \Omega^N$) be a measurable subset. Consider the set $\mathscr{I}_{\tilde{\Omega}}$ defined by $\mathscr{I}_{\tilde{\Omega}} = \left\{ f \in X : \tilde{\Omega} \subset \{\mathbf{x} \in \Omega : f(\mathbf{x}) = 0\} \right\}$. $\mathscr{P}$ will be called *irreducible* if the only measurable sets $\tilde{\Omega} \subset \Omega$ for which the set $\mathscr{I}_{\tilde{\Omega}}$ is invariant under $\mathscr{P}$ are $\tilde{\Omega} = \Omega$ (or Ω^N) and $\tilde{\Omega} = \varnothing$, the null set. A semigroup of operators $(\mathscr{T}(t))_{t \geq 0}$ on X will be called irreducible if $\mathscr{T}(t)$ is an irreducible operator for every $t > 0$. Suppose that A is the generator of the semigroup $(\mathscr{T}(t))_{t \geq 0}$ and $s(A) := \sup\{\mathrm{Re}(\lambda) : \lambda \in \mathrm{spec}(A)\}$. Then $(\mathscr{T}(t))_{t \geq 0}$ being irreducible is equivalent to $(\lambda \mathbb{I} - A)^{-1}$ mapping real-valued non-negative elements of X to strictly positive elements of X for every $\lambda > s(A)$ (Arendt *et al.*, 2006)[Definition C-III.3.1]. Note that the definitions in the cited reference are stated in a general framework of *Banach lattices*, for which $(\mathscr{T}(t))_{t \geq 0}$ being irreducible is equivalent to $(\lambda \mathbb{I} - A)^{-1}$ mapping positive elements of X to *quasi-interior* elements of X. However, for the spaces that we consider, quasi-interior elements are the same as functions that are positive almost everywhere on their domain of definition.

Theorem 3.4.2. *Let $\{q_e\}_{e \in \mathscr{E}}$ be a set of non-negative constants. Then* $\mathrm{spec}(\sum_{e \in \mathscr{E}} q_e \mathbf{Q}_e) \subset \bar{\mathbb{C}}_-$.

Proof. This follows from (Minc, 1988)[Theorem II.1.1] by noting that all the elements of the matrix $\mathbf{G}_\lambda = \lambda \mathbb{I} + \sum_{e \in \mathscr{E}} q_e \mathbf{Q}_e$ are non-negative for $\lambda > 0$ large enough. □

Theorem 3.4.3. *Let $\{q_e\}_{e \in \mathscr{E}}$ be a set of non-negative constants. Let $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$ such that $\mathbf{a} \geq c$ for some positive constant c. Suppose $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ such that $b_i = D_i \mathbf{1}$ is a positive constant function for each $i \in \mathscr{V}$. Then* $\mathrm{spec}(-\mathscr{M}_{\mathbf{b}} \mathscr{B}_{\mathbf{a}} + \sum_{e \in \mathscr{E}} q_e a_{S(e)} \mathscr{Q}_e) \subset \bar{\mathbb{C}}_-$.

Proof. Let $\mathscr{W} = -\mathscr{M}_{\mathbf{b}} \mathscr{B}_{\mathbf{a}} + \sum_{e \in \mathscr{E}} q_e a_{S(e)} \mathscr{Q}_e$. Let $\lambda \in \mathbb{C} \backslash \mathrm{spec}(-\mathscr{M}_{\mathbf{b}} \mathscr{B}_{\mathbf{a}} + \sum_{e \in \mathscr{E}} q_e a_{S(e)} \mathscr{Q}_e)$ be real and large enough such that $(\lambda \mathbb{I} - \mathscr{W})^{-1}$ is a positive operator, i.e., $(\lambda \mathbb{I} - \mathscr{W})^{-1} f \geq 0$ whenever $f \geq 0$. Such a λ necessarily exists because the semigroup $(\mathscr{U}(t))_{t \geq 0}$ generated by the operator $\mathscr{W}$ is positivity preserving from Lemma 3.3.3. Hence, the existence of

λ follows from the resolvent formula $(\lambda\mathbb{I}-\mathscr{W})^{-1}=\int_0^\infty e^{-\lambda t}\mathscr{U}(t)dt$ when λ is greater than the growth bound of the semigroup $(\mathscr{U}(t))_{t\geq 0}$, which is equal to the spectral growth bound, $s(\mathscr{W}):=\{\text{Re }\mu:\mu\in\text{spec}(\mathscr{W})\}$, of $\mathscr{W}$ since $(\mathscr{U}(t))_{t\geq 0}$ is analytic (Engel and Nagel, 2000)[Theorem II.1.10]. Let $R_\lambda=(\lambda\mathbb{I}-\mathscr{W})^{-1}$. The operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}}$ has a compact resolvent since $H^1_{a_i}(\Omega)$ is compactly embedded in $L^2_{a_i}(\Omega)$ for each $i\in\mathscr{V}$ (Schmüdgen, 2012)[Proposition 10.6]. The operator R_λ is compact and positivity preserving since $\mathscr{W}$ is a bounded perturbation of $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}}$. Additionally, the spectral radius of R_λ is positive since 0 is an eigenvalue of $\mathscr{W}$ (and hence $\frac{1}{\lambda}$ is an eigenvalue of R_λ). Therefore, from the *Krein-Rutman theorem* (Meyer-Nieberg, 2012)[Theorem 4.1.4], it follows that if r is the spectral radius of the operator R_λ, then there exists a positive nonzero element $\mathbf{h}\in\mathbf{L}^2_{\mathbf{a}}(\Omega)$ such that $r\mathbf{h}-R_\lambda\mathbf{h}=\mathbf{0}$. Then it follows that $\mathbf{h}\in\mathscr{D}(\mathscr{W})$ and $(\lambda-\frac{1}{r})\mathbf{h}-\mathscr{W}\mathbf{h}=\mathbf{0}$. For the sake of contradiction, suppose that $\lambda>\frac{1}{r}$. Then we have that

$$\alpha\int_\Omega h_i(\mathbf{x})d\mathbf{x}+\int_\Omega D_i(B_{a_i}h_i)(\mathbf{x})d\mathbf{x}-\sum_{e\in\mathscr{E}}\sum_{j=1}^N\int_\Omega q_e a_{S(e)}(\mathbf{x})Q_e^{ij}h_j(\mathbf{x})d\mathbf{x}=0$$

for each $i\in\mathscr{V}$, where $\alpha=\lambda-\frac{1}{r}$. This implies that

$$\alpha\int_\Omega h_i(\mathbf{x})d\mathbf{x}-\sum_{e\in\mathscr{E}}\sum_{j=1}^N\int_\Omega q_e a_{S(e)}(\mathbf{x})Q_e^{ij}h_j(\mathbf{x})d\mathbf{x}=0$$

for each $i\in\mathscr{V}$. But this implies that the matrix $\sum_{e\in\mathscr{E}}q_e k_{S(e)}\mathbf{Q}_e$, where the constants $\{k_i\}_{i\in\mathscr{V}}$ are such that

$$\int_\Omega a_i(\mathbf{x})h_i(\mathbf{x})d\mathbf{x}=k_i\int_\Omega h_i(\mathbf{x})d\mathbf{x},$$

has a positive eigenvalue α. This contradicts Theorem 3.4.2, since $\text{spec}(\sum_{e\in\mathscr{E}}q_e k_{S(e)}\mathbf{Q}_e)\subset\bar{\mathbb{C}}_-$. Hence, we cannot have any eigenvalues of $\mathscr{W}$ on the positive real axis. Since $\mathscr{W}$ generates a positive semigroup, and its spectrum is non-void, we know that the spectral growth $s(\mathscr{W})$ lies in the spectrum of $\mathscr{W}$ (Arendt *et al.*, 2006)[Theorem 1.1]. Hence, we can conclude that the spectrum of $\mathscr{W}$ lies in $\bar{\mathbb{C}}_-$. This concludes the proof. □

Proposition 3.4.4. *Let $\mathscr{G}$ be strongly connected. Let $\mathbf{f} \in \mathbf{L}^\infty(\Omega)$ be such that $\mathbf{f} \geq c$ for some positive constant c. Let $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ such that $b_i = D_i\mathbf{1}$ is a positive constant function for each $i \in \mathscr{V}$. Suppose $\mathbf{y}^0 \in \mathbf{L}^2(\Omega)$ such that $\mathbf{y}^0 \geq 0$ and $\sum_i \int_\Omega f_i(\mathbf{x})d\mathbf{x} = \sum_i \int_\Omega y_i^0(\mathbf{x})d\mathbf{x} = 1$. Let $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$ be such that $a_i = 1/f_i$ for each $i \in \mathscr{V}$. Then there exist positive parameters $\{q_e\}_{e\in\mathscr{E}}$ such that, if $(\mathscr{F}(t))_{t\geq 0}$ is the semigroup generated by the operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \sum_{e\in\mathscr{E}} q_e a_{S(e)}\mathscr{Q}_e$, then we have*

$$\|\mathscr{F}(t)\mathbf{y}^0 - \mathbf{f}\|_2 \;\leq\; Me^{-\lambda t} \tag{3.34}$$

for some positive constants M and λ and all $t \geq 0$.

Proof. Since the graph $\mathscr{G}$ is assumed to be strongly connected, from Proposition 2.3.4 we know that there exist positive parameters $\{p_e\}_{e\in\mathscr{E}}$ such that, if $u_e(t) = p_e$ for all $e \in \mathscr{E}$ and all $t \geq 0$, then the solution $\mu(t)$ of system (2.2) satisfies

$$\|\mu(t) - \mu^{eq}\|_2 \;\leq\; M_1 e^{-\lambda_1 t} \tag{3.35}$$

for some positive constants M_1 and λ_1 and all $t \geq 0$, where $\mu_k^{eq} = \int_\Omega f_k(\mathbf{x})d\mathbf{x}$ for each $k \in \mathscr{V}$ and $\mu^0 \in \mathscr{P}(\mathscr{V})$. In particular, 0 is a simple eigenvalue of the irreducible operator $\sum_{e\in\mathscr{E}} p_e\mathbf{Q}_e$ and μ^{eq} is the corresponding unique (up to a scalar multiple) and strictly positive eigenvector. Then 0 is an eigenvalue for the operator $\mathscr{W} = -\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \sum_e p_e a_{S(e)}\mathscr{Q}_e$ with the corresponding eigenvector $\mathbf{f}$, by construction. We will show that this eigenvalue is simple and is the dominant eigenvalue. Let $\mathbf{g} \in \mathbf{L}^2_{\mathbf{a}}(\Omega)$ such that $\mathbf{g}$ is not the zero element $\mathbf{0}$ and is non-negative a.e. in Ω^N. Defining $\mathbf{h} = (\lambda\mathbb{I} - \mathscr{W})^{-1}\mathbf{g}$ for some $\lambda > 0$ that is large enough, we have that

$$\lambda\int_\Omega h_i(\mathbf{x})d\mathbf{x} + \int_\Omega D_i(B_{a_i}h_i)(\mathbf{x})d\mathbf{x} - \sum_{e\in\mathscr{E}}\sum_{j=1}^N \int_\Omega p_e a_{S(e)}(\mathbf{x})Q_e^{ij}h_j(\mathbf{x})d\mathbf{x} = \int_\Omega g_i(\mathbf{x})d\mathbf{x}$$

for each $i \in \mathscr{V}$. This implies that

$$\lambda\int_\Omega h_i(\mathbf{x})d\mathbf{x} - \sum_{e\in\mathscr{E}}\sum_{j=1}^N \int_\Omega p_e a_{S(e)}(\mathbf{x})Q_e^{ij}h_j(\mathbf{x})d\mathbf{x} = \int_\Omega g_i(\mathbf{x})d\mathbf{x}$$

for each $i \in \mathscr{V}$, which implies that

$$\lambda \int_\Omega h_i(\mathbf{x})d\mathbf{x} - \sum_{e\in\mathscr{E}}\sum_{j=1}^{N}\int_\Omega p_e k_{S(e)} Q_e^{ij} h_j(\mathbf{x})d\mathbf{x} = \int_\Omega g_i(\mathbf{x})d\mathbf{x} \tag{3.36}$$

for each $i \in \mathscr{V}$ for some positive constants $k_i > 0$. The existence of such positive constants is guaranteed, since we assumed that $\mathbf{g}$ is non-negative and hence $\mathbf{h}$ is non-negative. However, $\sum_e p_e k_{S(e)} \mathbf{Q}_e$ generates an irreducible semigroup on $\mathbb{R}^N$ whenever $p_e > 0$ implies that $k_e > 0$ for all $e \in \mathscr{E}$. Hence, $(\lambda\mathbb{I} - \sum_e p_e k_{S(e)}\mathbf{Q}_e)^{-1}$ maps non-negative, nonzero elements of $\mathbb{R}^N$ to strictly positive elements of $\mathbb{R}^N$. This implies that $\int_\Omega h_i(\mathbf{x})d\mathbf{x} > 0$ for each $i \in \mathscr{V}$. From this, we can conclude that $h_i(\mathbf{x}) > 0$ for a.e. $\mathbf{x} \in \Omega$ for each $i \in \mathscr{V}$. To see this more explicitly, note that $\mathbf{h}$ must satisfy

$$\lambda h_i + D_i B_{a_i} h_i - G^{ii} a_i h_i = g_i + \sum_{j=1, j\neq i}^{N} G^{ij} a_j h_j \tag{3.37}$$

for each $i \in \mathscr{V}$, where $\mathbf{G} = \sum_{e\in\mathscr{E}} p_e \mathbf{Q}_e$. Let $\mathscr{M}_{a_i}$ be the multiplication operator, defined on $L^2(\Omega) = L^2_{a_i}(\Omega)$, that is associated with the function a_i. The spectrum of the operator $-D_i B_{a_i}$ lies in $\bar{\mathbb{C}}_-$. Consequently, since $a_i \geq \ell$ for some $\ell > 0$, so does the spectrum of the operator $-D_i B_{a_i} - \lambda\mathscr{M}_{a_i}$. Moreover, G_{ii} is negative. Hence, the inverse $R^i_\lambda = (\lambda\mathbb{I} + D_i B_{a_i} - G^{ii}\mathscr{M}_{a_i})^{-1} = (\lambda\mathscr{M}_{a_i}^{-1} + D_i B_{a_i}\mathscr{M}_{a_i}^{-1} - G^{ii}\mathbb{I})^{-1}\mathscr{M}_{a_i}^{-1}$ exists. The operator $-\lambda\mathscr{M}_{a_i}^{-1} - D_i B_{a_i}\mathscr{M}_{a_i}^{-1}$ generates an irreducible semigroup on $L^2(\Omega)$ (Ouhabaz, 2009)[Theorem 4.5] (see equation (4.8) in the cited reference for the class of operators considered); formally, $-B_{a_i}\mathscr{M}_{a_i}^{-1}$ is the operator $\nabla\cdot(\frac{1}{a_i}\nabla(\cdot))$. Hence, $(R^i_\lambda[g_i + \sum_{j=1,j\neq i}^N G^{ij}a_j h_j])(\mathbf{x})$ is strictly positive for a.e. $\mathbf{x} \in \Omega$ and each $i \in \mathscr{V}$, since $\sum_{j=1,j\neq i}^N G^{ij}$ and h_i are nonzero for each $i \in \mathscr{V}$. Therefore, $(\lambda\mathbb{I} - \mathscr{W})^{-1}$ maps nonzero, non-negative elements of $\mathbf{L}^2_{\mathbf{a}}(\Omega)$ to strictly positive elements of $\mathbf{L}^2_{\mathbf{a}}(\Omega)$. This implies that the semigroup generated by the operator $\mathscr{W}$ is irreducible. Now, we can use (Arendt *et al.*, 2006)[Corollary C-III.3.17] to establish that the eigenvalue 0 is simple and is the dominant eigenvalue. This follows from the cited corollary because $\mathscr{W}$ has a compact resolvent and generates an analytic semigroup, due to

the fact that it is a bounded perturbation of the operator $-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}}$, which itself has a compact resolvent and generates an analytic semigroup (Engel and Nagel, 2000)[Proposition III.1.12]. Additionally, we know from (Engel and Nagel, 2000)[Corollary III.1.19] that since $\mathscr{W}$ has a compact resolvent, its spectrum is discrete. Then the result follows from (Engel and Nagel, 2000)[Corollary V.3.3]. □

Irreducibility is not necessary, but only sufficient, for the simplicity of the dominant eigenvalue of a compact positive operator. The goal of the following proposition and theorem is to extend the result in Proposition 3.4.4 to a much larger set of equilibrium distributions, for which the resulting semigroup is not necessarily irreducible.

Proposition 3.4.5. *Let* $\mathbf{P} \in \mathbb{R}^{N\times N}$ *be essentially non-negative, i.e.,* $P^{ij} \geq 0$ *for all* $i \neq j$ *in* $\mathscr{V}$. *Let* $\mathscr{P}$ *be the linear bounded operator on* $\mathbf{L}^2(\Omega)$, *defined pointwise using* $\mathbf{P}$ *as* $(\mathscr{P}\mathbf{h})(\mathbf{x}) = \mathbf{P}\mathbf{h}(\mathbf{x})$ *for a.e.* $\mathbf{x} \in \Omega$ *for all* $\mathbf{h} \in \mathbf{L}^2(\Omega)$. *Suppose* $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ *such that* $b_i = D_i\mathbf{1}$ *is a positive constant function for each* $i \in \mathscr{V}$. *In addition, suppose that* $\mathrm{spec}(\mathbf{P})$ *lies in* $\mathbb{C}_-$. *If* $a_i = \mathbf{1}$ *for each* $i \in \mathscr{V}$, *then* $\mathrm{spec}(-\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \mathscr{P})$ *lies in* $\mathbb{C}_-$.

Proof. The proof follows the same line of argument as Theorem 3.4.3. Note that according to the Lie-Trotter product formula, $\mathscr{W} = -\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \mathscr{P}$ generates a positive semigroup since both $-\mathscr{B}_{\mathbf{a}}$ and $\mathscr{P}$ generate positivity preserving semigroups. Hence, if $\lambda > 0$ is large enough, then $R_\lambda = (\lambda - \mathscr{W})^{-1}$ is a positive operator. Moreover, R_λ is a compact operator and has a nonzero spectral radius r. From the Krein-Rutman theorem (Meyer-Nieberg, 2012)[Theorem 4.1.4], it follows that there exists a positive function $\mathbf{h} \in \mathbf{L}^2_{\mathbf{a}}(\Omega) = \mathbf{L}^2(\Omega)$ such that $r\mathbf{h} - R_\lambda\mathbf{h} = \mathbf{0}$. This implies that $\lambda - \frac{1}{r}$ is an eigenvalue of $\mathscr{W}$. However, this implies that $\int_\Omega -(B_{a_i}h_i) = 0$ for each $i \in \mathscr{V}$, and hence that

$$\left(\lambda - \frac{1}{r}\right)\int_\Omega h_i(\mathbf{x})d\mathbf{x} - \sum_{j=1}^{N}\int_\Omega P^{ij}h_j(\mathbf{x})d\mathbf{x} \;=\; 0$$

for each $i \in \mathscr{V}$. If $\lambda - \frac{1}{r} \geq 0$, then we arrive at a contradiction, since $\mathrm{spec}(\mathbf{P})$ lies in $\mathbb{C}_-$. Here, we have used the fact that $\mathbf{h}$ is a positive function, and therefore $\int_\Omega h_i(\mathbf{x})d\mathbf{x}$ cannot

be equal to 0 for each $i \in \mathscr{V}$. Hence, we cannot have any eigenvalues of $\mathscr{W}$ on the non-negative real axis of the complex plane. Since $\mathscr{W}$ generates a positive semigroup, and its spectrum is non-void, we know that the spectral growth $s(\mathscr{W})$ lies in the spectrum of $\mathscr{W}$ (Arendt *et al.*, 2006)[Theorem 1.1]. Hence, we can conclude that the spectrum of $\mathscr{W}$ lies in $\mathbb{C}_-$. This concludes the proof. □

Theorem 3.4.6. *Let $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ be strongly connected, and let Ω be an extension domain. Let $\mathbf{b} \in \mathbf{L}^\infty(\Omega)$ such that $b_i = D_i\mathbf{1}$ is a positive constant function for each $i \in \mathscr{V}$. Let $\mathbf{f} \in \mathbf{L}^\infty(\Omega)$ be non-negative such that $f_i \geq c\int_\Omega f_i(\mathbf{x})d\mathbf{x}$ for some positive constant $c > 0$. Let $\mathscr{V}_1 = \{i \in \mathscr{V} : \int_\Omega f_i(\mathbf{x})d\mathbf{x} > 0\}$. Additionally, consider the set $\mathscr{E}_1 = \{e \in \mathscr{E} \ : \ S(e), T(e) \in \mathscr{V}_1\}$. Suppose that the graph $\mathscr{G}_1 = (\mathscr{V}_1, \mathscr{E}_1)$ is strongly connected. Then there exist $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$ and spatially-dependent reaction coefficients $\{K_e(\mathbf{x})\}_{e\in\mathscr{E}} \in \mathbf{L}^\infty(\Omega)$ for which $-\mathscr{M}_\mathbf{b}\mathscr{B}_\mathbf{a} + \sum_{e\in\mathscr{E}} K_e\mathscr{Q}_e$ generates a positive semigroup $(\mathscr{S}(t))_{t\geq 0}$ on $\mathbf{L}^2_\mathbf{a}(\Omega)$ such that if $\mathbf{y}^0 \in \mathbf{L}^2_\mathbf{a}(\Omega)$ is a positive function and $\sum_{i\in\mathscr{V}}\int_\Omega f_i(\mathbf{x})d\mathbf{x} = \sum_{i\in\mathscr{V}}\int_\Omega y_i(\mathbf{x})d\mathbf{x}$, then*

$$\|\mathscr{S}(t)\mathbf{y}_0 - \mathbf{f}\| \ \leq \ Me^{-\lambda t} \tag{3.38}$$

for some positive constants M and λ and all $t \geq 0$.

Proof. Without loss of generality, we assume that the set $\mathscr{V}_1$ is of the form $\mathscr{V}_1 = \{1, 2, ..., \bar{N}\}$ for some integer $\bar{N} \leq N$. Define $\mu^{eq} \in \mathbb{R}^N_+$ such that $\mu^{eq}_i = \int_\Omega f_i(\mathbf{x})d\mathbf{x}$ for each $i \in \mathscr{V}$. Then from Proposition 2.3.4, it follows that there exist positive constants $\{q_e\}_{e\in\mathscr{E}}$ such that the solution $\mu(t)$ of the ODE system (2.2) converges exponentially to μ^{eq}. In particular, the matrix $\sum_{e\in\mathscr{E}} q_e\mathbf{Q}_e$ has 0 as a simple eigenvalue with μ^{eq} as the corresponding eigenvector, which is unique up to a scalar multiple. Let $\mathbf{G} = \sum_{e\in\mathscr{E}}\mathbf{Q}_e$. Then $\mathbf{G}$ is necessarily of the form

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 & \mathbf{G}_2 \\ \mathbf{0} & \mathbf{G}_3 \end{bmatrix}, \tag{3.39}$$

where $\mathbf{G}_1 \in \mathbb{R}^{\bar{N}\times\bar{N}}$, $\mathbf{G}_2 \in \mathbb{R}^{\bar{N}\times(N-\bar{N})}$, $\mathbf{G}_3 \in \mathbb{R}^{(N-\bar{N})\times(N-\bar{N})}$, and $\mathbf{0}$ is the zero element of $\mathbb{R}^{(N-\bar{N})\times\bar{N}}$. If $\mathbf{G}$ does not have the block triangular structure above, then there exist indices $i \in \mathscr{V}_1$ and $j \in \mathscr{V}\backslash\mathscr{V}_1$ such that $G^{ji} > 0$. But this implies that if $\mu^0 = \mu^{eq}$, then $\dot{\mu}_j(0) \neq 0$ for all $j \in \mathscr{V}$, hence contradicting that $\mathbf{G}\mu^{eq}$ is the zero element of $\mathbb{R}^N$. Moreover, since $\lim_{t\to\infty}\mu_j(t) = 0$ for all $j \in \mathscr{V}\backslash\mathscr{V}_1$ for any $\mu^0 \in \mathbb{R}^N$, we must have that $\text{spec}(\mathbf{G}_3)$ is in $\mathbb{C}_-$ and that 0 is a simple eigenvalue of $\mathbf{G}_1$. Now, let $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$ be such that $a_i = 1/f_i$ if $i \in \mathscr{V}_1$ and $a_i = k_i\mathbf{1}$ if $i \in \mathscr{V}\backslash\mathscr{V}_1$ for some positive constant k_i. Then consider the operator $\mathscr{W} = -\mathscr{M}_{\mathbf{b}}\mathscr{B}_{\mathbf{a}} + \sum_{e\in\mathscr{E}} q_e a_S(e)\mathscr{Q}_e$. This operator is of the form

$$\mathscr{W} = \begin{bmatrix} \mathscr{W}_1 & \mathscr{W}_2 \\ \mathbf{0} & \mathscr{W}_3 \end{bmatrix}, \tag{3.40}$$

where $\mathscr{W}_1 \in \mathscr{L}(X_1,X_1)$, $\mathscr{W}_2 \in \mathscr{L}(X_1,X_2)$, $\mathscr{W}_3 \in \mathscr{L}(X_2,X_2)$, and $\mathbf{0}$ is the zero element of $\mathscr{L}(X_2,X_1)$, with $X_1 = L^2_{a_1} \times ... \times L^2_{a_{\bar{N}}}$ and $X_2 = L^2_{a_{\bar{N}+1}} \times ... \times L^2_{a_N}$. From Proposition 3.4.5, it follows that $\text{spec}(\mathscr{W}_3)$ lies in $\mathbb{C}_-$. Moreover, from Theorem 3.4.2, it follows that 0 is a simple and dominant eigenvalue of $\mathscr{W}_1$ with the corresponding eigenvector $[f_1 \ ... \ f_{\bar{N}}]^T$. Then the result follows from (Engel and Nagel, 2000)[Corollary V.3.3]. □

3.5 Weighted Hypoelliptic Laplacians and their Semigroups

In this section, we generalize some of the semigroup generation results of Section 3.2 to a class of degenerate operators that are not necessarily elliptic. This generalization is relevant to applications in swarm robotics in which each agent of the swarm has non-holonomic constraints (Bloch, 2015) on its dynamics.

Before we proceed to state the results of this section, we define some new notation and motivation for the results. We refer the reader to (Lee, 2001) for the differential geometric terminologies used in this section.

In this section, Ω will denote an open, bounded, and connected subset with a smooth boundary of an $N-$dimensional simply connected Lie group G. The boundary of Ω is

denoted by $\partial\Omega$. In addition, $\int_\Omega f(\mathbf{x})d\mathbf{x}$ will denote the integral of a function $f:\Omega\to\mathbb{R}$ with respect to the Haar measure (Diestel and Spalsbury, 2014). We recall that when $G=\mathbb{R}^N$ with the standard group structure on $\mathbb{R}^N$, the Haar measure coincides with the Lebsegue measure.

Suppose that $e^{\mathbf{X}t}$ is the flow generated by a vector field $\mathbf{X}$. Then $\mathbf{X}$ defines a differential operator on the set of smooth functions $C^\infty(G)$ through the action

$$(\mathbf{X}f)(\mathbf{x})=\lim_{t\to 0}\frac{f(e^{t\mathbf{X}}(\mathbf{x}))-f(\mathbf{x})}{t} \tag{3.41}$$

for all $\mathbf{x}\in\Omega$.

Note that here we are using the differential geometric definition (Lee, 2001) of a vector field $\mathbf{X}$ as an associated differential operator acting on the space of smooth functions through the definition (3.41).

Let $\mathscr{X}=\{\mathbf{X}_1,...,\mathbf{X}_m\}$ be a collection of *left-invariant vector fields* (Lee, 2001) with $m\leq N$. We will assume that the collection of vector fields $\mathscr{X}$ satisfies the Lie Rank condition or the *Hormander condtion*, i.e., the Lie algebra spanned by the vector fields $\mathscr{X}$ has rank N. Given $a\in L^\infty(\Omega)$, with $a\geq c$ for a positive parameter $c>0$, we define the Horizontal Sobolev space $WH_a^1(\Omega)=\left\{f\in L^2(\Omega):\mathbf{X}_i(af)\in L^2(\Omega)\text{ for }1\leq i\leq m\right\}$. We equip this space with the Horizontal Sobolev norm $\|\cdot\|_{WH^1}$, given by $\|f\|_{WH_a^1}=\left(\|f\|_2^2+\sum_{i=1}^n\|\mathbf{X}_i(af)\|_2^2\right)^{1/2}$ for each $f\in WH_a^1(\Omega)$. Here, the derivative action of $\mathbf{X}_i$ on a function f is to be understood in the distributional sense.

A *horizontal curve* $\gamma:[0,1]\to\Omega$ connecting two points $\mathbf{x},\mathbf{y}\in\Omega$ is a Lipschitz curve in Ω such that there exist essentially bounded functions $a_i(t)$ such that

$$\dot{\gamma}(t)=\sum_{i=1}^m a_i(t)\mathbf{X}(\gamma(t)) \tag{3.42}$$

for almost every $t\in[0,1]$, such that $\gamma(0)=\mathbf{x}$ and $\gamma(1)=\mathbf{y}$. Then $\mathscr{X}$ defines a distance $d:\Omega\to\mathbb{R}_{\geq 0}$ on Ω given by

$$d(\mathbf{x},\mathbf{y})=\inf\{\int_0^1|\dot{\gamma}(t)|dt;\gamma\text{ is a horizontal curve connecting }\mathbf{x}\text{ and }\mathbf{y}\} \tag{3.43}$$

The metric d on Ω is known as the *Carnot-Caratheodory* or *Sub-Riemannian metric* (Bramanti, 2014). The topology induced by this metric on d coincides with the usual bi-invariant metric on G (Nhieu, 2001), which is the standard Euclidean metric when $G = \mathbb{R}^N$. We will assume that the radius $r(\Omega)$ of Ω, given by $r(\Omega) = \sup\{d(\mathbf{x},\mathbf{y}); \mathbf{x},\mathbf{y} \in G\}$, is finite.

Consider the following SDE,

$$\begin{aligned} d\mathbf{Z}(t) &= \sum_{i=1}^{m} u_i(\mathbf{Z}(t),t)\mathbf{X}_i dt + \sum_{i=1}^{m} \mathbf{X}_i \circ dW + \mathbf{n}(\mathbf{Z}(t))d\psi(t), \\ \mathbf{Z}(0) &= \mathbf{Z}_0, \end{aligned} \tag{3.44}$$

In the above SDE (3.44), the notation $\circ$ is used to mean that the SDE should be interpreted in the *sense of Stratonovich* (Karatzas and Shreve, 1998). We define $\Delta_H := \sum_{i=1}^{m} \mathbf{X}_i^2$ and will refer to this operator as the *Horizontal Laplacian* operator. Let ∇_H denote the horizontal gradient, defined by

$$\nabla_H(f) = \sum_{i=1}^{m} \mathbf{X}_i(f)\mathbf{X}_i \tag{3.45}$$

for all $f \in C^\infty(G)$. The associated probability density y of the random variable $\mathbf{Z}(t)$ evolves according to the PDE

$$\begin{aligned} y_t = \Delta_H y - \nabla_w \cdot (\textstyle\sum_{i=1}^{m} u_i(\mathbf{x},t)\mathbf{X}_i y) \quad & in \;\; \Omega \times [0,T] \\ \mathbf{n} \cdot (\nabla_H y - \textstyle\sum_{i=1}^{m} u_i(\mathbf{x},t)\mathbf{X}_i y) = 0 \quad & in \;\; \partial\Omega \times [0,T] \\ y(\cdot,0) = y^0 \quad & in \;\; \Omega, \end{aligned} \tag{3.46}$$

where $\nabla_w \cdot$ denotes the divergence operation with respect to the Haar measure that maps vector fields to functions.

The operator Δ_H is not elliptic in general, but only *hypoelliptic*. Particularly, if $f \in C_0^\infty(\Omega)$ has compact support K, then, due to the Lie Rank condition, if u is a function on Ω such that $\Delta_H u = f$, then u is smooth on K (Bramanti, 2014).

Let $f \in W^{1,\infty}(\Omega)$ be a positive function that is bounded from below by a positive number and for which $\int_\Omega f(\mathbf{x})d\mathbf{x} = 1$. If we set $u_i(\cdot,t) = \mathbf{X}_i(g)/g$ for each $i \in \{1,...,m\}$ and all $t \geq 0$,

then the PDE (3.46) becomes

$$\begin{aligned} y_t &= \Delta_H y - \nabla_w \cdot (\textstyle\sum_{i=1}^m \frac{\mathbf{X}_i(g)}{g} \mathbf{X}_i y) \quad \textit{in} \;\; \Omega \times [0,T] \\ \mathbf{n} \cdot (\nabla_H y - \textstyle\sum_{i=1}^m \frac{\mathbf{X}_i(g)}{g} \mathbf{X}_i y) &= 0 \quad \textit{in} \;\; \partial\Omega \times [0,T] \\ y(\cdot,0) &= y^0 \quad \textit{in} \;\; \Omega \,. \end{aligned} \tag{3.47}$$

When the Lie group G is *unimodular*, i.e., the left- and right-Haar measure coincide, $\nabla_w \cdot \nabla_H(\cdot) = \Delta_H(\cdot)$ (Agrachev *et al.*, 2009). Hence, if we set $y = g$, then

$$\Delta_H y - \nabla_w \cdot (\sum_{i=1}^m \frac{\mathbf{X}_i(g)}{g} \mathbf{X}_i y) = \Delta_H g - \nabla_w \cdot (\nabla_H g) = 0 \tag{3.48}$$

Thus, g is an equilibrium solution of the PDE (3.47). We can further show that g is the globally exponentially stable equilibrium solution of PDE (3.47) on the the set of square-integrable probability densities. Thus, if a swarm of robots is modeled according to the SDE (3.44), the state-feedback law $u_i(\cdot,t) = \mathbf{X}_i(g)/g$ can be used to stabilize the swarm to the target density g. This motivates us to study semigroup generation properties of the operator $\nabla_w \cdot (\frac{1}{a(\mathbf{x})} \nabla_H (a(\mathbf{x}) \cdot))$. Similarly, the operator $\Delta_H(a\cdot)$ can also be associated with a stochastic process on G whose probability density converges to $1/a$. Hence, we will also establish similar semigroup generation results for the operator $\Delta_H(a\cdot)$.

While there have been works on semigroups generated by hypoelliptic operators on manifolds without boundary (Jerison and Sánchez-Calle, 1986), or manifolds with boundary under the Dirichlet boundary (Varopoulos *et al.*, 2008; Robinson, 1991), there seems to be, to our knowledge, no existing work on semigroups generated by hypoelliptic operators with Neumann boundary condition such the one considered in (3.47).

Given $a \in L^\infty(\Omega)$ such that $a \geq c$ for some positive constant c, and $\mathscr{D}(\omega_a^H) = WH_a^1(\Omega)$, we define the sesquilinear form $\omega_a^H : \mathscr{D}(\omega_a^H) \times \mathscr{D}(\omega_a^H) \to \mathbb{C}$ as

$$\omega_a^H(u,v) = \sum_{i=1}^m \int_\Omega X_i(a(\mathbf{x})u(\mathbf{x})) \cdot X_i(a(\mathbf{x})\bar{v}(\mathbf{x})) d\mathbf{x} \tag{3.49}$$

for each $u \in \mathscr{D}(\omega_a)$. We associate with the form ω_a an operator $A_a^H : \mathscr{D}(A_a^H) \to L_a^2(\Omega)$, defined as $A_a^H u = v$ if $\omega_a^H(u,\phi) = \langle v, \phi \rangle_a$ for all $\phi \in \mathscr{D}(\omega_a^H)$ and for all $u \in \mathscr{D}(A_a^H) = \{g \in \mathscr{D}(\omega_a^H) : \exists f \in L_a^2(\Omega) \text{ s.t. } \omega_a^H(g,\phi) = \langle f, \phi \rangle_a \ \forall \phi \in \mathscr{D}(\omega_a^H)\}$.

Similarly, given $a \in L^\infty(\Omega)$ such that $a \geq c$ for some positive constant c and $\mathscr{D}(\sigma_a^H) = WH_a^1(\Omega)$, we define the sesquilinear form $\sigma_a^H : \mathscr{D}(\sigma_a^H) \times \mathscr{D}(\sigma_a^H) \to \mathbb{C}$ as

$$\sigma_a^H(u,v) = \sum_{i=1}^{m} \int_\Omega 1/(a(\mathbf{x})) X_i(a(\mathbf{x})u(\mathbf{x})) \cdot X_i(a(\mathbf{x})\bar{v}(\mathbf{x})) d\mathbf{x} \tag{3.50}$$

for each $u \in \mathscr{D}(\sigma_a^H) = WH_a^1(\Omega)$. As for the form ω_a^H, we associate an operator $B_a^H : \mathscr{D}(B_a^H) \to L_a^2(\Omega)$ with the form σ_a^H. We define this operator as $B_a u = v$ if $\sigma_a^H(u,\phi) = \langle v, \phi \rangle_a$ for all $\phi \in \mathscr{D}(\sigma_a^H)$ and for all $u \in \mathscr{D}(B_a^H) = \{g \in \mathscr{D}(\sigma_a^H) : \exists f \in L_a^2(\Omega) \text{ s.t. } \sigma_a^H(g,\phi) = \langle f, \phi \rangle_a \ \forall \phi \in \mathscr{D}(\sigma_a^H)\}$.

It is known that the space $WH^1(\Omega)$ is a Banach space and is dense and compactly embedded in $L^2(\Omega)$ (Nhieu, 2001). Thus, as in Lemma 3.2.3, we have the following result on the operators A_a^H and B_a^H.

Lemma 3.5.1. *The operators $A_a^H : \mathscr{D}(A_a^H) \to L_a^2(\Omega)$ and $B_a^H : \mathscr{D}(B_a^H) \to L_a^2(\Omega)$ are closed, densely-defined, and self-adjoint. Moreover, these operators have purely discrete spectra.*

We also know that if $f \in WH^1(\Omega)$, then $|f| \in WH^1(\Omega)$ (Garofalo and Nhieu, 1998). Thus, Corollary 3.2.5 extends to the following result.

Corollary 3.5.2. *Let $y^0 \in L_a^2(\Omega)$. Then $-A_a^H$ generates a semigroup of operators $(\mathscr{T}_a^{A^H}(t))_{t\geq 0}$. Additionally, the semigroup $(\mathscr{T}_a^{A^H}(t))_{t\geq 0}$ is positive. Finally, if $\|\mathscr{M}_a y^0\|_\infty \leq 1$, then $\|\mathscr{M}_a \mathscr{T}_a^{A^H}(t) y^0\|_\infty \leq 1$ for all $t \geq 0$.*

Using the same arguments as in the proof of Corollary 3.2.5, we have the following result.

Corollary 3.5.3. *The operator $-B_a^H$ generates a semigroup of operators $(\mathscr{T}_a^{B^H}(t))_{t\geq 0}$ on $L_a^2(\Omega)$. Moreover, the semigroup $(\mathscr{T}_a^{B^H}(t))_{t\geq 0}$ is positive.*

Next, we will establish the long-term stability properties of the semigroups associated with the sub-elliptic operators.

Lemma 3.5.4. *The semigroups $(\mathscr{T}_a^A(t))_{t\geq 0}$ and $(\mathscr{T}_a^B(t))_{t\geq 0}$ that are generated by the operators $-A_a$ and $-B_a$, respectively, are analytic. Additionally, these semigroups have the following mass conservation property: if $y^0 \geq 0$ and $\int_\Omega y^0(\mathbf{x})d\mathbf{x} = 1$, then $\int_\Omega (\mathscr{T}_a^A(t)y^0)(\mathbf{x})d\mathbf{x} = \int_\Omega (\mathscr{T}_a^B(t)y^0)(\mathbf{x})d\mathbf{x} = 1$ for all $t \geq 0$. Moreover, 0 is a simple eigenvalue of the operators $-A_a$ and $-B_a$. Hence, if $y^0 \geq 0$ and $\int_\Omega y^0(\mathbf{x})d\mathbf{x} = \int_\Omega f(\mathbf{x})d\mathbf{x} = 1$, then the following estimates hold:*

$$\|\mathscr{T}_a^A(t)y^0 - f\|_a \leq M_0 e^{-\lambda t}\|y^0 - f\|_a, \tag{3.51}$$

$$\|\mathscr{T}_a^B(t)y^0 - f\|_a \leq \tilde{M}_0 e^{-\tilde{\lambda} t}\|y^0 - f\|_a \tag{3.52}$$

for some positive constants $M_0, \tilde{M}_0, \lambda, \tilde{\lambda}$ and all $t \geq 0$.

Proof. The proof of analyticity of the semigroups follows along the lines of the proof of Lemma 3.2.7.

In order to establish the stability properties of the semigroups, we will identify the eigenvectors associated with the eigenvalue 0. In the proof of Lemma 3.2.7, we used the Poincaré inequality to establish the uniqueness of the eigenvector of constant functions, corresponding to the eigenvalue 0 of the Laplacian Δ. It is not clear if the Poincaré inequality holds for the operator Δ_H. Hence, instead of using a Poincaré inequality, we will prove that the kernel of the operator Δ_H consists only of constant functions. Suppose $u \in D(A^H)$ is such that $A^H u = \mathbf{0}$, where $A^H := A^H_{\mathbf{1}}$. Since the operator A^H satisfies the Lie Rank condition, from regularity results due to Hormander (Robinson, 1991; Bramanti, 2014), we can infer that u is locally smooth everywhere in Ω. Then we know that, for a given horizontal curve $\gamma : [0,1] \to \Omega$,

$$u(\gamma(1)) - u(\gamma(0)) = \int_0^1 \sum_{i=1}^m a_i(t)\mathbf{X}u(\gamma(t))dt = 0 \tag{3.53}$$

where $a_i(t)$ are the essentially bounded functions associated with the curve $\gamma(t)$ according to (3.42). Note that we require the local smoothness of u to make sense of the term $\int_0^1 \sum_{i=1}^m a_i(t)\mathbf{X}u(\gamma(t))dt$. Due to the Lie Rank condition, we can choose $\gamma(t)$ to be such that $\gamma(0)$ and $\gamma(1)$ are given initial and final conditions in Ω. Hence, we have that u is constant everywhere on Ω. This implies that $\Delta_H f = \mathbf{0}$, and hence $A_a^H a = B_a^H a = \mathbf{0}$. □

3.6 Stabilization of a System of Hypoelliptic Reaction-Diffusion Equations to Target Probability Densities with Disconnected Supports

In Sections 3.2-3.3, the probability densities that we stabilized were assumed to be uniformly bounded from below by a positive number. Without this assumption, the semigroups that were constructed to establish the controllability and stability results would not be irreducible. In this section, we will introduce a semilinear PDE model for stabilization of a swarm to probability densities that possibly have supports that are disconnected.

As in Section 3.5, Ω will denote an open bounded subset of a Lie group, and we have a collection of left-invariant vector fields $\mathscr{X} = \{\mathbf{X}_1, ..., \mathbf{X}_m\}$ satisfying the Lie Rank condition. Let $A^H := A_{\mathbf{1}}^H = \Delta_H$ be the operator defined in Section 3.5, where $\mathbf{1}$ denotes the function that is equal to 1 almost everywhere on Ω.

We will consider the following PDE model

$$\begin{aligned}
(y_1)_t &= -A^H y - u_1(\mathbf{x},t)y_1 + u_2(\mathbf{x},t)y_2 \quad && in\ \ \Omega \times [0,T] \\
(y_2)_t &= u_1(\mathbf{x},t)y_1 + u_2(\mathbf{x},t)y_2 && in\ \ \Omega \times [0,T] \\
\mathbf{y}(\cdot,0) &= \mathbf{y}^0 && in\ \ \Omega \\
\mathbf{n}\cdot\nabla y_1 &= 0 && in\ \ \partial\Omega \times [0,T].
\end{aligned} \tag{3.54}$$

This PDE model is the forward equation of a hybrid switching process, as defined in Section 3.3. Let $\mathbf{L}^2(\Omega) = L^2(\Omega) \times L^2(\Omega)$ and $\mathbf{L}^\infty(\Omega) = L^\infty(\Omega) \times L^\infty(\Omega)$ with the standard norms inherited from the spaces $L^2(\Omega)$ and $L^\infty(\Omega)$, respectively, as defined in Section 3.3.

We will consider the following problem in this section.

Problem 3.6.1. *Let $y^d \in L^\infty(\Omega)$ be a target probability density. Construct a mean-field feedback law $K_i : \mathbf{L}^2(\Omega) \to L^\infty(\Omega)$ such that if $u_i(\cdot,t) = K_i(\mathbf{y}(t))$ for all $i \in \{1,2\}$ and all $t \geq 0$, then the system* (3.54) *is globally asymptotically stable about the equilibrium $\mathbf{y}^d = [\mathbf{0} \;\; y^d]^T$.*

Before we address this problem, we make some additional assumptions about the domain Ω and the operator A^H. Particularly, we will **assume** that the domain Ω and/or the operator A^H satisfy one of the two following properties:

1. If $\Omega \neq G$, then Ω is a bounded subset of $\mathbb{R}^N$, equipped with the usual Lie group structure; $-A^H = \Delta$ is the Laplacian; and Ω is either a $C^{1,1}$ domain in the sense of Definition 3.1.1 or is convex.

2. The set Ω is a compact Lie group G without a boundary.

Given these assumptions, we have the following result due to Gaussian estimates proved by (Choulli and Kayser, 2015) for the Laplacian Δ, and by (Jerison and Sánchez-Calle, 1986) for sub-Laplacians Δ_H. We will use this result in the subsequent analysis.

Theorem 3.6.2. *Let $(\mathscr{T}^{A^H}(t))_{t\geq 0}$ be the semigroup generated by the operator $-A^H$. Let $y^0 \in L^2(\Omega)$ be non-negative. Then there exists a constant $C > 0$ and time $T > 0$, independent of y^0, such that $\mathscr{T}^{A^H}(t)y^0 \geq C\|y^0\|_1$ for all $t \geq T$.*

In order to address Problem 3.6.1, for each $i \in \{1,2\}$, we define the maps $F_i : L^2(\Omega) \to L^2(\Omega)$ given by

$$(F_i(f))(\mathbf{x}) = r_i(f(\mathbf{x}) - y^d(\mathbf{x})) \tag{3.55}$$

for almost every $\mathbf{x} \in \Omega$ and all $f \in L^2(\Omega)$, where $r_i : \mathbb{R} \to [0,c]$ are globally Lipschitz functions for some positive number c, such that the functions r_1 and r_2 have supports equal

to the intervals $[0,\infty)$ and $(-\infty,0]$, respectively. Our candidate mean-field feedback law K_i for addressing Problem 3.6.1 will be $K_i(\mathbf{y}) = F_i(y_1)$ for each $i \in \{1,2\}$.

Then the resulting *closed-loop* PDE is given by

$$\begin{aligned}
(y_1)_t &= -A^H y - F_1(y_2)y_1 + F_2(y_2)y_2 \qquad \text{in } \Omega \times [0,T] \\
(y_2)_t &= F_1(y_2)y_1 - F_2(y_2)y_2 \qquad \text{in } \Omega \times [0,T] \\
\mathbf{y}(\cdot,0) &= \mathbf{y}^0 \qquad \text{in } \Omega \\
\mathbf{n}\cdot\nabla(y) &= 0 \qquad \text{in } \partial\Omega \times [0,T].
\end{aligned} \tag{3.56}$$

In order to perform stability analysis of the PDE (3.56), we will need a suitable notion of a solution. Toward this end, we introduce the following notion of solutions for semilinear PDEs (Lunardi, 2012).

Definition 3.6.3. *Let $(\mathscr{T}^{A^H}(t))_{t\geq 0}$ be the semigroup generated by the operator $-A^H$. We will say that the PDE has a local mild solution if there exists $T > 0$ and $\mathbf{y} \in C([0,T];\mathbf{L}^2(\Omega))$ such that*

$$\begin{aligned}
y_1(\cdot,t) &= \mathscr{T}^{A^H}(t)y_1^0 - \int_0^t \mathscr{T}^{A^H}(t-s)\Big(F^1(y_2(\cdot,s))y_1(\cdot,s)\Big)ds \\
&\quad + \int_0^t \mathscr{T}^{A^H}(t-s)\Big(F^2(y_2(\cdot,s))y_2(\cdot,s)\Big)ds \\
y_2(\cdot,t) &= y_2^0 + \int_0^t F^1(y_2(\cdot,s))y_1(\cdot,s)ds - \int_0^t F^2(y_2(\cdot,s))y_2(\cdot,s)ds
\end{aligned} \tag{3.57}$$

for all $t \in [0,T]$.

We will say that the PDE (3.56) *has a unique global solution if it has unique mild solution for every $T > 0$.*

In order to establish the existence of solutions of the PDE (3.56), we consider the map $\mathbf{G} : \mathbf{L}^2(\Omega) \to \mathbf{L}^2(\Omega)$ defined by

$$\mathbf{G}(\mathbf{f}) = \begin{bmatrix} -F_1(f_2)f_1 + F_2(f_2)f_2 \\ +F_1(f_2)f_1 - F_2(f_2)f_2 \end{bmatrix}$$

for each $\mathbf{f} \in \mathbf{L}^2(\Omega)$. We will also need the operator $\mathbf{A} : D(A) :\to \mathbf{L}^2(\Omega)$ defined by

$$\mathbf{Ay} = \begin{bmatrix} A^H y_1 \\ \mathbf{0} \end{bmatrix}$$

for all $\mathbf{y} \in D(\mathbf{A}) = D(A^H) \times L^2(\Omega)$.

Then we have the following result.

Lemma 3.6.4. *The map* $\mathbf{G}$ *is locally Lipschitz continuous everywhere on* $\mathbf{L}^2(\Omega)$.

Proof. We only show that the map $\mathbf{y} \mapsto F_1(y_2)y_1$ from $\mathbf{L}^2(\Omega)$ to $L^2(\Omega)$ is locally Lipschitz everywhere on $\mathbf{L}^2(\Omega)$. The rest of the proof is a straightforward extension. Let $R > 0$ and $\mathbf{y}^0, \mathbf{p}, \mathbf{q} \in \mathbf{L}^2(\Omega)$ with $\|\mathbf{p} - \mathbf{y}^0\|_2 \leq R$ and $\|\mathbf{q} - \mathbf{y}^0\|_2 \leq R$. Then we have that

$$\begin{aligned} \|F_1(p_2)p_1 - F_1(q_2)q_1\|_2^2 &= \int_\Omega |r_1(p_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x}) - r_1(q_2(\mathbf{x}) - y^d(\mathbf{x}))q_1(\mathbf{x})|^2 d\mathbf{x} \\ &\leq \int_\Omega |r_1(p_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x}) - r_1(q_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x})|^2 d\mathbf{x} \\ &\quad + \int_\Omega |r_1(q_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x}) - r_1(q_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x})|^2 d\mathbf{x} \end{aligned}$$

Since function r_1 is a globally Lipschitz function that is bounded from above by a constant c, and from below by 0, we can conclude that

$$\begin{aligned} \|F_1(p_2)p_1 &- F_1(q_2)q_1\|_2^2 \\ &\leq C\|r_1(p_2(\cdot) - y^d(\cdot) - r_1(q_2(\cdot) - y^d(\cdot))\|_\infty \int_\Omega |p_1(\mathbf{x})|^2 d\mathbf{x} \\ &\quad + C \int_\Omega |r^1(q_2(\mathbf{x}) - y^d(\mathbf{x}))|^2 d\mathbf{x} \int_\Omega |p_1(\mathbf{x}) - q_1(\mathbf{x})|^2 d\mathbf{x} \\ &\leq C \int_\Omega |r_1(q_2(\mathbf{x}) - y^d(\mathbf{x}))p_1(\mathbf{x}) - r_1(q_2(\mathbf{x}) - y^d(\mathbf{x})|^2 d\mathbf{x} \int_\Omega |p_2(\mathbf{x})|^2 d\mathbf{x} \\ &\quad + C \int_\Omega |r_1(q_1(\mathbf{x}) - y^d(\mathbf{x}))|^2 d\mathbf{x} \int_\Omega |p_1(\mathbf{x}) - q_1(\mathbf{x})|^2 d\mathbf{x} \\ &\leq C \int_\Omega |r^1(q^2(\mathbf{x}) - y^d(\mathbf{x}))p^1(\mathbf{x}) - r^1(q^2(\mathbf{x}) - y^d(\mathbf{x}))|^2 d\mathbf{x} \\ &\quad + C \int_\Omega |p_1(\mathbf{x}) - q_1(\mathbf{x})|^2 d\mathbf{x} \\ &\leq C \int_\Omega |p_2(\mathbf{x}) - q_2(\mathbf{x})|^2 d\mathbf{x} + C \int_\Omega |p_1(\mathbf{x}) - q_1(\mathbf{x})|^2 d\mathbf{x} \end{aligned}$$

for some $C > 0$ depending only on the constants R and c. □

Using Lemma 3.6.4, we can conclude the following theorem on existence of a mild solution of the PDE (3.56) by applying standard results on existence of solutions of semilinear PDEs (Lunardi, 2012)[Theorem 7.1.2].

Theorem 3.6.5. *Let* $\mathbf{y}^0 \in L^2(\Omega)$. *There exists a unique local mild solution of the PDE* (3.56).

Proof. We have shown that the map $\mathbf{G}$ is locally Lipschitz everyhwhere on $\mathbf{L}^2(\Omega)$. □

Our next goal will be to construct global solutions of the PDE (3.56). Further ahead, we will show that the solutions of the PDE (3.56) remain essentially bounded if the initial condition is essentially bounded. Toward this end, we first establish this property for a related autonomous linear PDE.

Lemma 3.6.6. *Suppose that* $\mathbf{y} \in \mathbf{L}^\infty(\Omega)$. *Let* $\mathbf{a} \in \mathbf{L}^\infty(\Omega)$ *be non-negative. Consider the linear bounded operator* $B : \mathbf{L}^2(\Omega) \to \mathbf{L}^2(\Omega)$ *defined by*

$$(B\mathbf{y})(\mathbf{x}) = \begin{bmatrix} -a_1(\mathbf{x})y_1(\mathbf{x}) + a_2(\mathbf{x})y_2(\mathbf{x}) \\ a_1(\mathbf{x})y_1(\mathbf{x}) - a_2(\mathbf{x})y_2(\mathbf{x}) \end{bmatrix}$$

for almost every $\mathbf{x} \in \Omega$ *and all* $\mathbf{y} \in \mathbf{L}^2(\Omega)$. *Let* $(\mathscr{T}^{\mathbf{C}}(t))_{t\geq 0}$ *be the semigroup generated by the operator* $\mathbf{C} = -\mathbf{A} + \mathbf{B}$. *Then* $\|\mathscr{T}^{\mathbf{C}}(t)\mathbf{y}^0\|_\infty \leq e^{\|\mathbf{a}\|_\infty t}\|\mathbf{y}^0\|_\infty$ *for all* $t \geq 0$.

Proof. We know that the operator $\mathbf{A}$ generates a semigroup $(\mathscr{T}^{\mathbf{A}}(t))_{t\geq 0}$ given by

$$\mathscr{T}^{\mathbf{A}}(t) = \begin{bmatrix} \mathscr{T}^{A^H}(t) & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} \tag{3.58}$$

for all $t \geq 0$. Moreover, the semigroup $(\mathscr{T}^{\mathbf{A}}(t))_{t\geq 0}$ satisfies $\|\mathscr{T}^{\mathbf{A}}(t)\mathbf{y}^0\|_\infty \leq \|\mathbf{y}^0\|_\infty$ for all $\mathbf{y}^0 \in \mathbf{L}^\infty(\Omega)$ and $t \geq 0$ (Corollaries 3.2.5 and 3.5.2). Additionally, we know that the

semigroup $(\mathscr{T}^{\mathbf{B}}(t))_{t\geq 0}$ generated by the bounded operator $\mathbf{B}$ satisfies the estimate $\|\mathscr{T}^{\mathbf{B}}(t)\mathbf{y}^0\|_\infty \leq e^{\|\mathbf{a}\|_\infty t}\|\mathbf{y}^0\|_\infty$. Then the result follows from the *Lie-Trotter product formula* (Engel and Nagel, 2000)[Corollary III.5.8], by noting that $\mathscr{T}^{\mathbf{C}}(t) = \lim_{N\to 0}(\mathscr{T}^{\mathbf{A}}(\frac{t}{N})\mathscr{T}^{\mathbf{B}}(\frac{t}{N}))^N$, where the limit holds in the strong operator topology, for all $t \geq 0$. □

Now we can show that the $\mathbf{L}^\infty-$ estimate proved in the last lemma can be extended to a class of non-autonomous linear systems that can be treated as autonomous linear systems over certain intervals of time.

Lemma 3.6.7. *Suppose that* $\mathbf{y}^0 \in L^\infty(\Omega)$ *and* $T > 0$*. For a positive constant c, let* $a_1, a_2 \in L^2(0,T;L^2(\Omega))$ *be non-negative and piecewise constant with respect to time with* $\|a_1(t)\|_\infty \leq c$ *and* $\|a_2(t)\|_\infty \leq c$ *for all* $t \in [0,T]$*. Then suppose that* $\mathbf{y} \in C([0,T];\mathbf{L}^2(\Omega))$ *is given by*

$$\begin{aligned} y_1(\cdot,t) &= \mathscr{T}^{A^H}(t)y_1^0 - \int_0^t \mathscr{T}^{A^H}(t-s)\Big(a_1(\cdot,s)y_1(\cdot,s)\Big)ds \\ &\quad + \int_0^t \mathscr{T}^{A}(t-s)\Big(a_2(\cdot,s)y_2(\cdot,s)\Big)ds \\ y_2(\cdot,t) &= y_2^0 + \int_0^t a_1(\cdot,s)y_1(\cdot,s)ds - \int_0^t a_2(\cdot,s)y_2(\cdot,s)ds \end{aligned} \tag{3.59}$$

for all $t \in [0,T]$*. Then*

$$\|\mathscr{T}^{\mathbf{C}}(t)\mathbf{y}^0\|_\infty \leq e^{ct}\|\mathbf{y}^0\|_\infty \tag{3.60}$$

for all $t \in [0,T]$*.*

Proof. Let $(t_i)_{i=0}^m$ be a finite sequence of strictly increasing time instants of length $m+1 \in \mathbb{Z}_+$, with $t_0 = 0$, such that the functions a_1 and a_2 are constant over the intervals $[t_{i-1}, t_i)$ for each $i \in \{1,...,m\}$. Then, for each $i \in \{1,...,m\}$, consider the bounded operators $\mathbf{B}_i : \mathbf{L}^2(\Omega) \to \mathbf{L}^2(\Omega)$ and $\mathbf{C}_i : D(\mathbf{A}) \to \mathbf{L}^2(\Omega)$ defined by,

$$(\mathbf{B}_i\mathbf{y})(\mathbf{x}) = \begin{bmatrix} -a_1(\mathbf{x},t_{i-1})y_1(\mathbf{x}) + a_2(\mathbf{x},t_{i-1})y_2(\mathbf{x}) \\ a_1(\mathbf{x},t_{i-1})y_1(\mathbf{x}) - a_2(\mathbf{x},t_{i-1})y_2(\mathbf{x}) \end{bmatrix} \tag{3.61}$$

for almost every $\mathbf{x} \in \Omega$ and all $\mathbf{y} \in \mathbf{L}^2(\Omega)$, and $\mathbf{C}_i = \mathbf{A} + \mathbf{B}_i$, respectively. Then, for each $i \in \{1, ..., m\}$, $\mathbf{y}$ is given by,

$$\mathbf{y}(\cdot, t) = \mathscr{T}^{\mathbf{C}_i}(t - t_i)\mathscr{T}^{\mathbf{C}_{i-1}}(t_i - t_{i-1})...\mathscr{T}^{\mathbf{C}_1}(t_1) \tag{3.62}$$

for all $t \in [t_{i-1}, t_i]$. Then the result follows from Lemma 3.6.6. □

Lemma 3.6.8. *Suppose that $\mathbf{y}^0 \in L^\infty(\Omega)$ and $T > 0$. For a positive constant c, let $a_1, a_2 \in L^2(0,T;L^2(\Omega))$ be non-negative with $\|a_1(t)\|_\infty \leq c$ and $\|a_2(t)\|_\infty \leq c$ for almost every $t \in [0,T]$. Then suppose that $\mathbf{y} \in C([0,T];\mathbf{L}^2(\Omega))$ is given by*

$$\begin{aligned} y_1(\cdot,t) &= \mathscr{T}^{A^H}(t)y_1^0 - \int_0^t \mathscr{T}^{A^H}(t-s)\Big(a_1(\cdot,s)y_1(\cdot,s)\Big)ds \\ &\quad + \int_0^t \mathscr{T}^{A^H}(t-s)\Big(a_2(\cdot,s)y_2(\cdot,s)\Big)ds \\ y_2(\cdot,t) &= y_2^0 + \int_0^t a_1(\cdot,s)y_1(\cdot,s)ds - \int_0^t a_2(\cdot,s)y_2(\cdot,s)ds \end{aligned} \tag{3.63}$$

for all $t \geq 0$. Then

$$\|\mathbf{y}(\cdot,t)\|_\infty \leq e^{ct}\|\mathbf{y}^0\|_\infty \tag{3.64}$$

for all $t \in [0,T]$.

Proof. Given that $a_1, a_2 \in L^2(0,T;L^2(\Omega))$, we know that there exists sequence of piecewise (with respect to time) non-negative functions $(a_1^i)_{i=1}^\infty, (a_2^i)_{i=1}^\infty$ in $L^2(0,T;L^2(\Omega))$ such that $\lim_{i\to\infty}\|a_j^i - a_j\|_{L^2(0,T;L^2(\Omega))} = 0$, for $j = 1,2$ (Roubíček, 2013)[Proposition 1.36]. Moreover, for each $j \in \{1,2\}$, we can assume that $\|a_j^i(t)\|_\infty \leq c$ for all $t \in [0,T]$ and all $i \in \mathbb{Z}_+$. Consider the corresponding sequence $(\mathbf{y})_{i=1}^\infty$ in $C([0,T];\mathbf{L}^2(\Omega))$ defined by

$$\begin{aligned} y_1^i(\cdot,t) = &\quad \mathscr{T}^A(t)y_1^0 - \textstyle\int_0^t \mathscr{T}^A(t-s)\Big(a_1^i(\cdot,s)y_1^i(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^A(t-s)\Big(a_2^i(\cdot,s)y_2^i(\cdot,s)\Big)ds \\ y_2^i(\cdot,t) = &\quad y_2^0 + \textstyle\int_0^t a_1^i(\cdot,s)y_1^i(\cdot,s)ds - \int_0^t a_2^i(\cdot,s)y_2^i(\cdot,s)ds \end{aligned} \tag{3.65}$$

for each $i \in \mathbb{Z}_+$. Let $\mathbf{e}^i \in C([0,T];\mathbf{L}^2(\Omega))$ be given by $\mathbf{e}^i = \mathbf{y}^i - \mathbf{y}$ for each $i \in \mathbb{Z}_+$. Then,

from equations (3.63) and (3.65), we know that $\mathbf{e}^i$ satisfies

$$\begin{aligned} e_1^i(\cdot,t) &= -\int_0^t \mathscr{T}^A(t-s)\Big(a_1^i(\cdot,s)y_1^i(\cdot,s)\Big)ds+\int_0^t \mathscr{T}^A(t-s)\Big(a_2^i(\cdot,s)y_2^i(\cdot,s)\Big)ds \\ &\quad +\int_0^t \mathscr{T}^A(t-s)\Big(a_1(\cdot,s)y_1(\cdot,s)\Big)ds-\int_0^t \mathscr{T}^A(t-s)\Big(a_2(\cdot,s)y_2(\cdot,s)\Big)ds \\ &= -\int_0^t \mathscr{T}^A(t-s)\Big((a_1^i(\cdot,s)-a_1(\cdot,s))y_1^i(\cdot,s)\Big)ds \\ &\quad +\int_0^t \mathscr{T}^A(t-s)\Big(a_1(\cdot,s)(y_1(\cdot,s)-y_1^i(\cdot,s))\Big)ds \\ &\quad +\int_0^t \mathscr{T}^A(t-s)\Big((a_2^i(\cdot,s)-a_2(\cdot,s))y_2^i(\cdot,s)\Big)ds \\ &\quad -\int_0^t \mathscr{T}^A(t-s)\Big(a_2(\cdot,s)(y_2(\cdot,s)-y_2^i(\cdot,s))\Big)ds \end{aligned} \tag{3.66}$$

for all $t\in[0,T]$. Considering the fact the semigroup $\mathscr{T}^A(t-s)$ is contractive (Corollaries 3.2.5 and 3.5.2), and that a_j^i and y_j^i are uniformly bounded in $L^\infty((0,T)\times\Omega)$, we can conclude that there exists a constant $\alpha>0$ such that

$$\begin{aligned} \|e_1^i(\cdot,t)\|_2 &\le \alpha\|a_2^i-a_2\|_{L^2(0,T;L^2(\Omega))}\|y_1^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_1\|_{L^2(0,T;L^2(\Omega))}\|e_1^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_2^i-a_2\|_{L^2(0,T;L^2(\Omega))}\|y_2^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_2\|_{L^2(0,T;L^2(\Omega))}\|e_2^i\|_{L^2(0,T;L^2(\Omega))} \end{aligned} \tag{3.67}$$

for all $t\in[0,T]$. Similarly, we can conclude the estimate

$$\begin{aligned} \|e_2^i(\cdot,t)\|_2 &\le \alpha\|a_1^i-a_1\|_{L^2(0,T;L^2(\Omega))}\|y_1^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_1\|_{L^2(0,T;L^2(\Omega))}\|e_1^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_2^i-a_2\|_{L^2(0,T;L^2(\Omega))}\|y_2^i\|_{L^2(0,T;L^2(\Omega))} \\ &\quad +\alpha\|a_2\|_{L^2(0,T;L^2(\Omega))}\|e_2^i\|_{L^2(0,T;L^2(\Omega))} \end{aligned} \tag{3.68}$$

for all $t\in[0,T]$.

Then, by considering the sum $\|e_2^i(\cdot,t)\|_2+\|e_1^i(\cdot,t)\|_2$, combining the two inequalities (3.67) and (3.68), and applying the integral form of Gronwall's inequality (Evans, 1998),

we have that

$$\|e_2^i(\cdot,t)\|_2 + \|e_1^i(\cdot,t)\|_2 \leq C_2^i(1 + C_1^i t e^{C_1^i t}) \tag{3.69}$$

for all $t \in [0,T]$, where $C_1^i = \max\{2\|a_2\|_{L^2(0,T;L^2(\Omega))}, 2\|a_2\|_{L^2(0,T;L^2(\Omega))}\}$ and $C_2^i = 2\|a_1 - a_1^i\|_{L^2(0,T;L^2(\Omega))}\|y_1\|_{L^2(0,T;L^2(\Omega))} + 2\|a_2 - a_2^i\|_{L^2(0,T;L^2(\Omega))}\|y_2\|_{L^2(0,T;L^2(\Omega))}$, for all $i \in \mathbb{Z}_+$. From the inequality (3.69), we can infer that

$$\lim_{i\to\infty} \|\mathbf{e}^i\|_{C([0,T];\mathbf{L}^2(\Omega))} = 0$$

Considering the estimate (3.60), and the fact that the set

$$R_c := \{\mathbf{u} \in C([0,T];\mathbf{L}^2(\Omega)); \|\mathbf{u}(t)\|_\infty \leq c \ \forall t \in [0,T]\} \tag{3.70}$$

is a closed subset of $C([0,T];\mathbf{L}^2(\Omega))$ for every $c > 0$, we can conclude that $\mathbf{y}$ satisfies the estimate (3.64). □

From the above lemma, we can conclude the following theorem on global existence of solutions of the PDE (3.56).

Theorem 3.6.9. *Suppose that* $\mathbf{y}^0 \in L^\infty(\Omega)$. *Then the PDE* (3.56) *has a unique global mild solution.*

Next, our goal will be to prove that $\mathbf{y}^d$ is the globally asymptotically stable equilibrium of the system (3.56). We prove some preliminary results for this.

Lemma 3.6.10. *Suppose that* $y^0 \in L^\infty(\Omega)$ *and* $T > 0$. *Let* $a \in L^\infty(\Omega)$ *be non-negative. Consider the multiplication operator* $B : \mathbf{L}^2(\Omega) \to \mathbf{L}^2(\Omega)$ *defined by*

$$(By)(\mathbf{x}) = -a(\mathbf{x})y(\mathbf{x})$$

for all $\mathbf{x} \in \Omega$ *and all* $y \in L^2(\Omega)$. *Let* $(\mathscr{T}^C(t))_{t\geq 0}$ *be the semigroup generated by the operator* $C = -A^H + B$. *Then* $\|\mathscr{T}^C(t)y^0\|_\infty \leq \|y^0\|_\infty$ *for all* $t \geq 0$.

Proof. We know that if $(\mathscr{T}^{A^H}(t))_{t\geq 0}$ is the semigroup generated by the operator $-A^H$, then from Corollaries 3.2.5 and 3.5.2, $\|\mathscr{T}^C(t)\mathbf{y}^0\|_\infty \leq \|y^0\|_\infty$ for all $t \geq 0$. Moreover, B generates the multiplication semigroup $(e^{-a(\cdot)t})_{t\geq 0}$. Then the result follows from the Lie-Trotter formula (Engel and Nagel, 2000). □

Lemma 3.6.11. *Let $T > 0$. Let $f, a \in L^2(0,T;L^2(\Omega))$ be non-negative such that $\|f(t)\|_\infty$ and $\|a(t)\|_\infty$ are bounded by a constant $C > 0$ almost everywhere on $t \in [0,T]$. Suppose that $e \in C([0,T];L^2(\Omega))$ is given by*

$$e(\cdot,t) = \quad -\int_0^t \mathscr{T}^{A^H}(t-s)\Big(a(\cdot,s)e(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^{A^H}(t-s)f(\cdot,s)ds$$

for all $t \in [0,T]$. Then $e(\cdot,t)$ is non-negative for all $t \in [0,T]$.

Proof. The proof follows a similar line of argument as the proof of Lemma 3.6.8. Therefore, we only provide a sketch of the proof. As in the proof of Lemma 3.6.8, for a given $a \in L^2(0,T;L^2(\Omega))$ we can construct a sequence $(a^i)_{i=1}^\infty$ in $L^2(0,T;L^2(\Omega))$ that is piecewise constant in time, and converging in $L^2(0,T;L^2(\Omega))$ with $\|a^i(t)\|_\infty$ bounded almost everywhere on $[0,T]$ by $C > 0$. Let $(e_i)_{i=1}^\infty$ in $C([0,T];L^2(\Omega))$ be given by

$$e_i(\cdot,t) = \quad -\int_0^t \mathscr{T}^{A^H}(t-s)\Big(a^i(\cdot,s)e(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^{A^H}(t-s)f(\cdot,s)ds$$

for all $t \in [0,T]$. Using Lemma 3.6.10, we can conclude that $(e_i)_{i=1}^\infty$ is non-negative for each $i \in \mathbb{Z}_+$. Then, using the fact that the sequences $(e_i)_{i=1}^\infty$ and $(a_i)_{i=1}^\infty$ are uniformly bounded in the spaces $C([0,T];L^2(\Omega))$ and $L^\infty((0,T)\times\Omega)$, respectively, and applying Gronwall's inequality, the result follows. □

We can use the above lemma to prove the following result, which will enable us to show further on that the decay of the solution $\mathbf{y}$ of the PDE (3.56) toward 0 can be controlled by the decay of the solution of a related linear PDE.

Theorem 3.6.12. ***(Comparison Principle)***

Let $T > 0$. Let $y^0 \in L^2(\Omega)$ and $f, g \in L^2(0,T;L^2(\Omega))$ be non-negative such that $\|f(t)\|_\infty$ and $\|g(t)\|_\infty$ are bounded by a constant $C_1 > 0$ almost everywhere on $t \in [0,T]$. Let $C = -A^H - \|g\|_\infty \mathbf{I}$. Let $y(\cdot,t)$ be given by

$$y(\cdot,t) = \mathscr{T}^{A^H}(t)y^0 - \int_0^t \mathscr{T}^{A^H}(t-s)\Big(g(\cdot,s)y(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^{A^H}(t-s)f(\cdot,s)ds \quad (3.71)$$

for all $t \in [0,T]$. Then $y(\cdot,t) \geq \mathscr{T}^C(t)y^0$ for all $t \in [0,T]$, where $(\mathscr{T}^C(t))_{t\geq 0}$ is the semigroup generated by the operator C.

Proof. Let $\tilde{y}(\cdot,t) = \mathscr{T}^C(t)y^0$ for all $t \geq 0$. Then, we know that

$$\tilde{y}(\cdot,t) = \mathscr{T}^{A^H}(t)y^0 - \int_0^t \mathscr{T}^{A^H}(t-s)\|g\|_\infty \tilde{y}(\cdot,s)ds \quad (3.72)$$

for all $t \in [0,T]$. Let $e = y - \tilde{y}$. Then we have that

$$\begin{aligned} e(\cdot,t) &= -\int_0^t \mathscr{T}^{A^H}(t-s)\Big(g(\cdot,s)y(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^{A^H}(t-s)f(\cdot,s)ds \\ &\quad + \int_0^t \mathscr{T}^{A^H}(t-s)\|g\|_\infty \tilde{y}(\cdot,s)ds \\ &= -\int_0^t \mathscr{T}^{A^H}(t-s)\Big((g(\cdot,s) - \|g\|_\infty)e(\cdot,s)\Big)ds + \int_0^t \mathscr{T}^{A^H}(t-s)f(\cdot,s)ds \\ &\quad - \int_0^t \mathscr{T}^{A^H}(t-s)\Big(g(\cdot,s) - \|g\|_\infty)e(\cdot,s)\Big)ds \\ &\quad + \int_0^t \mathscr{T}^{A^H}(t-s)\Big((\|g\|_\infty - g(\cdot,s))\tilde{y}(\cdot,s)\Big)ds \end{aligned} \quad (3.73)$$

for all $t \in [0,T]$. Then the result follows from the non-negativity of e, which is a consequence of Lemma 3.6.11. □

Using the above comparison principle, we can prove the following result, which will be used later to establish the strict positivity of solutions of (3.56) over at least a small time interval.

Theorem 3.6.13. ***(Positive Lower Bound of Solutions)*** *Let $T > 0$. Let $y^0 \in L^2(\Omega)$ and $f, g \in L^2(0,T;L^2(\Omega))$ be non-negative such that $\|f(t)\|_\infty$ and $\|g(t)\|_\infty$ are bounded by a constant $C_1 > 0$ almost everywhere on $t \in [0,T]$. Let $y(\cdot,t)$ be given by*

$$y(\cdot,t) = \mathscr{T}^A(t)y^0 - \int_0^t \mathscr{T}^A(t-s)\Big(g(\cdot,s)y(\cdot,s)\Big)ds \int_0^t \mathscr{T}^A(t-s)f(\cdot,s)ds \tag{3.74}$$

for all $t \in [0,T]$. Then there exist $\tau, \varepsilon, \delta > 0$, independent of y^0 and $T > 0$, such that if $\tau + \delta < T$, then $y(\cdot,t) \geq \varepsilon \|y^0\|_1$ for all $t \geq [\tau, \tau+\delta]$.

Proof. We know from Theorem 3.6.2 that there exists a constant $C > 0$ and time $T > 0$, independent of y^0, such that $\mathscr{T}^{A^H}(t)y^0 \geq C\|y^0\|_1$ for all $t \geq T$. Let $C = -A^H - \|g\|_\infty \mathbf{I}$. Then the semigroup $(\mathscr{T}^C(t))$ generated by the operator C is given by $\mathscr{T}^{A^H}(t) = e^{-\|g\|_\infty t}\mathscr{T}^{A^H}(t)$ for all $t \geq 0$. Then the result follows from Theorem 3.6.12. □

In the following theorem, we establish a fundamental result that the PDE (3.56) conserves mass and maintains positivity.

Theorem 3.6.14. *Let $\mathbf{y} \in \mathbf{L}^\infty(\Omega)$ be non-negative. Then the unique global mild solution of the PDE* (3.56) *is non-negative, and $\|\mathbf{y}(\cdot,t)\|_1 = \|\mathbf{y}^0\|$ for all $t \geq 0$.*

Proof. The conservation of mass is a simple consequence of taking the inner product of the solution of (3.56) with a constant function. The positivity property of solutions follows from (Duprez and Perasso, 2017)[Theorem 1] by noting that, if $\lambda > 0$ is large enough, then $G(\mathbf{y}) + \lambda \mathbf{y} \geq 0$ for all $\mathbf{y} \in \mathbf{L}^2(\Omega)$ that are non-negative. □

From this point on, we will need some new notation. For a function $f \in L^2(\Omega)$, we define $f_+ := \frac{|f|+f}{2}$, the projection of the function f onto the set of non-negative functions in $L^2(\Omega)$, and $f_- := -\frac{|f|-f}{2}$, the projection of the function f onto the set of non-positive functions in $L^2(\Omega)$. Given these definitions, we have the following result on partial monotonicity of solutions of the PDE (3.56).

Proposition 3.6.15. ***(Partial Monotonicity of Solutions)*** *Let* $\mathbf{y} \in \mathbf{L}^{\infty}(\Omega)$ *be positive. The unique global mild solution of the PDE* (3.56) *satisfies*

$$(y^d - y_2(\cdot,t))_+ \leq (y^d - y_2(\cdot,s))_+ \tag{3.75}$$

$$(y^d - y_2(\cdot,t))_- \geq (y^d - y_2(\cdot,s))_- \tag{3.76}$$

for all $t \geq s \geq 0$.

Proof. We will only prove the first inequality (3.75). Since $\mathbf{y}^0 \in \mathbf{L}^{\infty}(\Omega)$, we know that $y_2 \in C([0,1];L^2(\Omega))$ and $\|y_2(t)\|_{\infty}$ is uniformly bounded over $[0,T]$. Assume that $y^d - y_2^0$ is non-zero and non-negative on a set $\Omega_1 \subseteq \Omega$ of positive measure. For the sake of contradiction, suppose that there exists $t_2 \in (0,T]$ such that $y_2(\cdot,t_2)$ is greater than y^d on a subset of Ω_1 that has positive Lebesgue measure. Then, due to the fact that $y_2 \in C([0,T];L^2(\Omega))$, there must exist $t_1 \in (0,t_2)$ and a measurable set $\Omega_2 \subset \Omega_1$ of positive Lebesgue measure, such that for each $s \in [t_1,t_2]$, $y_2(\mathbf{x},s) \geq y^d(\mathbf{x})$ for almost every $\mathbf{x} \in \Omega_2$, with $y_2(\mathbf{x},t_2) \neq y_2(\mathbf{x},s)$ for almost every $\mathbf{x} \in \Omega_2$ and a subset of $[t_1,t_2]$ with positive Lesbesgue measure. However, we know that

$$y_2(\cdot,t) = y_2(\cdot,t_1) + \int_s^t F_1(y_2(\cdot,\tau))y_1(\cdot,\tau)d\tau - \int_s^t F_2(y_2(\cdot,\tau))y_1(\cdot,\tau)d\tau \tag{3.77}$$

for all $t \in [t_1,t_2]$. This implies that

$$\begin{aligned} y_2(\mathbf{x},t) &= y_2(\mathbf{x},t_1) + \int_s^t F_1(y_2(\mathbf{x},\tau))y_1(\mathbf{x},\tau)d\tau - \int_s^t F_2(y_2(\mathbf{x},\tau))y_1(\mathbf{x},\tau)d\tau \\ &= y_2(\mathbf{x},t_1) - \int_s^t r_2(y_2(\mathbf{x},\tau) - y^d(\mathbf{x}))y_1(\mathbf{x},\tau)d\tau \end{aligned} \tag{3.78}$$

for almost every $\mathbf{x} \in \Omega_2$ and for all $t \in [t_1,t_2]$. Since the functions y_1 and r_2 are both non-negative, we arrive at a contradiction that $y_2(\mathbf{x},t) \leq y_2(\mathbf{x},t_1)$ for almost every $\mathbf{x} \in \Omega_1$ and for all $t \in [t_1,t_2]$. Hence, we must have that

$$y_2(\mathbf{x},t) = y_2(\mathbf{x},t_1) + \int_s^t r_1(y_2(\mathbf{x},\tau) - y^d(\mathbf{x}))y_1(\mathbf{x},\tau)d\tau \tag{3.79}$$

for almost every $\mathbf{x} \in \Omega_1$ and for all $t \in [0,T]$. This implies that y_2 is non-decreasing with time, and that it is less than or equal to y^d almost everywhere on Ω_1. This proves the first inequality (3.75).

Using a similar argument, based on the fact that r_1 and r_2 are non-negative bounded functions, we can arrive at the second inequality (3.76). □

Using the above proposition, we will establish global asymptotic stability of the system (3.56) in the L^1 norm. Toward this end, we first establish marginal stability of the system about the equilibrium distribution $\mathbf{y}^d$.

Theorem 3.6.16. *(L^1-Lyapunov Stability) Let $\mathbf{y}^0 \in L^\infty(\Omega)$ be non-negative and $\int_\Omega \mathbf{y}^0(\mathbf{x})d\mathbf{x} = 1$. For every $\varepsilon > 0$, if*

$$\|\mathbf{y}^0 - \mathbf{y}^d\|_1 \leq \varepsilon, \tag{3.80}$$

then the solution $\mathbf{y}(\cdot,t)$ of the system (3.56) satisfies

$$\|\mathbf{y}(\cdot,t) - \mathbf{y}^d\|_1 \leq 2\varepsilon \tag{3.81}$$

for all $t \geq 0$.

Proof. We know that the solution y satisfies $\int_\Omega \mathbf{y}(\cdot,)d\mathbf{x} = \int_\Omega y_1(\cdot,t)d\mathbf{x} + \int_\Omega y_2(\cdot,t)d\mathbf{x} = 1$ for all $t \in [0,T]$. From Proposition 3.6.15, we know that $\|y_2(\cdot,t) - y^d\|_1$ is non-decreasing with time t. Hence, $\|y_2(\cdot,t) - y^d\|_1 \leq \varepsilon$ for all $t \geq 0$. Then we have,

$$\int_\Omega y_1(\mathbf{x},t)d\mathbf{x} + \int_\Omega (y_2(\mathbf{x},t) - y^d(\mathbf{x}))d\mathbf{x} = 1 - \int_\Omega y^d(\mathbf{x})d\mathbf{x} \tag{3.82}$$

for all $t \geq 0$. This implies that

$$\begin{aligned} \int_\Omega y_1(\mathbf{x},t)d\mathbf{x} &\leq -\int_\Omega (y_2(\mathbf{x},t) - y^d(\mathbf{x}))d\mathbf{x} \\ &\leq \|y_2(\cdot,t) - y^d\|_1 \\ &\leq \varepsilon \end{aligned}$$

for all $t \geq 0$. This concludes the proof. □

Proposition 3.6.17. *Let* $\mathbf{y}^0 \in \mathbf{L}^\infty(\Omega)$ *be non-negative and* $\|\mathbf{y}^0\|_1 = 1$. *Then the solution* $\mathbf{y}$ *of the PDE* (3.56) *satisfies* $\lim_{t\to\infty} \|(y_1(\cdot,t) - y^d)_+\|_\infty = 0$.

Proof. Suppose that, for the sake of contradiction, this is not true. Then, due to the partial monotonicity property of the solution $\mathbf{y}$ as stated in Proposition 3.6.15, there exists a subset $\Omega_1 \subseteq \Omega$ of positive measure, and a parameter $\varepsilon > 0$, such that $y_1(\mathbf{x},t) - y^d(\mathbf{x}) \geq \varepsilon$ for almost every $\mathbf{x} \in \Omega_1$ and all $t \geq 0$. However, we know that

$$\begin{aligned} y_2(\mathbf{x},t) &= y_2(\mathbf{x},t_1) - \int_s^t F^2(y_2(\mathbf{x},\tau))y_2(\mathbf{x},\tau)d\tau \\ &= y_2(\mathbf{x},t_1) - \int_s^t r^2(y_2(\mathbf{x},\tau) - y^d(\mathbf{x}))y_2(\mathbf{x},\tau)d\tau \end{aligned} \tag{3.83}$$

for almost every $\mathbf{x} \in \Omega_1$ and for all $t \geq 0$. We know that the function r_2 is non-zero and continuous on the open interval $(0,\infty)$. Hence, there must exist $\delta > 0$ such that

$$\begin{aligned} y_2(\mathbf{x},t) &\leq y_2(\mathbf{x},0) - \int_0^t \delta y_2(\mathbf{x},\tau)d\tau \\ &\leq y_2(\mathbf{x},0) - \delta \int_0^t (y^d(\mathbf{x}) + \varepsilon)d\tau \end{aligned} \tag{3.84}$$

for almost every $\mathbf{x} \in \Omega_1$ and for all $t \geq 0$. This leads to a contradiction. □

Finally, we can establish the attractivity of the equilibrium point $\mathbf{y}^d \in L^\infty(\Omega)$. We prove a preliminary lemma toward this end.

Lemma 3.6.18. *Let* $\mathbf{y}^0 \in L^\infty(\Omega)$ *be non-negative and* $\|\mathbf{y}^0\|_1 = 1$. *Then the solution* $\mathbf{y}$ *of the PDE* (3.56) *satisfies* $\lim_{t\to\infty} \|y_1(\cdot,t)\|_1 = 0$. *Hence,* $\lim_{t\to\infty} \|y_2(\cdot,t)\|_1 = 1$.

Proof. Suppose that this statement is not true. Then there exists $\varepsilon_1 > 0$ and a sequence of increasing time instants $(t_i)_{i=1}^\infty$ such that $\lim_{i\to\infty} t_i = \infty$ and $\|y_1(\cdot,t_i)\|_1 \geq \varepsilon_1$ for all $i \in \mathbb{Z}_+$. From Theorem 3.6.13, we know that this implies that there exist $\tau, \varepsilon_2, \delta > 0$ such that $y_1(\cdot,t) \geq \varepsilon_2 \|y_1^0\|_1 \geq \varepsilon_1\varepsilon_2$ for all $t \geq [t_i, t_i + \delta]$, for all $i \in \mathbb{Z}_+$. Without loss of generality, we can assume that $t_{i+1} - t_i > \delta$ for all $i \in \mathbb{Z}_+$. Let $\Omega_1 \subseteq \Omega$ be the subset of largest measure

such that $y_2^0(\mathbf{x}) \geq y^d(\mathbf{x})$ for all $\mathbf{x} \in \Omega_1$. Then, from the partial monotonicity property of the solution $\mathbf{y}$ (Proposition 3.6.15), we have that, for each $i \in \mathbb{Z}_+$,

$$\begin{aligned} y_2(\mathbf{x}, t_i + \delta) &= y_2(\mathbf{x}, 0) + \int_0^{t_i+\delta} F_1(y_2(\mathbf{x}, \tau)) y_1(\mathbf{x}, \tau) d\tau \\ &\geq y_2(\mathbf{x}, 0) + \sum_{j=1}^{i} \int_{t_i}^{t_i+\delta} r_1(y_2(\mathbf{x}, \tau) - y^d(\mathbf{x})) y_1(\mathbf{x}, \tau) d\tau \end{aligned} \tag{3.85}$$

for almost every $\mathbf{x} \in \Omega_1$. This implies that $\lim_{i\to\infty} \|(y_2(\cdot, t_i) - y^d)_-\|_\infty = 0$. However, we know that $\|\mathbf{y}(\cdot, t)\|_1 = 1$ for all $t \geq 0$. From this, along with the fact that $\lim_{t\to\infty} \|(y_1(\cdot, t) - y^d)_+\|_\infty = 0$ (Lemma 3.6.18) and the assumption that $\|y_1(\cdot, t_i)\|_1 \geq \varepsilon_1$ for all $i \in \mathbb{Z}_+$, we arrive at a contradiction. □

Theorem 3.6.19. ***(L^1-Global Attractivity)*** *Let $\mathbf{y}^0 \in L^\infty(\Omega)$ be non-negative and $\|\mathbf{y}^0\|_1 = 1$. Then $\lim_{t\to\infty} \|\mathbf{y}(\cdot, t) - \mathbf{y}^d\|_1 = 0$.*

Proof. Let $\Omega_1 = \{\mathbf{x} \in \Omega; y_2^0(\mathbf{x}) \geq y^d(\mathbf{x})\}$. Let $\Omega_2 = \Omega - \Omega_1$. From Lemma 3.6.18, we know that $\lim_{t\to\infty} \|y_2(\cdot, t)|_{\Omega_1} - y^d|_{\Omega_1}\|_\infty = 0$, where $\cdot|_{\Omega_1}$ denotes the restriction operation. This implies that $\lim_{t\to\infty} \|y_2(\cdot, t)|_{\Omega_1} - y^d|_{\Omega_1}\|_1 = 0$. From Lemma 3.6.18, we know that

$$\lim_{t\to\infty} \int_{\Omega_1} \big(y_2(\mathbf{x}, t) - y^d(\mathbf{x})\big) d\mathbf{x} + \int_{\Omega_2} \big(y_2(\mathbf{x}, t) - y^d(\mathbf{x})\big) d\mathbf{x} = 0 \tag{3.86}$$

This implies that

$$\lim_{t\to\infty} \int_{\Omega_2} \big(y_2(\mathbf{x}, t) - y^d(\mathbf{x})\big) d\mathbf{x} = 0 \tag{3.87}$$

From Proposition 3.6.15, we know that $y_2(\cdot, t) \leq y^d$ almost everywhere on Ω_2 and for all $t \geq 0$. Therefore, $\lim_{t\to\infty} \int_{\Omega_2} |y_2(\mathbf{x}, t) - y^d(\mathbf{x})| d\mathbf{x} = 0$. Hence, we can conclude that

$$\lim_{t\to\infty} \|y_2(\cdot, t) - y^d\|_1 = \lim_{t\to\infty} \int_{\Omega_1} |y_2(\mathbf{x}, t) - y^d(\mathbf{x})| d\mathbf{x} + \int_{\Omega_2} |y_2(\mathbf{x}, t) - y^d(\mathbf{x})| d\mathbf{x} = 0 \tag{3.88}$$

From this equation, along with the fact that $\lim_{t\to\infty} \|y_1(\cdot, t)\|_1 = 0$, we arrive at our result.

□

Chapter 4

CONTROLLABILITY AND OPTIMAL CONTROL OF DISCRETE-TIME NONLINEAR SYSTEMS TO TARGET MEASURES

In this chapter, we consider a variation of the *optimal transport problem* (Villani, 2008). The objective of this problem is to construct a map such that a given probability measure is *pushed forward* to a target probability measure in some optimal manner. Initially motivated by resource allocation problems in economics, this problem has potential applications in many engineering problems involving the control of large-scale distributed systems (Djehiche *et al.*, 2016) using mean-field models, in which these measures could represent the distribution of an ensemble of agents such as a swarm of robots or the distribution of nodes in an electric power grid (Bagagiolo and Bauso, 2014) or a wireless network (Tembine, 2014).

In the original formulation of optimal transport, the dynamics of the agents are simplistic from a control-theoretic point of view. There have been some recent efforts to extend classical optimal transport theory to the case where the cost functions and transport maps are subject to dynamical constraints arising from control systems. Toward this end, (Hindawi *et al.*, 2011) considers the optimal transport problem for linear time-invariant systems with linear quadratic cost functions. For a smaller class of cost functions, the case of linear time-varying systems is addressed in (Chen *et al.*, 2017). There have also been efforts to extend the theory to nonlinear driftless control-affine systems in the framework of *sub-Riemannian optimal transport* (Agrachev and Lee, 2009; Figalli and Rifford, 2010; Khesin *et al.*, 2009).

The original optimal transport problem, i.e., the *Monge problem*, searches for a deterministic map that maps a given measure to a target measure. In view of the analytical

difficulties involved in this original formulation of Monge, Kantorovich introduced a relaxed version of the problem in 1942, in which the map is allowed to be stochastic. This form of relaxation, which is used to convexify nonlinear control problems, has a rich history in control theory in the context of *Young measures* or *relaxed control* (Florescu and Godet-Thobie, 2012; Young, 1980). In recent years, such a measure-based convexification of optimization problems has been used for numerical synthesis of control laws (Lasserre *et al.*, 2008; Vaidya *et al.*, 2010).

In this chapter, we use a similar relaxation procedure to consider the optimal transport problem for discrete-time nonlinear control systems with a compact set of admissible controls. Before considering the issue of optimality, we consider the problem of controllability. First, we prove that controllability of the original control system implies controllability of the control system induced on the space of probability measures. Next, we show that we can frame the control-constrained optimal transport problem of controllable nonlinear systems as a linear programming problem, as in the Kantorovich formulation of the optimal transport problem. Such a linear programming based approach to solving optimal control problems is classical in optimal control of discrete-time stochastic systems, also known as Markov decision problems (Hernández-Lerma and Lasserre, 2012).

4.1 Notation

Let X be a finite-dimensional manifold (for example, the Euclidean space $\mathbb{R}^M$) equipped with a metric. The set of admissible control inputs will be denoted by U. We will assume that the set U is a compact subset of a metric space. We will denote by $\mathscr{B}(X)$, $\mathscr{B}(U)$, and $\mathscr{B}(X \times U)$ the collection of Borel measurable sets of X, U, and $X \times U$, respectively. The space of Borel probability measures on the sets X and U will be denoted by $\mathscr{P}(X)$ and $\mathscr{P}(U)$, respectively. For a metric space Y, let $C_b(Y)$ be the set of bounded continuous functions on Y. We will say that a sequence of measures $(\mu_n)_{n=1}^{\infty} \in \mathscr{P}(Y)$ con-

verges narrowly to a limit measure $\mu \in \mathscr{P}(Y)$ if the sequence $\int_Y f(\mathbf{y})d\mu_n(\mathbf{y})$ converges to $\int_Y f(\mathbf{y})d\mu(\mathbf{y})$ for every $f \in C_b(Y)$. The topology on $\mathscr{P}(Y)$ corresponding to this convergence will be referred to as the narrow topology. For a set $M \subset X$ and $p \in \mathbb{Z}_+$, we will define the set $D_M^p = \left\{ \sum_{i=1}^p c_i \delta_{\mathbf{y}_i};\ \mathbf{y}_i \in M,\ c_i \in [0,1] \text{ for } i \in \{1,...,p\},\ \sum_{i=1}^p c_i = 1 \right\}$, where $\delta_{\mathbf{x}}$ is the Dirac measure concentrated at the point $\mathbf{x} \in X$. We will also define the set $D_M = \cup_{p \in \mathbb{Z}_+} D_M^p$. The support of a measure $\mu \in \mathscr{P}(X)$ will be denoted by supp $\mu = \{\mathbf{x} \in X;\ \mathbf{x} \in N_{\mathbf{x}}$ implies that $\mu(N_{\mathbf{x}}) > 0,$ where $N_{\mathbf{x}}$ is a neighborhood of $\mathbf{x}\}$. We define $\mathscr{Y}(X,U)$ as the set of stochastic feedback laws, i.e., maps of the form $K : X \times \mathscr{B}(U) \to \mathbb{R}$, where $K(\cdot,A)$ is Borel measurable for each $A \in \mathscr{B}(U)$ and $K(\mathbf{x},\cdot) \in \mathscr{P}(U)$ for each $\mathbf{x} \in X$. For a continuous map $F : Y \to X$, the pushforward map $F_\# : \mathscr{P}(Y) \to \mathscr{P}(X)$ is defined by

$$(F_\#\mu)(A) = \mu(F^{-1}(A)) = \int_Y \mathbf{1}_A(F(\mathbf{y}))d\mu(\mathbf{y})$$

for each $A \in \mathscr{B}(X)$, where $\mathbf{1}_B$ denotes the indicator function of the set $B \in \mathscr{B}(X)$ and $\mu \in \mathscr{P}(Y)$.

Problem Formulation

Now we are ready to state the problems addressed in this section. Consider the nonlinear discrete-time control system

$$\begin{aligned} &\mathbf{x}_{n+1} = T(\mathbf{x}_n, \mathbf{u}_n),\ n = 0,1,... \\ &\mathbf{x}_0 \in X, \end{aligned} \tag{4.1}$$

where $\mathbf{x}_n \in X$ for each $n \in \mathbb{Z}_+$, $(\mathbf{u}_i)_{i=0}^\infty$ is a sequence in a compact set U, and $T : X \times U \to X$ is a continuous map with respect to the topologies $\mathscr{T}(X)$, $\mathscr{T}(U)$, and $\mathscr{T}(X) \times \mathscr{T}(U)$ defined on X, U, and $X \times U$, respectively. Then this nonlinear control system induces a

control system on the space of measures $\mathscr{P}(X)$, given by

$$\mu_{n+1} = T(\cdot, \mathbf{u}_n)_{\#}\mu_n,\ n = 0, 1, ...$$
$$\mu_0 \in \mathscr{P}(X). \tag{4.2}$$

The first problem of interest is the following.

Problem 4.1.1. ***(Controllability problem with deterministic control)*** *Let $N \in \mathbb{Z}_+$ be a specified final time. Given an initial measure $\mu_0 \in \mathscr{P}(X)$ and a target measure $\mu^f \in \mathscr{P}(X)$, does there exist a sequence of feedback laws $\mathbf{v}_n : X \to U$ such that the closed-loop system satisfies*

$$\mu_{n+1} = T_{\#}^{cl,n}\mu_n,\ \ n = 0, 1, ..., N-1,$$
$$\mu_N = \mu^f,$$

where $T_{\#}^{cl,n} : \mathscr{P}(X) \to \mathscr{P}(X)$ is the pushforward map corresponding to the closed-loop map $T^{cl,n} : X \to X$ defined by $T^{cl,n}(\mathbf{x}) = T(\mathbf{x}, \mathbf{v}_n(\mathbf{x}))$ for all $\mathbf{x} \in X$?

This problem is unsolvable in general. For instance, consider the case when $X = \mathbb{R}$, $U = [-1, 1]$, $T(\mathbf{x}, \mathbf{u}) = \mathbf{x} + \mathbf{u}$ for each $(\mathbf{x}, \mathbf{u}) \in X \times U$, $\mu_0 = \delta_0$ is the Dirac measure concentrated at the point $0 \in \mathbb{R}$, and $\mu^f = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_{+1}$ is the sum of Dirac measures concentrated at -1 and 1, respectively. This example does not admit any solutions to the controllability problem because a deterministic map cannot take the measure concentrated at the point 0 and distribute it onto measures concentrated at -1 and $+1$. However, there might be several important cases where the problem does admit a solution. For example, when $X = \mathbb{R}^M$, $U = \mathbb{R}^M$ (which is not compact, in contrast to the assumptions made in this section), $T(\mathbf{x}, \mathbf{u}) = \mathbf{u}$ for all $(\mathbf{x}, \mathbf{u}) \in X \times U$, and $N = 1$, this problem is equivalent to the classical optimal transport problem (Villani, 2008), for which solutions are known to exist when the initial and final measures are absolutely continuous with respect to the Lebesgue

measure and have a finite second moment. On the other hand, this problem is expected to be highly challenging for general nonlinear control systems without any further constraints on the control set U, which is only assumed to be compact, given a final time $N \geq 1$. Hence, to make the problem analytically tractable, we consider the following relaxed problem.

Problem 4.1.2. *(**Controllability problem with stochastic control**) Given a final time $N \in \mathbb{Z}_+$, an initial measure $\mu_0 \in \mathscr{P}(X)$, and a target measure $\mu^f \in \mathscr{P}(X)$, determine whether there exists a sequence of stochastic feedback laws $K_n \in \mathscr{Y}(X,U)$ such that the closed-loop system satisfies*

$$\begin{aligned} \mu_{n+1} &= T_{\#}^{cl,n}\mu_n, \;\; n = 0,1,...,N-1, \\ \mu_N &= \mu^f, \end{aligned} \tag{4.3}$$

where the closed-loop pushforward map $T_{\#}^{cl,n}$ is given by

$$(T_{\#}^{cl,n}\mu)(A) = \int_X \int_U \mathbf{1}_A(T(\mathbf{x},\mathbf{u}))K_n(\mathbf{x},d\mathbf{u})d\mu(\mathbf{x}). \tag{4.4}$$

Problem 4.1.2 can be considered a relaxation of Problem 4.1.1 in the sense that deterministic control laws $\mathbf{v} : X \to U$ are just special types of stochastic control laws identified through the mapping $\mathbf{v}(\mathbf{x}) \mapsto \delta_{\mathbf{v}(\mathbf{x})}$.

After addressing Problem 4.1.2, we will address the following optimization problem.

Problem 4.1.3. *(**Fixed-time, fixed-endpoint optimal control problem**) Suppose that $c : X \times U \to \mathbb{R}$ is a continuous map. Given a final time $N \in \mathbb{Z}_+$, an initial measure $\mu_0 \in \mathscr{P}(X)$, and a target measure $\mu^f \in \mathscr{P}(X)$, determine whether the following optimization problem admits a solution:*

$$\min_{\substack{\mu_m \in \mathscr{P}(X) \\ K_m \in \mathscr{Y}(X,U)}} \sum_{m=0}^{N-1} \int_X \int_U c(\mathbf{x},\mathbf{u})K_m(\mathbf{x},d\mathbf{u})d\mu_m(\mathbf{x}) \tag{4.5}$$

subject to the constraints

$$\mu_{n+1} = T_{\#}^{cl,n}\mu_n, \;\; n = 0,1,...,N-1,$$
$$\mu_N = \mu^f. \tag{4.6}$$

4.2 Controllability

In this section, we will address Problem 4.1.2. Toward this end, we present the following definitions, which will be needed to define sufficient conditions under which Problem 4.1.2 admits a solution. Let $R_1^{\mathbf{x}} = \{T(\mathbf{x},\mathbf{u});\ \mathbf{u} \in U\}$ be the set of reachable states from $\mathbf{x} \in X$ at the first time step. Then we inductively define the set $R_m^{\mathbf{x}} = \cup_{\mathbf{y} \in R_{m-1}^{\mathbf{x}}} \{T(\mathbf{y},\mathbf{u});\ \mathbf{u} \in U\}$ for each $m \in \mathbb{Z}_+ - \{1\}$.

Instead of proving that we can always find a sequence of stochastic feedback laws K_n such that the system of equations (4.3) is satisfied, we will consider the alternative "convexified problem" in which we look for measures ν_n on the product space $\mathscr{P}(X \times U)$ such that, for given initial and target measures $\mu_0, \mu^f \in \mathscr{P}(X)$, the following constraints are satisfied:

$$\mu_{n+1} = T_{\#}\nu_n, \;\; n = 0,1,...,N-1, \tag{4.7}$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathscr{B}(X)$ and $\mu_N = \mu^f$. We will first solve Problem 4.1.2 for the special case of Dirac measures, and then extend the result to general measures using a density-based argument that is standard in measure-theoretic probability.

Now we are ready to present several results that address Problem 4.1.2.

Proposition 4.2.1. *Let $\mu_0 = \delta_{\mathbf{x}_0}$ for some $\mathbf{x}_0 \in X$. Let $\mu^f \in D_M^p$ for a compact subset M of X, for some $p \in \mathbb{Z}_+$, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}_0}$. *Then there exists a sequence of measures $(\nu_m)_{m=0}^{N-1} \in \mathscr{P}(X \times U)$ such that*

$$\mu_{n+1} = T_{\#}\nu_n, \;\; n = 0,1,...,N-1, \tag{4.8}$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathscr{B}(X)$ and $\mu_N = \mu^f$.

Proof. Let $\mu^f = \sum_{i=1}^p c^i \delta_{\mathbf{y}^i}$, where $\sum_{i=1}^p c^i = 1$, for some $\mathbf{y}^i \in X$. By assumption, supp $\mu^f \subseteq R_N^{\mathbf{x}_0}$. Hence, for each $i \in \{1,...,p\}$, there exists a sequence of inputs $(\mathbf{u}^i)_{n=0}^N$ such that the nonlinear discrete-time control system

$$\begin{aligned} \mathbf{x}_{n+1}^i &= T(\mathbf{x}_n^i, \mathbf{u}_n^i), \ \ n = 0,1,...,N-1, \\ \mathbf{x}_0^i &= \mathbf{x}^0 \end{aligned} \tag{4.9}$$

satisfies $\mathbf{x}_N = \mathbf{y}^i$ for all $i \in \{1,...,p\}$. We define $\nu_n^i = \delta_{(\mathbf{x}_{n-1}^i, \mathbf{u}_n^i)} \in \mathscr{P}(X \times U)$. Note that $(T_\# \nu_n^i)(A) = \delta_{\mathbf{x}_n^i}(A)$ for all $A \in \mathscr{B}(X)$ and all $i \in \{1,...,p\}$. Then the result follows from the linearity of the operator $T_\# : \mathscr{P}(X \times U) \to \mathscr{P}(X)$ by setting $\nu_n = \sum_{i=1}^p c^i \nu_n^i$ for all $n \in \{0,1,...,N-1\}$. In particular, for this choice of ν_n, we have that $(T_\# \nu_n) = \sum_{i=1}^p c^i \mu_{n+1}^i$ for each $n \in \{0,1,...,N-1\}$, and hence that $(T_\# \nu_{N-1}) = \sum_{i=1}^p c^i \delta_{\mathbf{y}^i} = \mu^f$. □

The next result follows immediately from Proposition 4.2.1.

Lemma 4.2.2. *Let $\mu_0 \in D_A^p$ and $\mu^f \in D_A^q$ for a compact subset A of X, for some $p,q \in \mathbb{Z}_+$, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}}$ *for each* $\mathbf{x} \in$ supp μ_0. *Then there exists a sequence of measures $(\nu_m)_{m=0}^{N-1} \in \mathscr{P}(X \times U)$ such that*

$$\mu_{n+1} = T_\# \nu_n, \ \ n = 0,1,...,N-1, \tag{4.10}$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathscr{B}(X)$, and $\mu_N = \mu^f$.

Proof. Let $\mu_0 = \sum_{i=1}^p c^i \delta_{\mathbf{y}^i}$, where $\sum_{i=1}^p c^i = 1$, for some $\mathbf{y}^i \in X$. By assumption, supp $\mu^f \subseteq \cup_{i=1}^p R_N^{\mathbf{y}^i}$. From Proposition 4.2.1, there exist measures $\nu_n^i \in \mathscr{P}(X \times U)$ such that if $\eta_0^i = \mu_0$, then

$$\eta_{n+1}^i = T_\# \nu_n^i, \ \ n = 0,1,...,N-1, \tag{4.11}$$

with $\nu_n^i(A \times U) = \eta_n^i(A)$ for all $A \in \mathscr{B}(X)$, and $\eta_N^i = \mu^f$. The result follows by setting $\nu_n = \sum_{i=1}^p c^i \nu_n^i$ for all $n \in \{0,1,...,N-1\}$. □

In order to prove the next proposition, we recall a well-known result, which follows from (Pedersen, 2012)[Proposition 2.5.7], that probability measures can be approximated using linear combinations of Dirac measures.

Theorem 4.2.3. *Let Y be a locally compact Hausdorf space Y. Then the set of elements in $\mathscr{P}(Y)$ with support contained in a compact subset $M \subseteq Y$ is a convex and narrowly compact subset of $\mathscr{P}(Y)$. Additionally, the set D_M is narrowly dense in the subset of $\mathscr{P}(Y)$ with supports contained in M.*

Proposition 4.2.4. *Let $\mu_0, \mu^f \in \mathscr{P}(X)$ be Borel probability measures with compact supports, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}}$ *for each* $\mathbf{x} \in$ supp μ_0. *Then there exists a sequence of measures $(\nu_m)_{m=0}^{N-1} \in \mathscr{P}(X \times U)$ such that*

$$\mu_{n+1} = T_{\#}\nu_n, \;\; n = 0, 1, ..., N-1, \tag{4.12}$$

with $\nu_n(A \times U) = \mu_n(A)$ for all $A \in \mathscr{B}(X)$, and $\mu_N = \mu^f$.

Proof. Let $A = \cup_{\mathbf{x} \in \text{supp } \mu_0} R_m^{\mathbf{x}}$. Clearly, the set A is compact. From Theorem 4.2.3, we know that there exist sequences of measures $(\mu_0^i)_{i=1}^{\infty}, (\mu^{f,i})_{i=1}^{n\infty} \in D_A$ such that $(\mu_0^i)_{i=1}^{\infty}$ and $(\mu^{f,i})_{i=1}^{\infty}$ narrowly converge to μ_0 and μ^f, respectively. Then it follows from Lemma 4.2.2 that there exists a sequence of probability measures $(\nu_n^i)_{i=1}^{\infty}$ in $\mathscr{P}(X \times U)$ such that

$$\mu_{n+1}^i = T_{\#}\nu_n^i, \;\; n = 0, 1, ..., N-1, \tag{4.13}$$

with $\nu_n^i(A \times U) = \mu_n^i(A)$ for all $A \in \mathscr{B}(X)$ and $\mu_N^i = \mu^{f,i}$ for all $i \in \mathbb{Z}_+$. Since the map $T : X \times U \to X$ is continuous, the map $T_{\#}$ is narrowly continuous. This imples that, for each $n \in \{0, 1, ..., N-1\}$, there exists a limit measure $\nu_n \in \mathscr{P}(X \times U)$ such that $T_{\#}\nu_n^i$ narrowly converges to a unique limit $T_{\#}\nu_n$ as $i \to \infty$. Using the fact that the map $T_{\#} : \mathscr{P}(X \times U)$ is narrowly continuous, the last statement also implies that the sequence of marginal measures $\nu_n^i(\cdot \times U) = \mu_n^i$ narrowly converges to the unique limit μ_n for each $n \in \{0, 1, ..., N-1\}$. $\square$

From the above proposition, we obtain one of the main results of this section.

Theorem 4.2.5. *Let $\mu_0, \mu^f \in \mathscr{P}(X)$ be Borel probability measures with compact supports, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}}$ *for each* $\mathbf{x} \in$ supp μ_0. *Then there exists a sequence of stochastic feedback laws $(K_n)_{m=1}^{N-1} \in \mathscr{Y}(X,U)$ such that the system of equations* (4.3) *is satisfied, and hence the measure μ^f can be reached from the measure μ_0.*

Proof. Note that X and U are separable. Hence, the product σ-algebra on $X \times U$ is equal to $\mathscr{B}(X \times U)$. Then, given a measure $\nu \in \mathscr{P}(X \times Y)$, from the *disintegration theorem* (Florescu and Godet-Thobie, 2012)[Theorem 3.2] there exists a measure $\mu \in \mathscr{P}(X)$ and stochastic feedback law $K \in \mathscr{Y}(X,U)$ such that

$$\int_{A\times B} d\nu(\mathbf{x},\mathbf{u}) = \int_A \int_B K(\mathbf{x}, d\mathbf{u}) d\mu(\mathbf{x}) \tag{4.14}$$

for all $A \in \mathscr{B}(X)$ and all $B \in \mathscr{B}(U)$. Then the result follows from Proposition 4.2.4. In particular, using the measures $(\nu_m)_{m=0}^{N-1} \in \mathscr{P}(X \times U)$, by disintegration, the stochastic feedback laws $(K_m)_{m=0}^{N-1} \in \mathscr{Y}(X,U)$ can be constructed such that the system of equations (4.3) holds true. □

Remark 4.2.6. ***(Conservatism of controllability result)*** *Theorem 4.2.5 gives a sufficient, but not necessary, condition on system* (4.1) *for Problem 4.1.2 to admit a solution: namely, that each point in the support of the target measure be reachable from each point in the support of the initial measure. The controllability result in Theorem 4.2.5 is conservative because we do not, in general, require this condition. To see this explicitly, consider the trivial example where $X = \mathbb{R}$, $U = \{0\}$, and $T(x,u) = x + u$. Suppose we define the initial and target measures as $\mu_0 = \mu^f = \frac{1}{2}\delta_{x_1} + \frac{1}{2}\delta_{x_2}$ for some $x_1 \neq x_2$ in $\mathbb{R}$. Then it is straightforward to see that the target measure is reachable from the initial measure. However, the system is nowhere controllable in $\mathbb{R}$. More specifically, the points x_1 and x_2 are not reachable from each other.*

4.3 Optimal Control

This section addresses Problem 4.1.3. As in the proof of the controllability result in Theorem 4.2.5, we will apply the disintegration theorem (Florescu and Godet-Thobie, 2012)[Theorem 3.2] to the correspondence between elements of $\mathscr{Y}(X,U)$ and elements of $\mathscr{P}(X\times U)$ with a given marginal. Hence, the optimization problem (4.5)-(4.6) can be convexified by replacing stochastic feedback laws $K_n \in \mathscr{Y}(X,U)$ with elements $\nu_n \in \mathscr{P}(X\times U)$ and by enforcing appropriate constraints on the marginals of the measures ν_n. These modifications allow us to frame the optimization problem in Problem 4.1.3 as an equivalent infinite-dimensional linear programming problem:

$$\min_{\substack{\mu_{m+1}\in\mathscr{P}(X),\\ \nu_m\in\mathscr{P}(X\times U)}} \sum_{m=0}^{N-1}\int_{X\times U} c(\mathbf{x},\mathbf{u})d\nu_m(\mathbf{x},\mathbf{u}) \tag{4.15}$$

subject to the constraints

$$\begin{aligned}\mu_{n+1} &= T_{\#}\mu_n,\ \ n=0,1,...,N-1,\\ \mu_N &= \mu^f,\\ \pi_{\#}\nu_n &= \mu_n,\end{aligned} \tag{4.16}$$

where $\pi : X\times U \to X$ is the projection map defined by $\pi(\mathbf{x},\mathbf{u}) = \mathbf{x}$ for all $\mathbf{x}\in X$ and all $\mathbf{u}\in U$. Here, the constraints $\pi_{\#}\nu_n = \mu_n$ ensure that, for each $n\in\{1,...,N\}$, $\nu_n(A\times U) = (\pi_{\#}\nu_n)(A) = \mu_{n-1}(A)$ for all $A\in\mathscr{B}(X)$. Hence, we have the following result.

Theorem 4.3.1. *Let $\mu_0,\mu^f\in\mathscr{P}(X)$ be Borel probability measures with compact supports, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}}$ *for each* $\mathbf{x}\in$ supp μ_0. *Then the optimization problem* (4.15)-(4.16) *has a solution* (μ_{n+1},ν_n), $n=0,...,N-1$.

Proof. The proof follows the standard compactness-based arguments in optimization. From Theorem 4.2.5, we know that the set of measures satisfying constraints (4.16) is non-empty. Moreover, the map $c : X\times U\to\mathbb{R}$ is continuous. Since T is continuous, measures with

compact support are pushed forward to measures with compact support. This implies that for any choice of measure ν_n, supp μ_{n+1} is contained in a compact set since supp μ_0 is contained in a compact set. Therefore, $\sum_{m=0}^{N-1} \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\nu_m(\mathbf{x},\mathbf{u})$ is bounded from below on the set of admissible measures. Hence, there exists a minimizing sequence of measures $(\mu_{n+1}^i, \nu_n^i)_{i=1}^{\infty}$, with $(\mu_{n+1}^i, \nu_n^i) \in \mathscr{P}(X) \times \mathscr{P}(X\times U)$ for each $n \in \{0,1,...,N-1\}$, that satisfies the constraints (4.16). By *minimizing*, we mean that the sequence of measures $(\mu_{n+1}^i, \nu_n^i)_{i=1}^{\infty}$

$$\lim_{i\to\infty} \sum_{m=0}^{N-1} \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\nu_m^i(\mathbf{x},\mathbf{u}) = \inf_{\mu_{m+1}\in\mathscr{P}(X),\ \nu_m\in\mathscr{P}(X\times U)} \ \sum_{m=0}^{N-1} \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\nu_m(\mathbf{x},\mathbf{u}), \tag{4.17}$$

with the infimum taken over the constraint set (4.16). We now confirm that there exist measures (μ_{n+1}^*, ν_n^*) that achieve this infimum. We recall that the support of the measures (μ_{n+1}, ν_n) is compact for all $n \in \{0,1,...,N-1\}$. Therefore, it trivially follows that there exists a compact set Q such that $\mu_{n+1}(Q) > 1-\varepsilon$ and $\nu_n(Q\times U) > 1-\varepsilon$. This implies that the set of measures that satisfy the constraints (4.16) is *tight* (Billingsley, 2013), and therefore is relatively compact, i.e, every sequence of measures (μ_{n+1}^i, ν_n^i) contains a narrowly converging subsequence (μ_{n+1}^j, ν_n^j). The map $\gamma \mapsto \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\gamma(\mathbf{x},\mathbf{u})$, a map from $\mathscr{P}(X\times U)$ to $\mathbb{R}$, is narrowly continuous. Hence, there exist limit measures (μ_{n+1}^*, ν_n^*) such that $\sum_{m=0}^{N} \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\nu_m^*(\mathbf{x},\mathbf{u}) = \inf_{\mu_{m+1}\in\mathscr{P}(X),\ \nu_m\in\mathscr{P}(X\times U)} \ \sum_{m=0}^{N-1} \int_{X\times U} c(\mathbf{x},\mathbf{u}) d\nu_m(\mathbf{x},\mathbf{u})$, subject to the constraints (4.16). This concludes the proof. □

By disintegration of the measures ν_m in Theorem 4.3.1, it is straightforward to conclude the following result.

Theorem 4.3.2. *Let $\mu_0, \mu^f \in \mathscr{P}(X)$ be Borel probability measures with compact supports, such that* supp $\mu^f \subseteq R_N^{\mathbf{x}}$ *for each* $\mathbf{x} \in$ supp μ_0. *Then the optimization problem in Problem 4.1.3 has a solution* $(\mu_{n+1}, K_n),\ n = 0,...,N-1$.

4.4 Numerical Optimization

In this section, we briefly describe a numerical approach to solving the optimization problem in Problem 4.1.3. In both the examples that we consider in Section 4.5, the state space X is taken to be a compact subset of $\mathbb{R}^2$. This subset X is partitioned into $n_x \in \mathbb{Z}_+$ sets, $\tilde{X} = \{\Omega_1, ..., \Omega_{n_x}\}$, whose union is X and whose intersections have zero Lesbesgue measure. The set of control inputs U is approximated as a set of $n_u \in \mathbb{Z}_+$ discrete elements, $\tilde{U} = \{\gamma_1, ..., \gamma_{n_u}\}$, where $\gamma_i \in U$ for each i. We then use the *Ulam-Galerkin method* (Bollt and Santitissadeekorn, 2013) to construct an approximating controlled Markov chain on a finite state space $V = \{1, ..., n_x\}$. In the uncontrolled setting, this method is a classical technique used to construct approximations of pushforward maps induced by dynamical systems, also known as Perron-Frobenius operators.

We define the *controlled* transition probabilities for the Markov chain on V as follows:

$$\tilde{p}_{ij}^k = \frac{\tilde{m}(T_k^{-1}(\Omega_j) \cap \Omega_i)}{\tilde{m}(\Omega_i)},$$

where $\tilde{m}$ is the Lebesgue measure and $T_k = T(\cdot, \gamma_k)$. The quantity $\tilde{p}_{ij}^k$ is the probability of the system state entering the set Ω_j in the next time step, given that this state is uniformly randomly distributed over the set Ω_i (identified with $i \in V$) and the control input is chosen to be γ_k. We also define an equivalent of the stochastic feedback law K_n in the discretized case that we consider. Toward this end, we denote by $\lambda_n^{k,i}$ the probability of choosing the control input γ_k, given that the system state is in Ω_i at time n. We define the variables $\tilde{v}_n^{k,i} = \tilde{\mu}_n^i \lambda_n^{k,i}$, where $\tilde{\mu}_n^i$ is the probability of the state being in Ω_i at time step time n. Additionally, let $\tilde{c}_{i,k} = \int_{\Omega_i} c(\mathbf{x}, \gamma_k) d\mathbf{x}$ be the average cost of the state being in Ω_i and the control input given by γ_k.

Given these parameters and specified initial and target measures $\tilde{\mu}_0, \tilde{\mu}^f \in \mathscr{P}(\tilde{X})$, we can define the finite-dimensional equivalent of the linear programming problem (4.15)-(4.16)

as follows:

$$\min_{\tilde{\mu}^i_{m+1},\tilde{v}^{k,i}_m \in \mathbb{R}_{\geq 0}} \sum_{m=0}^{N-1} \sum_{i=1}^{n_x} \sum_{k=1}^{n_u} \tilde{c}_{i,k} \tilde{v}^{k,i}_m \tag{4.18}$$

subject to the constraints

$$\begin{gathered} \tilde{\mu}^j_{n+1} = \textstyle\sum_{k=1}^{n_u} \sum_{i=1}^{n_x} \tilde{p}^k_{ij} \tilde{v}^{k,i}_n, \\ \tilde{\mu}^j_N = (\tilde{\mu}^f)^j, \\ \textstyle\sum_{i=1}^{n_x} \tilde{\mu}^i_{n+1} = 1, \ \sum_{k=1}^{n_u} \tilde{v}^{k,j}_n = \tilde{\mu}^j_n, \end{gathered} \tag{4.19}$$

for $n \in \{0,...,N-1\}$ and $j \in \{1,...,n_x\}$.

After solving this linear programming problem, we can extract the control laws $\lambda^{k,i}_n$ by setting $\lambda^{k,i}_n = \frac{\tilde{v}^{k,i}_n}{\tilde{\mu}^i_n}$ if $\tilde{\mu}^i_n \neq 0$ and $\lambda^{k,i}_n = 0$ otherwise. The resulting Markov chain evolves according to the equation $\tilde{\mu}^j_{n+1} = \sum_{k=1}^{n_u} \sum_{i=1}^{n_x} \tilde{p}^k_{ij} \lambda^{k,i}_n \tilde{\mu}^i_n$.

4.5 Simulation Examples

In this section, we apply the numerical optimization procedure described in the previous section to two examples. Neither example can be solved by classical optimal transport methods, due to the nonlinearity of the control system (Example 1) or the bounds on the control set (Examples 1 and 2). In both examples, we define the cost function as $c(\mathbf{x},\mathbf{u}) = \|\mathbf{x}\|^2 + \|\mathbf{u}\|^2$, where $\|\cdot\|$ represents the 2-norm.

Example 1: Unicycles in a Time-Periodic Double Gyre We consider the system

$$\mathbf{x}_{n+1} = F(\mathbf{x}_n) + G(\mathbf{u}), \tag{4.20}$$

where $\mathbf{x}_n = [x_n \ y_n]^T \in X$, $\mathbf{u} = [u^1 \ u^2]^T \in U$, and $G(\mathbf{u}) = [u^1 \cos(u^2) \ \ u^1 \sin(u^2)]^T$. The phase space is $X = [0,2] \times [0,1]$, and the set of control inputs is $U = [-1,1] \times [0,2\pi]$. The final

time is set to $N = 10$. To define the map $F : X \to X$, we consider the double-gyre system:

$$\dot{x} = -\pi A \sin(\pi f(x,t)) \cos(\pi y), \tag{4.21}$$

$$\dot{y} = \pi A \cos(\pi f(x,t)) \sin(\pi y) \frac{df(x,t)}{dx}, \tag{4.22}$$

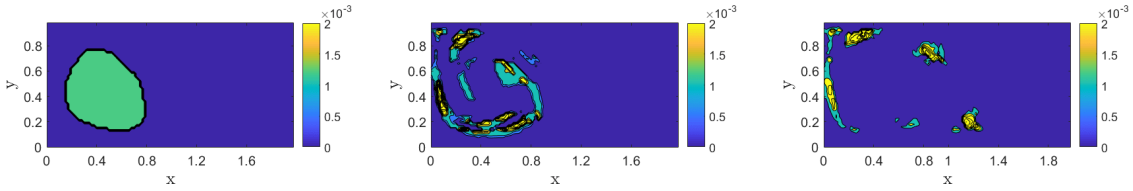
where $f(x,t) = \beta \sin(\omega t)x^2 + (1 - 2\beta \sin(\omega t))x$ is the time-periodic forcing in the system. The map F is defined by setting $F(\mathbf{x})$ equal to the solution of equations (4.21)-(4.22), integrated over the time period τ. In this example, we define $A = 0.25$, $\beta = 0.25$, and $\omega = 2\pi$, which results in $\tau = 1$. The set X is not invariant for all choices of control inputs in U. Hence, since this set must be approximatable by a finite set, we define $F(\mathbf{x}) + G(\mathbf{u}) \triangleq \mathbf{x}$ if $F(\mathbf{x}) + G(\mathbf{u}) \notin X$ for some $(\mathbf{x}, \mathbf{u}) \in X \times U$. The initial and target measures are chosen to be uniform over certain *almost-invariant sets* (Bollt and Santitissadeekorn, 2013) in the left and right halves of the domain, respectively. The optimal transport shown in Fig. 4.1 exploits *lobe dynamics*, i.e., the control inputs push the initial measure onto regions bounded by stable and unstable manifolds. As a result, the measure is transported into the right half of the domain under the action of F.

Example 2: Double-Integrator System In this example, we consider the following system:

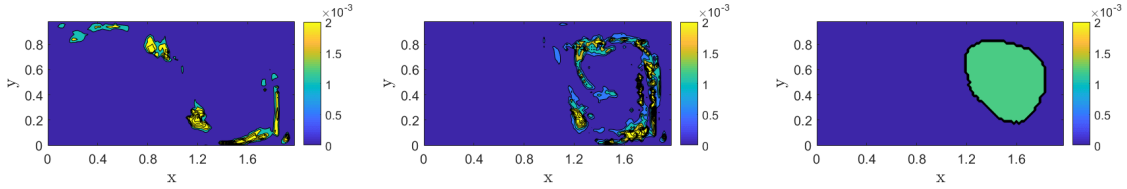
$$x_{n+1} = x_n + 0.15 y_n, \tag{4.23}$$

$$y_{n+1} = y_n + u, \tag{4.24}$$

with $[x_n \; y_n]^T \in X = [0,1]^2$ and $u \in U = [-0.25, 0.25]$. The final time is set to $N = 15$. For unbounded control inputs, this control system can be verified to be globally controllable using the Kalman rank condition. For compact control sets, controllability is harder to verify without numerical computation. The initial measure is taken to be the Dirac measure concentrated at $[0 \; 0]^T \in X$. The target measure is a linear combination of Gaussian distributions that are centered at the coordinates $[0.8 \; 0.1]^T$ and $[0.8 \; 0.8]^T$, as shown in Fig. 4.2d.



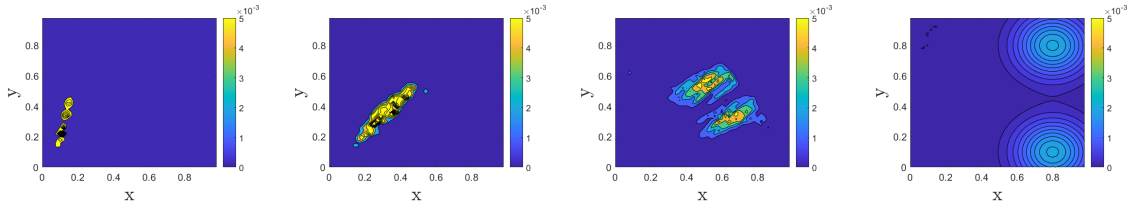
(a) $n = 0$ (Initial Measure) (b) $n = 2$ (c) $n = 4$



(d) $n = 6$ (e) $n = 8$ (f) $n = 10$ (Final Measure)

Figure 4.1: Solution of the Optimal Transport Problem at Several Times n for Unicycles in a Double-Gyre Flow Model

Measures at three intermediate times are shown in Fig. 4.2a-4.2c. The control map adds a "drift" term $0.15y_n$ to equation (4.23), which makes the system controllable despite the fact that it is underactuated. Figure 4.2 confirms that this drift drives the initial measure exactly to the target measure at $N = 15$.



(a) $n = 4$ (b) $n = 8$ (c) $n = 12$ (d) $n = 15$

Figure 4.2: Solution of the Optimal Transport Problem at Several Times n for a Double-Integrator System

Chapter 5

COMPUTATIONAL OPTIMAL TRANSPORT OF CONTROL-AFFINE SYSTEMS

In this chapter, we consider multi-agent systems in which each agent's state $\mathbf{x}(t)$ is governed by a continuous-time nonlinear control system. The distribution of the agents' states is described by a time-varying measure over the phase space of a single agent. Specifically, we consider nonlinear control-affine systems of the form,

$$\dot{\mathbf{x}}(t) = g_0(\mathbf{x}(t), t) + \sum_{i=1}^{n} u_i(t) g_i(\mathbf{x}(t)), \tag{5.1}$$

where X is the $d-$dimensional phase space, and n is the number of control inputs and $\{g_i\}_{i=0}^{n}$ are smooth vector fields on X. Our aim is to compute controls u_i such that a cost of transporting a measure μ_{t_0} to μ_{t_f} over the time-horizon $[t_0, t_f]$ is minimized. This cost is given by the integral over phase-space and time,

$$C = \int_X \int_{t_0}^{t_f} \sum_{i=1}^{n} |u_i(\mathbf{x}, t)|^2 dt \, d\mu_t(\mathbf{x}). \tag{5.2}$$

In contrast to Chapter 4, where the goal was to transport the measures using stochastic feedback laws, the objective in this chapter is to construct deterministic feedback laws $u_i(\mathbf{x}, t)$.

5.1 Preliminaries

We briefly review concepts from control systems theory, optimal transport, and set-oriented numerical methods relevant to the discussion in Section 5.2. Specifically, we motivate the developments of Section 5.2 by relating the continuous and discrete (graph-based) concepts of optimal transport in controlled dynamical systems.

Optimal Transport in Controlled Dynamical Systems The Monge-Kantorovich optimal transport (OT) problem (Villani, 2003) is concerned with mapping of an initial measure μ_0 on a space X to a final measure μ_1 on a space Y. In the original formulation, it involves solving for a measurable transport map $T : X \to Y$, which pushes forward μ_0 to μ_1 in an optimal manner. The cost of transport per unit mass is prescribed by a function $c(\mathbf{x}, T(\mathbf{x}))$. Hence, the optimization problem is

$$\inf_T \int c(\mathbf{x}, T(\mathbf{x})) d\mu_0(\mathbf{x}), \tag{5.3}$$
$$\text{s.t. } T_{\#}\mu_0 = \mu_1,$$

where $T_{\#}$ is the pushforward of T, i.e. $(T_{\#}\mu)(A) = \mu(T^{-1}(A))$ for every A. In a "relaxed" version of this problem due to Kantorovich, the optimization problem is to obtain an optimal joint distribution $\pi(X \times Y)$ on the product space $X \times Y$, where the marginal of π on X is μ_0 and on Y is μ_1. We denote by $\prod(\mu_0, \mu_1)$ the set of all measures on product space with the marginals μ_0 and μ_1 on X and Y, respectively. Hence, the relaxed problem is

$$\inf_{\pi(X\times Y)\in\prod(\mu_0,\mu_1)} \int c(\mathbf{x}, \mathbf{y}) d\pi(\mathbf{x}, \mathbf{y}). \tag{5.4}$$

For the case of quadratic costs, i.e., $c(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|^2$, the support of the optimal distribution $\pi(X \times Y)$ is the graph of the optimal map T obtained from the solution of problem (5.3). The square root of the optimal cost obtained as the solution of this problem is called the $2-$Wasserstein distance, and we denote it by $W_2(\mu_0, \mu_1)$. We concern ourselves with only quadratic costs in this section.

An alternative fluid dynamical interpretation of the OT problem was provided by Benamou and Brenier (Benamou and Brenier, 2000). In this approach, the optimization problem is formulated in terms of an advection field $u(\mathbf{x}, t)$ and the initial and final *densities* $(\rho_0(\mathbf{x}), \rho_1(\mathbf{x}))$ of a single agent. The core idea is to obtain the optimal map T as a result of advection over a time period (t_0, t_f) by an optimal advection field $u(\mathbf{x}, t)$. It can be shown

that the optimization problem (5.3) (with $X = Y = \mathbb{R}^d$) with quadratic cost is equivalent to the following problem:

$$W_2^2(\mu_0,\mu_1) = \inf_{u(\mathbf{x},t),\rho(\mathbf{x},t)} \int_{\mathbb{R}^d}\int_{t_0}^{t_f} \rho(\mathbf{x},t)|\mathbf{u}(\mathbf{x},t)|^2 dtdx, \tag{5.5}$$

$$\text{s.t. } \frac{\partial \rho(\mathbf{x},t)}{\partial t} + \nabla\cdot(\rho(\mathbf{x},t)\mathbf{u}(\mathbf{x},t)) = 0, \tag{5.6}$$

$$\rho(\mathbf{x},t_0) = \rho_0(\mathbf{x}),\ \rho(\mathbf{x},t_f) = \rho_1(x).$$

The motion of a single agent is governed by the ordinary differential equation of the single integrator,

$$\dot{\mathbf{x}}(t) = \mathbf{u}(\mathbf{x},t). \tag{5.7}$$

By a change of variables from (ρ,u) to $(\rho,m \triangleq \rho u)$, the optimization problem (5.5), (5.6) can be put into a form where its convexity can be proved easily. The transformed convex optimization problem is

$$\inf_{\rho(\mathbf{x},t)\geq 0, m(\mathbf{x},t)} \int_{\mathbb{R}^d}\int_{t_0}^{t_f} \frac{|m(\mathbf{x},t)|^2}{\rho(\mathbf{x},t)} dtd\mathbf{x}, \tag{5.8}$$

$$\text{s.t. } \frac{\partial \rho(\mathbf{x},t)}{\partial t} + \nabla\cdot(m(\mathbf{x},t)) = 0,\ t_0 \leq t \leq t_f,$$

$$\rho(\mathbf{x},t_0) = \rho_0(\mathbf{x}),\ \rho(x,t_f) = \rho_1(\mathbf{x}).$$

The basic theory of generalization to general nonlinear controlled dynamical systems $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t),\mathbf{u}(t))$ has been developed in (Agrachev and Lee, 2009; Rifford, 2014). This problem can be interpreted as finding an optimal control which steers an initial scalar density to a final density, where the scalar transport occurs according to a controlled dynamical system $\mathbf{f}(\mathbf{x}(t),\mathbf{u}(t))$.

Transfer Operator and Infinitesimal Generator Consider the flow-map $\phi_{t_0}^{t_0+T} : X \to X$ on a d-dimensional phase space X. This map may be obtained as a time-T map of the flow

of a possibly time-dependent dynamical system,

$$\dot{\mathbf{x}} = \mathbf{f}(\mathbf{x}, t). \tag{5.9}$$

The corresponding Perron-Frobenius transfer operator (Lasota and Mackey, 1994) $P_{t_0}^{t_0+T}$ is a linear operator which pushes forward measures in phase space according to the dynamics of the trajectories under $\phi_{t_0}^{t_0+T}$. Let $\mathbf{B}(X)$ denote $\sigma-$algebra of Borel sets in X. Then, for any measure μ,

$$P_{t_0}^{t_0+T}\mu(A) = \mu((\phi_{t_0}^{t_0+T})^{-1}(A)) \;\; \forall A \in \mathbf{B}(X). \tag{5.10}$$

The transfer operator lifts the evolution of the dynamical systems from phase space X to the space of measures $\mathbf{M}(X)$. Numerical approximation of P, denoted by $\hat{P}$, may be viewed as a transition matrix of an N-state Markov chain (Bollt and Santitissadeekorn, 2013). For computation, we partition the phase space volume of interest into N $d-$dimensional connected, positive volume subsets, $B_1, B_2, \ldots, B_N$ with piecewise smooth boundaries ∂B_i. Usually, these subsets are hyperrectangles. The matrix $\hat{P} = \{\hat{p}_{ij}\}$ is numerically computed via the Ulam-Galerkin method (Ulam, 2004; Bollt and Santitissadeekorn, 2013) as follows:

$$\hat{p}_{ij} = \frac{\bar{m}\left((\phi_{t_0}^{t_0+T})^{-1}(B_i) \cap B_j\right)}{\bar{m}(B_j)}, \tag{5.11}$$

where $\bar{m}$ is the Lebesgue measure. The action of the transfer operator over a finite time T can also be defined naturally on densities in the case of Lebesgue absolutely continuous measures. However, we are more interested in capturing the continuous-time behavior of the dynamical system (5.9) in the space of densities. The continuity equation for the system in equation (5.9) is given by

$$\frac{d\mu}{dt} = -\nabla \cdot (f(x,t)\mu). \tag{5.12}$$

For the numerical approach used in this section, we briefly consider equation (5.12) in an operator-theoretic framework, as an abstract ordinary differential equation in the space

of measures, formally. Equation (5.12) can be expressed as

$$\dot{\mu}(t) = \mathscr{A}(t)\mu \;\; ; \;\; \mu(s) = \mu_s \in \mathbf{M}(X), \tag{5.13}$$

where $\mathscr{A}(t) : D(\mathscr{A}(t) \to \mathbf{M}(X))$, $D(\mathscr{A}(t)) \subset \mathbf{M}(X)$ and the solution, $\mu(t)$, of equation (5.13) can be expressed using a two-parameter semigroup of operators $(\mathscr{U}(t,s)_{s,t\in\mathbb{R},t\geq s}$ as $\mu(t) = \mathscr{U}(t,s)\mu_s$. The divergence operation is to be understood in the sense of duality of $M(X)$ with $C(X)$ (assuming X is compact). Here $C(X)$ refers to the space of continuous functions on X. The Perron-Frobenius operator is related to this two-parameter semigroup of operators as $\mathscr{U}(T,t_0) = P_{t_0}^{t_0+T}$ for given parameters t_0 and T. In general, guaranteeing the existence of a strongly continuous two-parameter semigroup based on the time-dependent generator $\mathscr{A}(t)$ is quite involved. See, for example, (Engel and Nagel, 2000; Fattorini, 1984). In contrast, the theory is more well-developed for the case when $\mathscr{A}(t) \equiv \mathscr{A}$ (the vector field $f(\mathbf{x})$ is time-independent). In this case, the solution, $\mu(t)$, can be expressed by a one-parameter semigroup of bounded operators, $(\mathscr{T}(t))_{t\geq 0}$, as $\mu(t) = \mathscr{T}(t-s)\mu_s$. Here, the generator $\mathscr{A}$ and $\mathscr{T}(t)$ are related by the formula

$$\mathscr{A}\mu = \lim_{h\to 0^+} \frac{\mathscr{T}(h)\mu - \mu}{h} \;\text{ for each }\; \mu \in D(\mathscr{A}). \tag{5.14}$$

As in the case of the Perron-Frobenius operator, one can also consider the semigroup and its generator on a space of densities, or equivalently, on a space of measures that are absolutely continuous with respect to a reference measure with additional regularity restrictions.

Ulam's method for approximating Perron-Frobenius operators using Markov matrices extends to numerical approximations of semigroups corresponding to the continuity equation. Analogously, one approximates the generator of the semigroup using transition rate matrices, which generate approximating semigroups on a finite state space. We recall this method as shown in (Froyland *et al.*, 2013). We denote by $\bar{B}_i$ the closure of B_i. The operator $\mathscr{A}(t)$ is approximated by defining elements of the time-varying transition rate matrix

$\{A_{ij}(t)\}$, which are computed as follows:

$$A_{ij}(t) = \begin{cases} \frac{1}{\bar{m}(B_j)} \int_{\bar{B}_i \cap \bar{B}_j} \max\{\mathbf{f}(\mathbf{x},t)\cdot \mathbf{n}_{ij}, 0\} dm_{d-1}(\mathbf{x}) & i \neq j, \\ -\sum_{k\neq i} \frac{\bar{m}(\bar{B}_k)}{\bar{m}(\bar{B}_i)} A_{ik}(t) \text{ otherwise,} \end{cases} \tag{5.15}$$

where $\mathbf{n}_{ij}$ is the unit normal vector pointing out of B_i into B_j if $\bar{B}_i \cap \bar{B}_j$ is a $(d-1)$-dimensional face, and the zero vector otherwise, andm_{d-1} denotes the $d-1$-dimensional measure. Note that in (Froyland *et al.*, 2013), the authors also considered the perturbed version of the operator, $-\nabla\cdot(\mathbf{f}(\mathbf{x},t)\cdot) : -\nabla\cdot(\mathbf{f}(\mathbf{x},t)\cdot) + \frac{\varepsilon^2}{2}\Delta$. This was mainly to exploit the spectral properties of the perturbed operator and the corresponding semigroup. However, in this work, the perturbed operator does not offer any visible advantages. Hence, we work with approximations of the operator, $-\nabla\cdot(\mathbf{f}(\mathbf{x},t)\cdot)$, alone. Nevertheless, we note that the discretization will introduce some numerical diffusion.

Monge-Kantorovich Transport on Graphs Now consider a directed graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$ on X, where the vertices $\mathscr{V}$ represent the subsets B_i as before, and the directed edges $\mathscr{E}$ are obtained from the topology of X. For each pair of neighboring vertices, two edges are constructed, one in each direction.

A continuous-time advection on such a graph can be described as (Berman *et al.*, 2009; Chapman, 2015),

$$\frac{d}{dt}\mu(t,v) = \sum_{e=(w\to v)} U(t,e)\mu(t,w) - \sum_{e=(v\to w)} U(t,e)\mu(t,v), \tag{5.16}$$

where $\mu(t,v)$ is the time-varying measure on a vertex v, and $U(t,e)$ is the flow on an edge e. Here we use the notation $e = (v \to w)$ to represent the edge e directed from a vertex v to w. The notion of optimal transport has been extended to such a continuous-time discrete-space setting recently (Maas, 2011; Gigli and Maas, 2013; Mielke, 2013; Solomon *et al.*, 2016). Following (Solomon *et al.*, 2016), one can formulate a quadratic-cost optimal

transport problem on $\mathscr{G}$ as follows. First, define an advective inner product between two flows U_1, U_2 as

$$\langle U_1, U_2 \rangle_\mu = \sum_{e=(v\to w)} \left(\frac{\mu(v)}{\mu(w)} . \frac{\mu(v)+\mu(w)}{2} \right) U_1(e) U_2(e). \tag{5.17}$$

Then the corresponding optimal transport distance between a set of measures (μ_0, μ_1) supported on $\mathscr{V}$ can be written as

$$\tilde{W}_N(\mu_0, \mu_1) = \inf_{U(t,e)\geq 0, \mu(t,v)\geq 0} \int_0^1 \|U(t,.)\|_{\mu(t,.)} dt, \tag{5.18}$$

such that equation (5.16) holds, and

$$\mu(0,v) = \mu_0(v),\ \mu(1,v) = \mu_1(v)\ \ \forall v \in V.$$

Here $\|U(t,.)\|_{\mu(t,.)} \triangleq \sqrt{\langle U, U \rangle_\mu}$. This approach is motivated by the previously discussed Benamou-Brenier approach for optimal transport on continuous spaces, and results in the following advection-based convex optimization problem:

$$\tilde{W}_N(\mu_0, \mu_1)^2 = \inf_{J(t,e)\geq 0, \mu(t,v)\geq 0} \int_0^1 \sum_{e=(v\to w)} \frac{J(t,e)^2}{2} \left(\frac{1}{\mu(t,v)} + \frac{1}{\mu(t,w)} \right) dt, \tag{5.19}$$

$$\mu(0,v) = \mu_0(v),\ \mu(1,v) = \mu_1(v)\ \ \forall v \in V, \tag{5.20}$$

$$\frac{d}{dt}\mu(t,.) = D^T J(t,.), \tag{5.21}$$

where $J(t,e) \triangleq \mu(t,v)U(t,e)$ for $e = (v \to w)$, and $D \in \mathbb{R}^{|\mathscr{E}|\times|\mathscr{V}|}$ is the linear flow operator computing $\mu(w) - \mu(v)$ for each $e = (v \to w) \in \mathscr{E}$. Specifically, $D^T(i,j)$ equals $+1$ if the jth edge points into the ith vertex, -1 if the jth edge points out of the ith vertex, and 0 if the jth edge is not connected to the ith vertex. Hence, equation (5.21) is a rewriting of equation (5.16) in terms of $J(t,.)$. The change of variables from U to J is analogous to the change of variables in Benamou-Brenier formulations, as discussed earlier in this section. Conceptually, one can regard the problem described by equations (5.19-5.21) as the graph-based analogue of the optimal transport problem (5.8). Recall that this corresponds to

single-integrator dynamics $\dot{x} = u(t)$. In the next section, we use this interpretation, and generalize this graph-based framework to nonlinear dynamical systems of the form given in equation (5.1).

5.2 Problem Setup and Computational Approach

Formulation of Optimal Transport Problem on Graphs Let $M \subset \mathbb{R}^d$ be an open bounded connected subset of an Euclidean space with piecewise smooth boundary. For a collection of analytic time-invariant vector fields $\{\mathbf{g}_i\}_{i=1}^n$ and possibly time-varying vector field $\mathbf{g}_0$ on M, consider the control affine system of the form

$$\dot{\mathbf{x}}(t) = \mathbf{g}_0(\mathbf{x}(t),t) + \sum_{i=1}^{n} u_i(t)\mathbf{g}_i(\mathbf{x}(t)), \qquad x(0) = x_0. \tag{5.22}$$

Then given the densities ρ_0 and ρ_1 on M, the corresponding optimal transport problem of interest is the following:

$$\inf_{u(\mathbf{x},t),\rho(\mathbf{x},t)} \int_{\mathbb{R}^n} \int_{t_0}^{t_f} \sum_{i=1}^{n} \rho(\mathbf{x},t)|u_i(\mathbf{x},t)|^2 dt d\mathbf{x}, \tag{5.23}$$

$$\text{s.t. } \frac{\partial \rho(x,t)}{\partial t} + \nabla \cdot (\rho(\mathbf{x},t)\mathbf{g}_0(\mathbf{x},t)) + \sum_{i=1}^{n} \nabla \cdot (\rho(x,t)u_i(\mathbf{x},t)\mathbf{g}_i(\mathbf{x})) = 0,\ x \in M, \tag{5.24}$$

$$\vec{n} \cdot (\mathbf{g}_0(x,t)\rho(\mathbf{x},t) + \sum_{i} u_i(x,t)\mathbf{g}_i(x)\rho(\mathbf{x},t)) = 0 \ \ a.e.\ \ x \in \partial M,$$

$$\rho(\mathbf{x},t_0) = \rho_0(x),\ \rho(\mathbf{x},t_f) = \rho_1(\mathbf{x}).$$

Here, $\vec{n}$ is the outward normal vector at the boundary of M, and we have assumed zero mass flux boundary conditions.

We approximate the optimal transport problem using a sequence of optimal transport problems on graphs. A key tool is to approximate the (time-varying) generator of the semigroup corresponding to equation (5.24) using generator approximations on a finite state space (Froyland *et al.*, 2013), as discussed in Section 5.1. Hence, we approximate

solutions of optimal transport problems on a Euclidean space using solutions of optimal transport problems on graphs.

Construction of Graph $\mathscr{G}$: Toward this end, we partition M into m d-dimensional connected, positive volume subsets $P_m = \{B_1, B_2, ..., B_m\}$. Additionally, we assume that the boundaries ∂B_i are piecewise smooth. Then we can consider the optimal transport problem on a graph $\mathscr{G} = (\mathscr{V}, \mathscr{E})$, where the the cardinality of $\mathscr{V}$ is m and the connectivity of the graph is determined by the topology of M and the partition P_m. More specifically, $\mathscr{V} = \{1, 2, ..., m\}$ and an element $e = (v \to w) \in \mathscr{E}$ for $v, w \in \mathscr{G}$ and $v \neq w$ if $\bar{B}_v \cap \bar{B}_w$ has nonzero $(d-1)$-dimensional measure. The graph $\mathscr{G}$ is *strongly connected*, i.e., for any two vertices $v_0, v_T \in \mathscr{V}$, there exists a directed path $(v_1, v_2, ..., v_r)$ of r vertices in $\mathscr{V}$ such that $(v_i \to v_{i+1}) \in \mathscr{E}$ for each $i \in \{1, 2, ..., r-1\}$. Moreover, this graph is also *symmetric*, that is, $e = (v \to w) \in \mathscr{E}$ implies that $\bar{e}$ defined by $\bar{e} = (w \to v)$ is also in $\mathscr{E}$.

In order to apply the approximation procedure from (Froyland *et al.*, 2013), we express the continuity equation (5.24) as a bilinear control system,

$$\dot{y}(t) = \mathscr{A}_0(t)y + \sum_{i=1}^{n} \mathscr{A}_i(\hat{u}_i(t)y(t)), \tag{5.25}$$

where $\mathscr{A}_0(t) = -\nabla \cdot (g_0(x,t) \cdot\ )$ for each $t \in [0,1]$, $\hat{u}_i(t) = u_i(\cdot, t)$, $y(t) = \rho(\cdot, t)$, $\mathscr{A}_i = -\nabla \cdot (g_i(x) \cdot\ )$. Note that the right-hand side of a bilinear system is traditionally expressed in the form $A(t)\rho(t) + u(t)B\rho(t)$ in the control theory literature (Elliott, 2009). The form in equation (5.25) is equivalent for systems on finite-dimensional state spaces, but not for general infinite-dimensional bilinear systems if $\hat{u}(t)$ is not a scalar for each $t \in [t_0, t_f]$. For example, in the continuity equation, one can see that $\mathbf{u}(\mathbf{x}, t)\nabla \cdot (\rho(x,t)) \neq \nabla \cdot (\mathbf{u}(\mathbf{x}, t)\rho(\mathbf{x}, t))$ in general. Hence, the form of equation (5.25) is more appropriate for expressing the system in equation (5.24).

In Section 5.1, it was discussed how generators of semigroups corresponding to the continuity equation can be used to define a approximating semigroup on a graph generated by appropriately constructed transition rate matrices. This method can be generalized to the controlled continuity equation, equation (5.24). A natural extension is to consider approximations of the control operators $\mathscr{A}_i$ using corresponding transition rate matrices, and analogously construct a controlled Markov chain on the space $\mathscr{V}$. However, we note that typically for a controlled Markov chain, the control parameters are constrained to be non-negative. Hence, a direct approximation of $\mathscr{A}_i$ using transition rate matrices and constraining $\hat{u}_i(t)$ to be positive would negate the possibility that the system can flow both backward and forward along the control vector fields, which is critical for controllability of the system. Hence, to account for this in the approximation procedure, we define a bilinear control system equivalent to the one in equation (5.25), but with positivity constraints on the control:

$$\dot{y}(t) = \mathscr{A}_0(t)y + \sum_{s\in\{+,-\}}\sum_{i=1}^{n} \mathscr{A}_i^s(\hat{u}_i^s(t)y(t)); \;\; \hat{u}_i^s(t) \geq 0 \tag{5.26}$$

where $\mathscr{A}_i^+ = -\mathscr{A}_i^- = \mathscr{A}_i$ for each $i \in \{1,2,...,n\}$.

Using the methodology introduced in Section 5.1, for each of the operators $\mathscr{A}_0$, $\mathscr{A}_i^s$, we construct the control operators on the graph $\mathscr{G}$, which are denoted by $A_0 : [0,T] \times \mathscr{E} \to \mathbb{R}^+$ and $A_i^s : \mathscr{E} \to \mathbb{R}^+$. (Recall that only g_0 is possibly time-varying, while g_i, $i > 0$, are all time-invariant.) The difference is that while generators in Section 5.1 were defined as *vertex-based* $|\mathscr{V}| \times |\mathscr{V}|$ transition rate matrices, here we construct *edge-based* vectors of size $|\mathscr{E}|$ in a natural way. Hence, A_0 is the edge-based version of the generator constructed from the vector field $g_0(\mathbf{x},t)$ using the formula in equation (5.15). For A_i^s, the corresponding transition rates are defined as

$$A_i^+(e) = A_i^+(v \to w) = \frac{1}{m(B_w)} \int_{\bar{B}_v \cap \bar{B}_w} \max\{\mathbf{g}_i(\mathbf{x}) \cdot \mathbf{n}_{vw}, 0\} dm_{d-1}(\mathbf{x}), \tag{5.27}$$

$$A_i^-(e) = A_i^-(v \to w) = \frac{1}{m(B_w)} \int_{\bar{B}_v \cap \bar{B}_w} \max\{-\mathbf{g}_i(\mathbf{x}) \cdot \mathbf{n}_{vw}, 0\} dm_{d-1}(\mathbf{x}), \tag{5.28}$$

for $i = 1, \ldots, n$, where $\mathbf{n}_{vw}$ is the unit normal vector pointing out of B_v into B_w at $\mathbf{x}$.

Construction of Control Graph $\mathscr{G}_c$ and Drift Graph $\mathscr{G}_0$: Let $\mathscr{P}(\mathscr{V})$ be the space of probability densities on the finite state space, $\mathscr{V}$. Then using the above parameter definitions, we consider the following flows on the graph $\mathscr{G}$,

$$\begin{aligned}\frac{d}{dt}\mu(t,v) = &\sum_{e=(w\to v)} A_0(t,e)\mu(t,w) - \sum_{e=(v\to w)} A_0(t,e)\mu(t,v) \\ &+ \sum_{s\in\{+,-\}}\sum_{i=1}^{n}\sum_{e=(w\to v)} A_i^s(e)U_i^s(t,e)\mu(t,w) \\ &- \sum_{s\in\{+,-\}}\sum_{i=1}^{n}\sum_{e=(v\to w)} A_i^s(e)U_i^s(t,e)\mu(t,v), \end{aligned} \tag{5.29}$$

where $\mu(t,\cdot) \in \mathscr{P}(\mathscr{V})$ for each $t \in [0,T]$, and $U_i^s(t,\cdot)$ are the edge-dependent non-negative "control" parameters that scale the transition rates, $A_i^s(e)$. We associate a set of edges $\mathscr{E}_i^s$ with the above controlled flow. For each $s \in \{+,-\}$ and $i \in \{1,2,...,n\}$, we set $e \in \mathscr{E}_i^s$ if $A_i^s(e) \neq 0$. Similarly, we define $\mathscr{E}_0$ by setting $e \in \mathscr{E}_0$ if $A_0(t,e) \neq 0$ for some $t \in [0,1]$. Using these definitions, we define the *control graph* $\mathscr{G}_c = (\mathscr{V}, \mathscr{E}_c)$ by setting $\mathscr{E}_c = \cup_{s\in\{+,-\}} \cup_{i=1}^{n} \mathscr{E}_i^s$, and the *drift graph* $\mathscr{G}_0 = (\mathscr{V}, \mathscr{E}_0)$. These definitions will be used in Section 5.2.

The above defined flows can be shown to correspond to the evolution of a time-inhomogeneous continuous-time Markov chain on the finite state space, $\mathscr{V}$. The evolution of the corresponding stochastic process $X(t) \in \mathscr{V}$ over an edge, $e = (w \to v) \in \mathscr{E}$, is defined by the conditional probabilities:

$$\mathbb{P}(X(t+h) = v | X(t) = w) = A_0(t,e) + \sum_{s\in\{+,-\}}\sum_{i=1}^{n}\sum_{e=(w\to v)} A_i^s(e)U_i^s(t,e) + o(h). \tag{5.30}$$

This leads us to the approximating optimal transport problem on a graph, motivated by

the formulation in Section 5.1:

$$\tilde{W}(\mu_0,\mu_1) = \inf_{U_i^s(t,e)\geq 0, \mu(t,v)\geq 0} \sum_{s\in\{+,-\}} \sum_{i=1}^{n} \int_0^1 \|U_i^s(t,.)\|_{\mu(t,.)} dt \tag{5.31}$$

such that equation (5.29) holds, and

$$\mu(0,v) = \mu_0(v),\ \mu(1,v) = \mu_1(v)\ \ \forall v \in \mathscr{V}$$

Again, the formulation in Section 5.1 motivates the following convex formulation of the above problem:

$$\tilde{W}(\mu_0,\mu_1)^2 = \inf_{J_i^s(t,e)\geq 0, \mu(t,v)\geq 0} \sum_{s\in\{+,-\}} \sum_{i=1}^{n} \int_0^1 \sum_{e=(v\to w)} \frac{J_i^s(t,e)^2}{2} \left(\frac{1}{\mu(t,v)} + \frac{1}{\mu(t,w)}\right) dt, \tag{5.32}$$

$$\mu(0,v) = \mu_0(v),\ \mu(1,v) = \mu_1(v)\ \ \forall v \in \mathscr{V},$$

$$\frac{d}{dt}\mu(t,.) = \sum_{e=(w\to v)} A_0(t,e)\mu(t,w) - \sum_{e=(v\to w)} A_0(t,e)\mu(t,v) + \sum_{s\in\{+,-\}} \sum_{i=1}^{n} (D_i^s)^{\intercal} J_i^s(t,.), \tag{5.33}$$

where $J_i^s(t,e) \triangleq \mu(t,v)U_i^s(t,e)$ for $e=(v\to w)$, $i=\{1,2,...,n\}$, and $D_i^s \in \mathbb{R}^{|\mathscr{E}_i^s|\times|\mathscr{V}|}$ is the linear flow operator computing $\mu(w)-\mu(v)$ for each $e=(v\to w)\in\mathscr{E}_i^s$.

Remark 5.2.1. *We note that the controlled advection equation (5.29) and the corresponding convex optimal transport problem (5.32) can be simplified if control vector fields are unidirectional across all boundaries ∂B_i. This can often be achieved by choosing the grid carefully, and making the subvolumes B_i small enough. If this condition holds, then we immediately see from equations (5.27-5.28) that for each edge $e=(v\to w)$, only one of the transition rates $A_i^+(e)$ and $A_i^-(e)$ is nonzero. Denote the nonzero transition rate by $A_i(e)$. It also follows that $A_i(e) = A_i(\bar{e})$, where $\bar{e} = (w\to v)$. Then the simplified version of equation*

(5.29) is

$$\frac{d}{dt}\mu(t,v) = \sum_{e=(w\to v)} A_0(t,e)\mu(t,w) - \sum_{e=(v\to w)} A_0(t,e)\mu(t,v) + \sum_{i=1}^{n}\sum_{e=(w\to v)} A_i(e)U_i(t,e)\mu(t,w) - \sum_{i=1}^{n}\sum_{e=(v\to w)} A_i(e)U_i(t,e)\mu(t,v). \quad (5.34)$$

This results in the following convex optimal transport problem,

$$\tilde{W}(\mu_0,\mu_1)^2 = \inf_{J_i(t,e)\geq 0,\mu(t,v)\geq 0} \sum_{i=1}^{n}\int_0^1 \sum_{e=(v\to w)} \frac{J_i(t,e)^2}{2}\left(\frac{1}{\mu(t,v)}+\frac{1}{\mu(t,w)}\right)dt, \quad (5.35)$$

$$\mu(0,v)=\mu_0(v),\ \mu(1,v)=\mu_1(v),\ \ \forall v\in\mathscr{V}$$

$$\frac{d}{dt}\mu(t,.) = \sum_{e=(w\to v)} A_0(t,e)\mu(t,w) - \sum_{e=(v\to w)} A_0(t,e)\mu(t,v) + \sum_{i=1}^{n}(D_i)^{\intercal}J_i(t,.). \quad (5.36)$$

Remark 5.2.2. *We note that equation (5.16), discussed in Section 5.1, can be seen as the special case of equation (5.34) with* $\mathbf{g}_0 \equiv 0$ *and* $\mathbf{g}_i = \hat{i}$ *(the ith unit vector). Hence, our formulation generalizes optimal transport on graphs from a single-integrator system to general nonlinear control-affine systems.*

Controllability Analysis of Flow over Graphs

In this section, we establish that the controlled Markov chain approximations (5.29) preserve the controllability properties of the system (5.22). In other words, we will show that if the underlying dynamical system (5.22) satisfies some controllability conditions, then the dynamical system (5.29) governing the evolution of measure on the graph $\mathscr{G}$ is also controllable in some precise sense. This will ensure the well-posedness of the graph optimal transport problem (5.31), since optimal transport is meaningful only if the set of possible transports between a pair of measures is non-empty.

Our procedure is as follows. First, in Theorem 5.2.7, we will prove that controllability of equation (5.22) results in the control graph $\mathscr{G}_c$ being strongly connected and equal to $\mathscr{G}$. In the subsequent theorems, we will show that the strongly connected property of

$\mathscr{G}_c = \mathscr{G}_c$ implies that the system defined by equation (5.29) is controllable between any pair of measures in the interior of $\mathscr{P}(\mathscr{V})$. This is first shown for the case of driftless systems (i.e., $\mathbf{g}_0 \equiv 0$) in Theorem 5.2.9, and then for systems with drift (i.e., $\mathbf{g}_0 \not\equiv 0$) in Theorem 5.2.10. Here, the interior of $\mathscr{P}(\mathscr{V})$ is defined as the set $\text{int}(\mathscr{P}(\mathscr{V})) = \{\mu \in \mathscr{P}(\mathscr{V}); \mu(v) > 0 \text{ for each } v \in \mathscr{V}\}$.

Without loss of generality, we consider the case when $t_0 = 0$ and $t_f = 1$. First, we recall a few standard notions from geometric control theory (Bloch, 2015).

Definition 5.2.3. *Given $\mathbf{x}_0 \in M$, we define $R(x_0, t)$ to be the set of all $\mathbf{y} \in M$ for which there exists an admissible control $\mathbf{u} = (u_1, u_2, ..., u_n)$ such that there exists a trajectory of system (5.22) with $\mathbf{x}(0) = \mathbf{x}_0$, $\mathbf{x}(t) = \mathbf{y}$. The **reachable set from $\mathbf{x}_0$ at time** T is defined to be*

$$R_T(\mathbf{x}_0) = \cup_{0 \le t \le T} R(\mathbf{x}_0, t). \tag{5.37}$$

Definition 5.2.4. *We say that the system (5.22) is **small-time locally controllable** from $\mathbf{x}_0$ if $\mathbf{x}_0$ is an interior point of $R_T(\mathbf{x}_0)$ for any $T > 0$.*

Definition 5.2.5. *Let $\mathbf{f} = (f_1, ..., f_d)$ and $\mathbf{g} = (g_1, ..., g_d)$ be two smooth vector fields on M. Then the **Lie bracket** $[\mathbf{f}, \mathbf{g}]$ is defined to be the vector field with components*

$$[\mathbf{f}, \mathbf{g}]^i = \sum_{j=1}^{d} \left(f^j \frac{\partial g^i}{\partial x^j} - g^j \frac{\partial f^i}{\partial x^j} \right). \tag{5.38}$$

Definition 5.2.6. *For a collection of vector fields $\{\mathbf{g}_i\}$, $\mathbf{Lie}\{\mathbf{g}_i\}$ refers to the smallest Lie subalgebra of a set of smooth vector fields on M that contains $\{g_i\}$. $\mathbf{Lie}_x\{\mathbf{g}_i\}$ refers to the span of all vector fields in $\mathbf{Lie}\{\mathbf{g}_i\}$ at $\mathbf{x} \in M$.*

Using these definitions, we have the following result.

Theorem 5.2.7. *Suppose that one of the following statements is true:*

1. $\mathbf{g}_0 \equiv 0$ *and* $\mathbf{Lie}_{\mathbf{x}}\left\{\mathbf{g}_i : i \in \{1,2,...,n\}\right\} = T_{\mathbf{x}}M$ *at each* $\mathbf{x} \in int(M)$.

2. $span\left\{\mathbf{g}_i(\mathbf{x}) : i \in \{1,2,...,n\}\right\} = T_{\mathbf{x}}M$ *at each* $\mathbf{x} \in int(M)$.

Then the graph $\mathscr{G}_c$ *associated with the system (5.29) is strongly connected and* $\mathscr{G}_c = \mathscr{G}$.

Proof. Let $v, w \in \{1,2,...,m\}$ be such that $v \neq w$ and $\bar{B_v} \cap \bar{B_w}$ has nonzero $(d-1)-$dimensional (Hausdorff) measure. Consider points $\mathbf{x}_0 \in \text{int}(B_v)$ and $\mathbf{x}_1 \in \text{int}(B_w)$. Due to the connectedness of M, there exists a continuous path $\gamma : [0,1] \to M$ such that $\gamma(0) = \mathbf{x}_0$, $\gamma(1) = \mathbf{x}_1$, and $\gamma(t) \in B_v \cup B_w \; \forall t \in [0,1]$. From the Lie bracket condition of the vector fields, it follows that the system is small-time locally controllable at every $\mathbf{x} \in \text{int}(M)$. Then, we can approximate the path γ as a trajectory of the control system, using a sequence of piecewise-constant control inputs.

To construct such a sequence, let us denote the flow map for time period t under an autonomous vector field X by e^{tX}. Then, for each $\varepsilon > 0$, there exists $k \in \mathbb{N}$ large enough, a sequence of time intervals $t_1, t_2, ..., t_k$ satisfying $\sum_{i=1}^{k} t_i = 1$, constant control inputs $u^1, u^2, ..., u^k \in \mathbb{R}$, a set of indices $\eta_i \in \{1,2,...,n\}$ selecting the corresponding control vector field $\mathbf{g}_{\eta_i}$, and an approximating path $f : [0,1] \to M$ satisfying $\|\gamma(z) - f(z)\|_2^2 \leq \varepsilon$ for all $z \in [0,1]$. The path $f(z)$ for $z \in [0,1]$ can be written using the concatenation of flows under the action of a chosen sequence of control vector fields:

$$f(\sum_{j=1}^{j} t_j + \tau) = e^{\tau u^{j+1} \mathbf{g}_{\eta_{j+1}}} \circ ... \circ e^{t_j u^j \mathbf{g}_{\eta_j}} \circ e^{t_1 u^1 \mathbf{g}_{\eta_1}} \mathbf{x}_0 \tag{5.39}$$

for each $j \in \{0,1,...,k\}$ and $\tau \in [t_j, t_{j+1}]$. Here, the case $j = 0$ means $f(\tau) = e^{\tau u^1 \mathbf{g}_{\eta_1}} \mathbf{x}_0$ for all $\tau \in [0, t_1]$.

Let $z^* \in (0,1)$ be such that $f(z^*) \in \partial B_v$ and there exists $c \in (0, z^*)$ small enough such that $f(z^* - c) \in \text{int}(B_v)$ and $f(z^* + c) \in \text{int}(B_w)$. Then, clearly $\mathbf{n}_{vw} \cdot \mathbf{g}_r(\mathbf{x}) \neq 0$ for some $r \in \{1,2,...,n\}$ and some x in an open neighborhood of $f(z^*)$ that is completely contained

in $B_v \cup B_w$, assuming that γ and ε are chosen appropriately (i.e. avoiding crossings of γ and f over corners of B_v and B_w). If not, then $f(z^* + \delta) \in \partial B_i$ for all $\delta \in (0, c]$, since the non-existence of such a point c with the desired property in the neighborhood of $f(z^*)$ implies that one cannot use a concatenation of flows associated with the control vector fields to leave the set ∂B_v, which leads to a contradiction to the assumed property of small-time local controllability. From continuity of the vector field $\mathbf{g}_r$, there exists a small enough neighborhood $N_{\mathbf{x}}$ of $\mathbf{x}$ such that $\mathbf{n}_{vw} \cdot \mathbf{g}_r(\mathbf{y}) \neq 0$ for all $\mathbf{y} \in N_{\mathbf{x}}$. Hence, this implies that $A_r^s(e) \neq 0$ for $e = v \to w$ for some $s \in \{+, -\}$. Due to continuity of the vector field $\mathbf{g}_r$ at x, it also follows that $A_r^s(e) = A_r^s(\bar{e})$. Hence, the connectivity of the graph $\mathscr{G}_c$ follows. Case 2 follows from the assumption that $span\left\{ \mathbf{g}_i(\mathbf{x}) : i \in \{1, 2, ..., n\} \right\} = T_{\mathbf{x}}M$ at each $x \in \text{int}(M)$. □

Remark 5.2.8. *The main obstruction in extending the above result for underactuated systems (span* $\left\{ \mathbf{g}_i(\mathbf{x}) : i \in \{1, 2, ..., n\} \right\} \neq T_{\mathbf{x}}M$ *for some* $\mathbf{x} \in M$*) with drift, i.e.* $\mathbf{g}_0 \not\equiv 0$*, is that usual tests for small-time local controllability of control systems with drift (Sussmann, 1987) require the initial condition to be an equilibrium point. Hence, starting at a non-equilibrium initial condition, one might need to make large excursions (in our case, possibly outside the domain M) in order to return to the initial condition. For example, consider the simplest control-affine system with drift, the double integrator:* $\ddot{\mathbf{x}} = u$*. Hence, given initial and target densities, the optimal transport problem on a bounded domain might not admit a solution for a system with drift if M is not taken to be large enough.*

In the following, we observe that equation (5.29) has a certain controllability property for the case when the underlying system is driftless (i.e., $\mathbf{g}_0 \equiv 0$). The proof follows from Theorem 2.2.5, where the controllability result was proved for the case when $A_i(t, e)$ is either equal to 0 or 1 for each $i \in \{1, 2, ..., n\}$ and each $e \in \mathscr{G}_c$, and $\mathscr{G}_c$ is only required to be

strongly connected.

Theorem 5.2.9. *Consider $\mu_0, \mu_1 \in int(\mathscr{P}(\mathscr{V}))$ and assume that $\mathscr{G}_c = \mathscr{G}$ is strongly connected and $A_0(t,e) = 0$ for every $e \in \mathscr{E}$ and all $t \in [0,1]$. Then there exist piecewise continuous $U_i^s(t,\cdot) \geq 0$ such that the solution of equation (5.29), $\mu(t,\cdot)$ satisfies $\mu(0,\cdot) = \mu_0$ and $\mu(1,\cdot) = \mu_1$.*

Theorem 5.2.9 leads to the following result for the case of systems with drift, i.e., $g_0 \not\equiv 0$.

Theorem 5.2.10. *Consider $\mu_0, \mu_1 \in int(\mathscr{P}(\mathscr{V}))$. Assume the graph $\mathscr{G}_c = \mathscr{G}$ is strongly connected, and $\mathscr{G}_0 \subseteq \mathscr{G}_c$. Then there exist $U_i^s(t,\cdot) \geq 0$ such that equation (5.29) satisfies $\mu(0,\cdot) = \mu_0$ and $\mu(1,\cdot) = \mu_1$.*

Proof. The graph $\mathscr{G}_c$ is connected. Since $\mathscr{G}_0 \subseteq \mathscr{G}$, we can choose $\tilde{U}_i^s(t,\cdot)$ such that the right-hand side of equation (5.29) is equal to 0 for all $t \in [0,1]$. Then, from the previous theorem, it follows that there exists a control $U_i^s(t,\cdot)$ of the form $U_i^s(t,\cdot) = \hat{U}_i^s(t,\cdot) + \tilde{U}_i^s(t,\cdot)$ such that equation (5.29) satisfies $\mu(0,\cdot) = \mu_0$ and $\mu(1,\cdot) = \mu_1$. Here, the parameters $\tilde{U}_i(t,\cdot)$ negate the effect of the drift field A_0, and $\hat{U}_i^s(t,\cdot)$ ensure that the density μ_0 is transported to μ_1, as in Theorem 5.2.9. □

5.3 Construction of Approximate Feedback Control Laws

Given the solution the optimal transport problem on the graph, we reconstruct the corresponding approximate feedback control laws $\{u_i(\mathbf{x},t)\}$ for the underlying dynamical system Eq. (5.22). Since the optimal transport problem is solved on the graph, the feedback control law is vertex-based. For any vertex v of the graph $\mathscr{G}$, all agents with their state $\mathbf{x}$ lying in the sub-volume B_v apply the following feedback law:

$$u_i(\mathbf{x},t) = \frac{\sum_{w\in\mathscr{N}_i^+(v)} U_i^+(v\to w,t)}{|\mathscr{N}_i^+(v)|} - \frac{\sum_{w\in\mathscr{N}_i^-(v)} U_i^-(v\to w,t)}{|\mathscr{N}_i^-(v)|} \quad \forall x\in B_v. \tag{5.40}$$

Here, $\mathscr{N}_i^s(v)$ refers to the the neighboring vertices of v in the graph $(\mathscr{V},\mathscr{E}_i^s)$ for each $s\in\{+,-\}$ and $i\in\{1,2....n\}$.

5.4 Numerical Implementation

We adapt the numerical scheme used in (Solomon *et al.*, 2016) to our setting, and use a staggered discretization scheme for pseudo-time discretization. We define

$$\mu_j(v) \triangleq \mu(t_j,v), \tag{5.41}$$

$$J_{i,j}^s(e) \triangleq J_i^s(t_j,e), \tag{5.42}$$

where $t_j = (j/k)t_f, j\in[0,1,2,\ldots,k]$ is the time discretization into k intervals. We take $t_0=0$. Here $J_{i,j}^s(e)$ represents the $s\in\{+,-\}$ flow due to $g_i(\mathbf{x})$ over edge $e=(v\to w)$, from vertex v at time t_j to vertex w at time t_{j+1}.

Hence, the optimization problem given in Eqs. (5.32) can be discretized as,

$$\tilde{W}(\mu_0,\mu_1)^2 = \inf_{J_{i,j}^s\geq 0,\mu_j\geq 0} \sum_{s\in\{+,-\}}\sum_{i=1}^{n}\sum_{j=1}^{k}\sum_{\substack{e=1\\ e=(v\to w)}}^{|\mathscr{E}_i^s|} (J_{i,j}^s(e))^2\left(\frac{1}{\mu_j(v)}+\frac{1}{\mu_{j+1}(w)}\right), \tag{5.43}$$

subject to the following constraints:

$$\frac{\mu_{j+1}-\mu_j}{\Delta t} = A_0(t_j)\mu_j + \sum_{s\in\{+,-\}}\sum_{i=1}^{n}(D_i^s)^{\intercal} J_s^{i,j}, \tag{5.44}$$

$$\mu_0 = \mu_{t_0}, \mu_k = \mu_{t_f}, \tag{5.45}$$

where we have used the vertex-based $m\times m$ transition rate matrix $A_0(t_j)$ as originally defined in Eq. (5.15). Here $\Delta t = \frac{t_f}{k}$. The cost function given by Eq. (5.43) is again of the

form "quadratic over linear," and the advection (Eq. (5.44)) imposes linear constraints. Hence the discretized problem is convex, and can be solved using many off-the-shelf convex solvers. The optimization problem is solved via the CVX (Grant *et al.*, 2008) modeling platform, an open-source software for converting convex optimization problems into a usable format for various solvers. We use the SCS (O'Donoghue *et al.*, 2013) solver, a first-order solver for large size convex optimization problems. This solver uses the Alternating Direction Method of Multipliers (ADMM) (Eckstein and Yao, 2012) to enable quick solution of very large convex optimization problems, with moderate accuracy.

The variables to be solved for in the optimization problem Eqs. (5.43-5.45) are vertex-based quantities μ_j and edge-based quantities $J^s_{i,j}$. The size of the optimization problem can be quantified in terms of the number of time-discretization steps k, the number of vertices $|\mathscr{V}| = m$, and the number of edges $|\mathscr{E}_c|$. The graph $\mathscr{G}_c$ is always sparse, since a typical vertex is at most connected to $2(n+1)d$ neighbors, and $m \gg n, m \gg d$. Hence, the variables in the optimization problem scale as $O(k(m+|\mathscr{E}|)) = O(n \cdot d \cdot k \cdot m)$.

In the examples that follow, the graph size m is chosen to be large enough so that the qualitative features of the optimal transport are well resolved, and do not change upon finer grid refinement. The time-discretization parameter k is chosen such that the optimal transport cost $\tilde{W}$ is insensitive to finer discretization.

5.5 Simulation Examples

Optimal Transport in the Grushin Plane

We first apply our framework to a non-holonomic control-affine system in which certain optimal transport solutions can be found analytically. We consider transport of measure in the Grushin system. In (Agrachev and Lee, 2009), the structure of optimal controls in this problem was analyzed. Using this structure, optimal transport to a delta measure at $(0,0)$

was computed. The system is described by

$$\dot{x}_1 = u_1, \tag{5.46a}$$

$$\dot{x}_2 = u_2 x_1. \tag{5.46b}$$

This system is a driftless system with control vector fields $\mathbf{g}_1(x_1,x_2) = [1\,0]^\intercal$, $\mathbf{g}_2(x_1,x_2) = [0\,x_1]^\intercal$. These do not span the tangent space $\mathbb{R}^2$, but their Lie algebra does, i.e. $Lie_{\mathbf{x}}\left\{ g_i : i \in \{1,2\} \right\} = \mathbb{R}^2$. This can be seen by noting that the Lie bracket $[\mathbf{g}_1,\mathbf{g}_2] = [0\ 1]^\intercal$, and hence $span\{[\mathbf{g}_1,\mathbf{g}_2],\mathbf{g}_1\} = \mathbb{R}^2$. Hence, this system satisfies condition 1 of Theorem 5.2.7. By Theorem 5.2.9, the corresponding numerical optimal transport problem for this driftless system is well-posed.

The optimal control cost $c(\mathbf{x},\mathbf{y})$ between initial and final states, $\mathbf{x} = (x_1,x_2)^\intercal, \mathbf{y} = (y_1,y_2)^\intercal$, is taken to be square of the sub-Riemannian distance $d(\mathbf{x},\mathbf{y}) = \inf_{\mathbb{U}_\mathbf{x}^\mathbf{y}} \int_0^1 \sqrt{u_1^2+u_2^2}dt$. Hence, the optimal control solutions are also geodesics in the sub-Riemannian space. The solutions of the optimal control problem are integral curves of the Hamiltonian H given by

$$H(x_1,x_2,p_1,p_2) = \frac{1}{2}(p_1^2 + x_1^2 p_2^2). \tag{5.47}$$

Here p_1, p_2 are the co-state variables. Note that since H is independent of x_2, H can be reduced to a Hamiltonian in (x_1,p_1), and the integral curves of H can be obtained using quadratures. The geodesics reaching $(0,\alpha)$ at $t=1$ are of the form

$$x_1(t) = \frac{a}{b}\sin(b(1-t)), \tag{5.48}$$

$$x_2(t) = \frac{a^2}{4b^2}(2b(1-t) - \sin(2b(1-t))) + \alpha. \tag{5.49}$$

A geodesic between a specified initial point $(\bar{x}_1,\bar{x}_2)$, and $(0,\alpha)$ can be obtained by inverting the Eqs. (5.48-5.49) at $t=0$ to solve for (a,b). For $t \leq \frac{\pi}{b}$, these geodesics are also global minimizers of the optimal control problem. Figure 5.1(a) shows some geodesics to the origin.

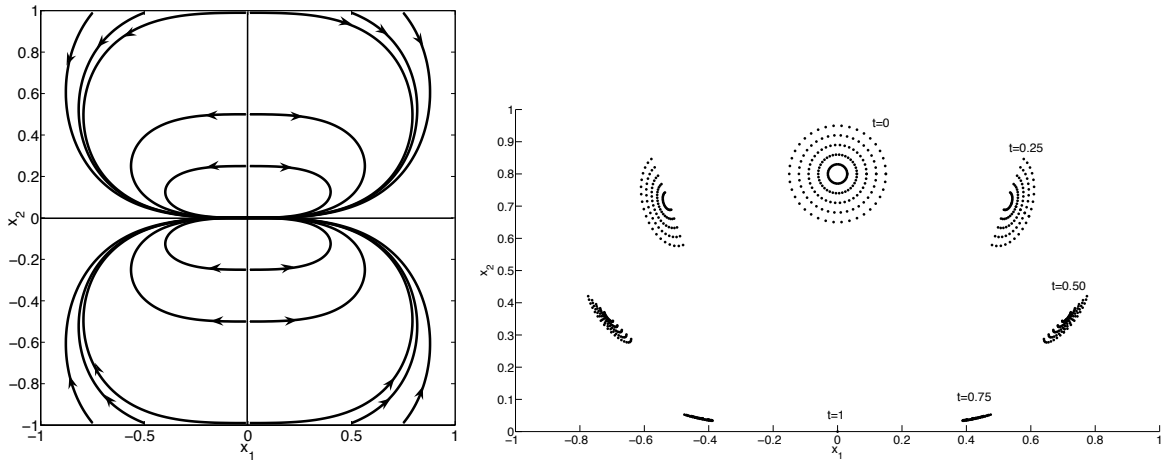


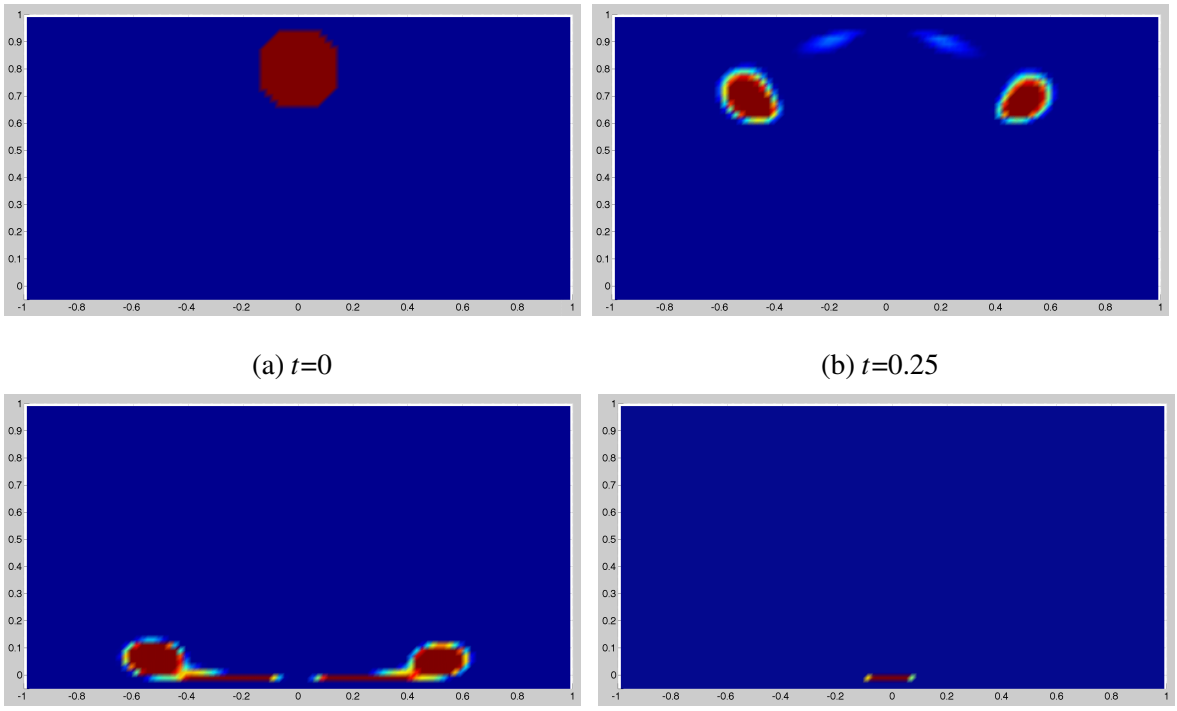
Figure 5.1: (a) Some minimizing geodesics to the origin in the Grushin plane. (b) Analytically computed optimal transport solution between a uniform measure whose support is the disk $\Omega = \{(x,y)|x^2+(y-.8)^2 < .15^2\}$, and a measure concentrated at the origin.

Now consider the optimal transport problem with $c(\mathbf{x},\mathbf{y}) = d^2$ from an initial measure μ_0 to final measure $\mu_1 = \delta_{(0,0)}$. See Fig. 5.1(b) for analytically computed transport in the case in which the initial measure is uniform over a disk.

Using the algorithm developed in Section 5.2, we compute optimal transport for this same case. We divide the phase space $X = [-1,1] \times [-1,1]$ into $m = 100^2$ boxes, and form the corresponding graph $\mathscr{G}$. The resulting solution is shown in Figure 5.2(a)-(d). It can be seen that the computed solution closely follows the analytical solution shown in Fig. 5.1.

Optimal Transport for Unicycle Model

Finally, we consider optimal transport for a three-dimensional non-holonomic system called the "unicycle" model. This system is a toy model for vehicle kinematics, and is used extensively in vehicle path planning and control (Murray and Sastry, 1993; Aicardi *et al.*, 1995). The states are the Cartesian coordinates $(x,y) \in \mathbf{R}^2$ and orientation $\theta \in \mathbf{S}^1$ of the unicycle. The system equations on $M = \mathbf{S}^1 \times \mathbf{R}^2$ are given by



(a) t=0 (b) t=0.25

(c) t=0.75 (d) t=0.95

Figure 5.2: (a)-(d) The optimal transport solution in the Grushin plane between a measure whose support is the disk $\Omega = \{(x,y)|x^2 + (y-.8)^2 < .15^2\}$, and the delta measure at the origin. The parameters are $m = 10^4$, $k = 75$.

$$\dot{\theta} = u_1,$$

$$\dot{x} = u_2 \cos\theta,$$

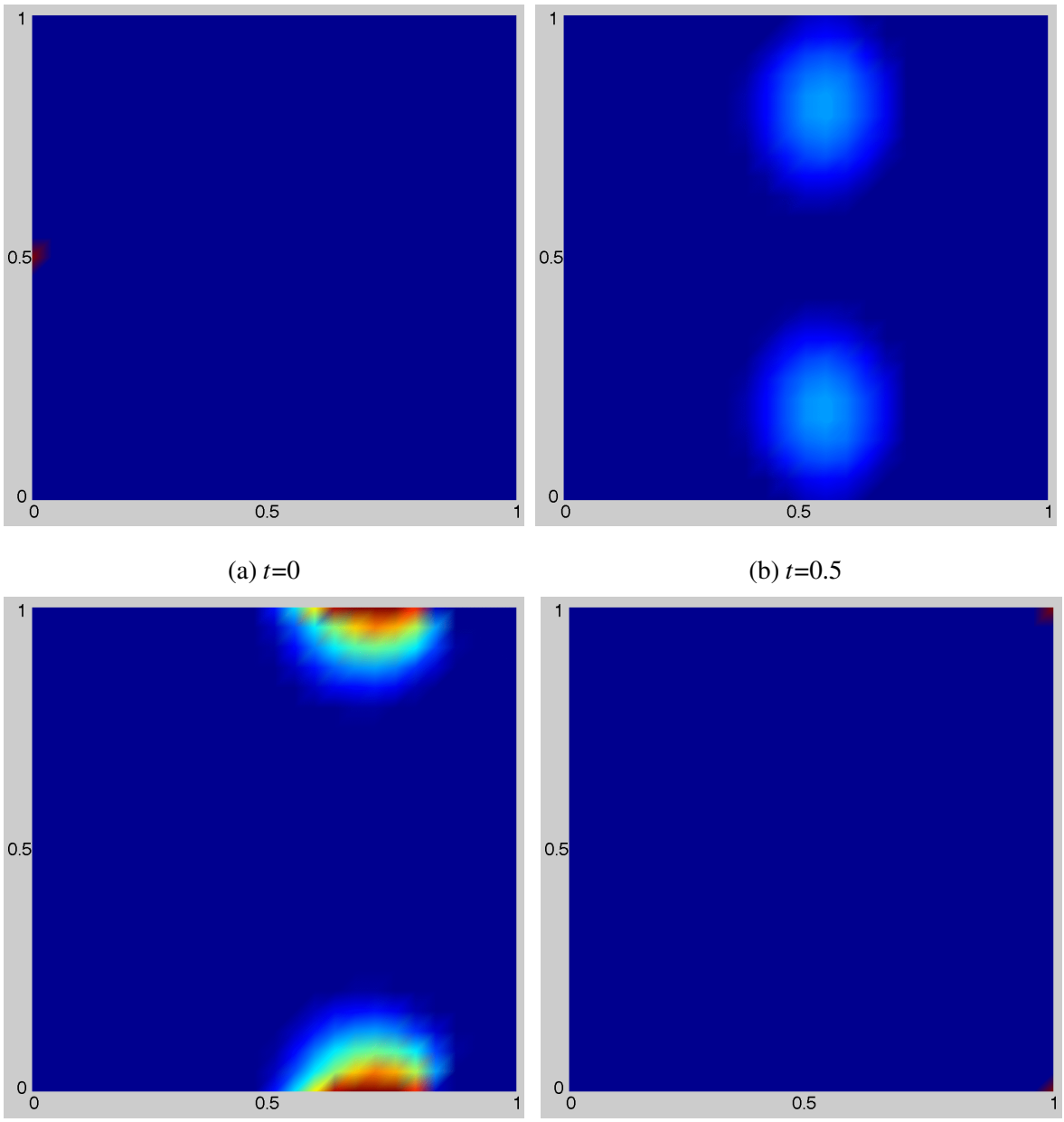
$$\dot{y} = u_2 \sin\theta,$$

where u_1 is the steering speed and u_2 is the translation speed. The above system is a driftless system with control vector fields $\mathbf{g}_1(\theta,x,y) = [1\ 0\ 0]^\intercal$, $\mathbf{g}_2(\theta,x,y) = [0\ \cos\theta\ \sin\theta]^\intercal$. These do not span the tangent space T_xM, but their Lie algebra does, i.e. $Lie_{\mathbf{x}}\left\{\mathbf{g}_i : i \in \{1,2\}\right\} = T_{\mathbf{x}}M$. This can be seen by noting that the Lie bracket $[\mathbf{g}_1,\mathbf{g}_2] = [0\ -\sin\theta\ \cos\theta]^\intercal$ does not lie in $span\{\mathbf{g}_1,\mathbf{g}_2\}$. Hence, this system satisfies condition 1 of Theorem 5.2.7. By Theorem

Figure 5.3: Initial and final measures shown on the (x,y) plane for optimal transport for the unicycle model. The green arrows indicate the third coordinate θ. μ_0 is supported on $(0,0.5,0)$, and μ_1 is supported on $(1,0,0)$ and $(1,1,0)$.

5.2.9, the corresponding optimal transport problem for this driftless system is well-posed.

To study the optimal transport problem for the unicycle model, take the control cost to be quadratic, i.e. $d(z_1,z_2) = \inf_{\mathbb{U}_{z_1}^{z_2}} \int_0^1 \sqrt{u_1^2+u_2^2}dt$. We compute optimal transport solutions for two scenarios. In the first case, μ_0 is chosen to be the uniform measure supported on a box containing $(0,0.5,0)$, and μ_1 is chosen to be the uniform measure supported on a union of boxes containing $(1,0,0)$ and $(1,1,0)$. In the second case, μ_0 is chosen to be the uniform measure supported on a box containing $(0,0.5,0)$, and μ_1 is chosen to be a uniform measure supported on a union of boxes containing $(1,1,\frac{\pi}{2})$ and $(1,0,\frac{3\pi}{2})$. We use $m = 25^3$ boxes to discretize the 3D phase space M, and $t_f = 1$ with $k = 20$ equally-spaced time steps, for both cases. The computation in CVX takes about 6×10^4 seconds. The initial and final measures are depicted in Fig. 5.3. The optimal transport solution is shown in Fig. 5.4. Since the final orientation is prescribed to be along the $x-$axis, this leads to a splitting of the measure half-way in the transport, and steering of the two halves horizontally to their final positions.



(a) t=0 (b) t=0.5

(c) t=0.7 (d) t=1

Figure 5.4: The optimal transport solution for the unicycle model shown in the $x-y$ plane. The grid size is $m = 25^3$, and $k = 20$.

Chapter 6

CONCLUSION AND FUTURE WORK

In this chapter, we conclude this thesis and mention some possible directions for future work.

In Chapter 2, we presented several fundamental results on controllability and stabilizability properties of forward equations of CTMCs that are associated with strongly connected graphs. We proved a sufficient condition for controllability of control-affine systems that extends the classical rank conditions for controllability to the case where the control inputs have non-negativity constraints. We applied this condition to a system in which only a subset of the transition rates are control inputs. We proved asymptotic controllability of distributions that are not strictly positive, with target densities equal to zero for some states. We also characterized the stationary distributions that are stabilizable using time-independent and time-dependent control inputs. Further, we constructed decentralized, density-dependent feedback laws that stabilize the forward equation, with control inputs that equal zero at equilibrium. A possible direction for future work is to generalize the controllability and stabilization results of Chapter 2 to more general CRN models of the form (1.12). Such models are nonlinear even when the control inputs are independent of the probability distributions. Hence, the corresponding stabilization and controllability problem is expected to be much more complicated to address.

In Chapter 2, we also addressed the problem of herding a robotic swarm to a desired distribution among a set of states using a leader agent that produces a repulsive effect on swarm members in its current state. We utilized a mean-field model of the swarm in our approach and constructed a switching feedback controller for the leader agent. We proved that this controller can stabilize the swarm to target probability distributions that

are positive everywhere. Future work will focus on designing feedback laws and optimal control strategies for the leader agent that improve system performance criteria such as the rate of the follower agents' convergence to the target distribution and the robustness of this convergence to disturbances, such as environmental factors (e.g., wind) and inter-agent collisions.

In Chapter 3, we proved controllability properties of a system of advection-diffusion-reaction (ADR) PDEs with zero-flux boundary condition that is defined on certain smooth domains. In contrast to previous work, we established controllability of the PDEs with bounded control inputs. Our approach to establishing controllability using spectral properties of the elliptic operators under consideration is also novel. In our opinion, this approach to proving controllability of ADR PDEs is simpler than methods that have previously been employed in similar works, discussed in Section 1.2. In addition, we provided constructive solutions to the problem of asymptotically stabilizing a class of hybrid-switching diffusion processes (HSDPs) to target non-negative stationary distributions. A possible direction for future work is to extend the arguments in this chapter to the case where the corresponding HSDP has diffusion and velocity control parameters in only a small subset of the discrete behavioral states. Lastly, while diffusive movement by agents was modeled as Brownian motion in this chapter, future work could focus on alternative diffusion models which do not implicitly assume that agents can move from one location to another at arbitrarily large speeds.

To consider scenarios where the dynamics of each agent in a swarm is nonlinear, in Chapter 4 we presented a relaxed version of the optimal transport problem for discrete-time nonlinear systems. We showed that under mild assumptions on the controllability of the original control system, the extended system on the space of measures is controllable. This enabled us to prove the existence of solutions of an optimal transport problem for nonlinear systems evolving in discrete time. A possible direction for future work is to

explore conditions under which deterministic feedback maps exist for the optimal transport problem.

In Chapter 5, we developed a graph-based computational framework for continuous-time optimal transport over nonlinear dynamical systems. In the control systems setting, this framework generalizes the graph-based approximations of the optimal transport problem for single-integrator systems to nonlinear control-affine systems. This is accomplished by exploiting recent work on approximations of infinitesimal generators associated with nonlinear dynamical systems using infinitesimal generators on graphs. The controllability of measures over graphs is related to the connectivity of the "controlled" graph, and is proved to be a consequence of controllability of the underlying control system. This work opens up new directions in the design of efficient feedback control strategies for multi-agent and swarm systems with agents that have nonlinear dynamics.

REFERENCES

Acikmese, B. and D. S. Bayard, "A Markov chain approach to probabilistic swarm guidance", in "American Control Conference (ACC)", pp. 6300–6307 (IEEE, 2012).

Açıkmeşe, B. and D. S. Bayard, "Markov chain approach to probabilistic guidance for swarms of autonomous agents", Asian Journal of Control **17**, 4, 1105–1124 (2015).

Agassounon, W., A. Martinoli and K. Easton, "Macroscopic modeling of aggregation experiments using embodied agents in teams of constant and time-varying sizes", Autonomous Robots **17**, 2-3, 163–192 (2004).

Agrachev, A., U. Boscain, J.-P. Gauthier and F. Rossi, "The intrinsic hypoelliptic Laplacian and its heat kernel on unimodular Lie groups", Journal of Functional Analysis **256**, 8, 2621–2655 (2009).

Agrachev, A. and P. Lee, "Optimal transportation under nonholonomic constraints", Transactions of the American Mathematical Society **361**, 11, 6019–6047 (2009).

Agranovich, M. S., *Sobolev spaces, their generalizations and elliptic problems in smooth and Lipschitz domains* (Springer, 2015).

Aicardi, M., G. Casalino, A. Bicchi and A. Balestrino, "Closed loop steering of unicycle like vehicles via Lyapunov techniques", IEEE Robotics & Automation Magazine **2**, 1, 27–35 (1995).

Albani, D., T. Manoni, D. Nardi and V. Trianni, "Dynamic UAV swarm deployment for non-uniform coverage", in "Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems", pp. 523–531 (International Foundation for Autonomous Agents and Multiagent Systems, 2018).

Ambrosio, L., G. Savaré and L. Zambotti, "Existence and stability for Fokker–Planck equations with log-concave reference measure", Probability Theory and Related Fields **145**, 3-4, 517–564 (2009).

Annunziato, M. and A. Borzì, "Optimal control of probability density functions of stochastic processes", Mathematical Modelling and Analysis **15**, 4, 393–407 (2010).

Annunziato, M. and A. Borzì, "A Fokker–Planck control framework for multidimensional stochastic processes", Journal of Computational and Applied Mathematics **237**, 1, 487–507 (2013).

Annunziato, M. and A. Borzi, "A Fokker–Planck control framework for stochastic systems", EMS Surveys in Mathematical Sciences **5**, 1, 65–98 (2018).

Arendt, W., A. Grabosch, G. Greiner, U. Groh, H. P. Lotz, U. Moustakas, F. Neubrander and U. Schlotterbeck, *One-parameter semigroups of positive operators*, vol. 1184 (Springer, 2006).

Aubin, J.-P. and H. Frankowska, *Set-valued analysis* (Springer Science & Business Media, 2009).

Bagagiolo, F. and D. Bauso, "Mean-field games and dynamic demand management in power grids", Dynamic Games and Applications **4**, 2, 155–176 (2014).

Bakry, D., I. Gentil and M. Ledoux, *Analysis and geometry of Markov diffusion operators*, vol. 348 (Springer Science & Business Media, 2013).

Bandyopadhyay, S., S.-J. Chung and F. Y. Hadaegh, "Probabilistic and distributed control of a large-scale swarm of autonomous agents", IEEE Transactions on Robotics **33**, 5, 1103–1123 (2017).

Bass, R. F. and P. Hsu, "Some potential theory for reflecting brownian motion in holder and lipschitz domains", The Annals of Probability **19**, 2, 486–508 (1991).

Benamou, J.-D. and Y. Brenier, "A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem", Numerische Mathematik **84**, 3, 375–393 (2000).

Bensoussan, A., J. Frehse and P. Yam, *Mean field games and mean field type control theory*, vol. 101 (Springer, 2013).

Berman, A. and R. J. Plemmons, *Nonnegative matrices in the mathematical sciences*, vol. 9 (SIAM, 1994).

Berman, S., Á. Halász, M. A. Hsieh and V. Kumar, "Optimized stochastic policies for task allocation in swarms of robots", IEEE Transactions on Robotics **25**, 4, 927–937 (2009).

Berman, S., V. Kumar and R. Nagpal, "Design of control policies for spatially inhomogeneous robot swarms with application to commercial pollination", in "Robotics and Automation (ICRA), 2011 IEEE International Conference on", pp. 378–385 (IEEE, 2011).

Bertozzi, A. L., T. Laurent and J. Rosado, "Lp theory for the multidimensional aggregation equation", Communications on Pure and Applied Mathematics **64**, 1, 45–83 (2011).

Billingsley, P., *Convergence of Probability Measures* (John Wiley & Sons, 2013).

Blaquiere, A., "Controllability of a Fokker-Planck equation, the Schrödinger system, and a related stochastic optimal control (revised version)", Dynamics and Control **2**, 3, 235–253 (1992).

Bloch, A. M., *Nonholonomic mechanics and control*, vol. 24 (Springer, 2015).

Bodnar, M. and J. J. L. Velazquez, "Derivation of macroscopic equations for individual cell-based models: a formal approach", Mathematical Methods in the Applied Sciences **28**, 15, 1757–1779 (2005).

Bogachev, V. I., *Measure theory*, vol. 1 & 2 (Springer Science & Business Media, 2007).

Bollt, E. M. and N. Santitissadeekorn, *Applied and Computational Measurable Dynamics* (SIAM, 2013).

Bortolussi, L., J. Hillston, D. Latella and M. Massink, "Continuous approximation of collective system behaviour: A tutorial", Performance Evaluation **70**, 5, 317–349 (2013).

Boyd, S., L. El Ghaoui, E. Feron and V. Balakrishnan, *Linear matrix inequalities in system and control theory*, vol. 15 (SIAM, 1994).

Bramanti, M., *An invitation to hypoelliptic operators and Hörmander's vector fields* (Springer, 2014).

Brambilla, M., E. Ferrante, M. Birattari and M. Dorigo, "Swarm robotics: a review from the swarm engineering perspective", Swarm Intelligence **7**, 1, 1–41 (2013).

Breiten, T., K. Kunisch and L. Pfeiffer, "Control strategies for the Fokker- Planck equation", ESAIM: Control, Optimisation and Calculus of Variations **24**, 2, 741–763 (2018).

Bressan, A. and D. Zhang, "Control problems for a class of set valued evolutions", Set-Valued and Variational Analysis **20**, 4, 581–601 (2012).

Caines, P. E., M. Huang and R. P. Malhamé, "Mean field games", in "Handbook of Dynamic Game Theory", edited by G. Z. e. T. Basar, pp. 1–28 (Springer, 2017).

Canuto, C., F. Fagnani and P. Tilli, "A Eulerian approach to the analysis of rendezvous algorithms", in "Proceedings of the 17th IFAC world congress (IFAC08)", pp. 9039–9044 (2008).

Carmona, R. and F. Delarue, *Probabilistic Theory of Mean Field Games with Applications I-II* (Springer, 2018).

Carrillo, J. A., Y.-P. Choi and M. Hauray, "The derivation of swarming models: mean-field limit and Wasserstein distances", in "Collective dynamics from bacteria to crowds", pp. 1–46 (Springer, 2014).

Carrillo, J. A., M. Fornasier, G. Toscani and F. Vecil, "Particle, kinetic, and hydrodynamic models of swarming", in "Mathematical modeling of collective behavior in socio-economic and life sciences", pp. 297–336 (Springer, 2010).

Chapman, A., "Advection on graphs", in "Semi-Autonomous Networks", pp. 3–16 (Springer, 2015).

Chattopadhyay, I. and A. Ray, "Supervised self-organization of homogeneous swarms using ergodic projections of Markov chains", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **39**, 6, 1505–1515 (2009).

Chen, Y., T. T. Georgiou and M. Pavon, "Optimal transport over a linear dynamical system", IEEE Transactions on Automatic Control **62**, 5, 2137–2152 (2017).

Cheng, D., "Controllability of switched bilinear systems", IEEE Transactions on Automatic Control **50**, 4, 511–515 (2005).

Chesi, G., *Domain of attraction: analysis and control via SOS programming*, vol. 415 (Springer Science & Business Media, 2011).

Chib, S. and E. Greenberg, "Understanding the Metropolis-Hastings algorithm", The American Statistician **49**, 4, 327–335 (1995).

Choulli, M. and L. Kayser, "Gaussian lower bound for the Neumann Green function of a general parabolic operator", Positivity **19**, 3, 625–646 (2015).

Clarke, F. H., Y. S. Ledyaev, E. D. Sontag and A. I. Subbotin, "Asymptotic controllability implies feedback stabilization", IEEE Transactions on Automatic Control **42**, 10, 1394–1407 (1997).

Colombo, R. M. and N. Pogodaev, "Confinement strategies in a model for the interaction between individuals and a continuum", SIAM Journal on Applied Dynamical Systems **11**, 2, 741–770 (2012).

Dai Pra, P., "A stochastic control approach to reciprocal diffusion processes", Applied Mathematics and Optimization **23**, 1, 313–329 (1991).

de Oca, M. A. M., E. Ferrante, A. Scheidler, C. Pinciroli, M. Birattari and M. Dorigo, "Majority-rule opinion dynamics with differential latency: a mechanism for self-organized collective decision-making", Swarm Intelligence **5**, 3-4, 305–327 (2011).

Deshmukh, V., K. **Elamvazhuthi**, S. Biswal, Z. Kakish and S. Berman, "Mean-field stabilization of Markov chain models for robotic swarms: Computational approaches and experimental results", IEEE Robotics and Automation Letters **3**, 3, 1985–1992 (2018).

Diestel, J. and A. Spalsbury, *The joys of Haar measure* (American Mathematical Soc., 2014).

Djehiche, B., A. Tcheukam and H. Tembine, "Mean-field-type games in engineering", arXiv preprint arXiv:1605.03281 (2016).

Duprez, M. and A. Perasso, "Criterion of positivity for semilinear problems with applications in biology", Positivity **21**, 4, 1383–1392 (2017).

Eckstein, J. and W. Yao, "Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results", RUTCOR Research Reports **32** (2012).

El Chamie, M., Y. Yu, B. Açıkmeşe and M. Ono, "Controlled Markov processes with safety state constraints", IEEE Transactions on Automatic Control **64**, 3, 1003–1018 (2019).

Elliott, D. L., *Bilinear control systems: matrices in action*, vol. 169 (Springer Science & Business Media, 2009).

Engel, K.-J. and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, vol. 194 (Springer Science & Business Media, 2000).

Eren, U. and B. Açıkmeşe, "Velocity field generation for density control of swarms using heat equation and smoothing kernels", IFAC-PapersOnLine **50**, 1, 9405–9411 (2017).

Ethier, S. N. and T. G. Kurtz, *Markov processes: characterization and convergence*, vol. 282 (John Wiley & Sons, 2009).

Evans, L. C., "Partial differential equations", Graduate Studies in Mathematics **19** (1998).

Fattorini, H. O., *The Cauchy problem*, vol. 13517 (Cambridge University Press, 1984).

Fattorini, H. O., *Infinite dimensional optimization and control theory*, vol. 54 (Cambridge University Press, 1999).

Figalli, A. and L. Rifford, "Mass transportation on sub-riemannian manifolds", Geometric And Functional Analysis **20**, 1, 124–159 (2010).

Filippov, A. F., *Differential equations with discontinuous righthand sides: control systems*, vol. 18 (Springer Science & Business Media, 2013).

Finch, S. R., *Mathematical constants* (Cambridge University Press, 2003).

Finotti, H., S. Lenhart and T. Van Phan, "Optimal control of advective direction in reaction-diffusion population models", Evolution Equations & Control Theory **1**, 1 (2012).

Fleig, A. and R. Guglielmi, "Optimal control of the Fokker–Planck equation with space-dependent controls", Journal of Optimization Theory and Applications **174**, 2, 408–427 (2017).

Florescu, L. C. and C. Godet-Thobie, *Young Measures and Compactness in Measure Spaces* (Walter de Gruyter, 2012).

Foderaro, G., S. Ferrari and T. A. Wettergren, "Distributed optimal control for multi-agent trajectory optimization", Automatica **50**, 1, 149–154 (2014).

Froyland, G., O. Junge and P. Koltai, "Estimating long-term behavior of flows without trajectory integration: The infinitesimal generator approach", SIAM Journal on Numerical Analysis **51**, 1, 223–247 (2013).

Galstyan, A., T. Hogg and K. Lerman, "Modeling and mathematical analysis of swarms of microscopic robots", in "Proceedings 2005 IEEE Swarm Intelligence Symposium, 2005. SIS 2005.", pp. 201–208 (IEEE, 2005).

Gardiner, C., *Stochastic methods*, vol. 4 (springer Berlin, 2009).

Garofalo, N. and D.-M. Nhieu, "Lipschitz continuity, global smooth approximations and extension theorems for sobolev functions in carnot-carathéodory spaces", Journal d'Analyse Mathématique **74**, 1, 67–97 (1998).

Gast, N. and B. Gaujal, "Markov chains with discontinuous drifts have differential inclusion limits", Performance Evaluation **69**, 12, 623–642 (2012).

Geng, J., "Homogenization of elliptic problems with Neumann boundary conditions in non-smooth domains", Acta Mathematica Sinica, English Series **34**, 4, 612–628 (2018).

Gigli, N. and J. Maas, "Gromov–Hausdorff convergence of discrete transportation metrics", SIAM Journal on Mathematical Analysis **45**, 2, 879–899 (2013).

Goebel, R., R. G. Sanfelice and A. R. Teel, *Hybrid dynamical systems: modeling, stability, and robustness* (Princeton University Press, 2012).

Grant, M., S. Boyd and Y. Ye, "CVX: Matlab software for disciplined convex programming", (2008).

Grillo, G., M. Muratori and M. M. Porzio, "Porous media equations with two weights: smoothing and decay properties of energy solutions via Poincaré inequalities", Discrete & Continuous Dynamical Systems - A **33**, 8, 3599–3640 (2013).

Grisvard, P., *Elliptic problems in nonsmooth domains* (SIAM, 2011).

Haghighat, B., R. Thandiackal, M. Mordig and A. Martinoli, "Probabilistic modeling of programmable stochastic self-assembly of robotic modules", in "2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)", pp. 4656–5663 (IEEE, 2017).

Halász, A., M. A. Hsieh, S. Berman and V. Kumar, "Dynamic redistribution of a swarm of robots among multiple sites", in "2007 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)", pp. 2320–2325 (IEEE, 2007).

Hamann, H., *Space-time continuous models of swarm robotic systems: Supporting global-to-local programming*, vol. 9 (Springer Science & Business Media, 2010).

Hamann, H. and H. Wörn, "A framework of space–time continuous models for algorithm design in swarm robotics", Swarm Intelligence **2**, 2-4, 209–239 (2008).

Hernández-Lerma, O. and J. B. Lasserre, *Discrete-Time Markov Control Processes: Basic Optimality Criteria*, vol. 30 (Springer Science & Business Media, 2012).

Hindawi, A., J.-B. Pomet and L. Rifford, "Mass transportation with LQ cost functions", Acta applicandae mathematicae **113**, 2, 215–229 (2011).

Hsieh, M. A., Á. Halász, S. Berman and V. Kumar, "Biologically inspired redistribution of a swarm of robots among multiple sites", Swarm Intelligence **2**, 2-4, 121–141 (2008).

Huang, M., P. E. Caines and R. P. Malhamé, "Large-population cost-coupled LQG problems with nonuniform agents: individual-mass behavior and decentralized ε-Nash equilibria", IEEE Transactions on Automatic Control **52**, 9, 1560–1571 (2007).

Jerison, D. and C. E. Kenig, "The functional calculus for the Laplacian on Lipschitz domains", Journées Equations aux Dérivées Partielles pp. 1–10 (1989).

Jerison, D. S. and A. Sánchez-Calle, "Estimates for the heat kernel for a sum of squares of vector fields", Indiana University mathematics journal **35**, 4, 835–854 (1986).

Karatzas, I. and S. E. Shreve, "Brownian motion", in "Brownian Motion and Stochastic Calculus", pp. 47–127 (Springer, 1998).

Khalil, H. K., *Nonlinear Systems* (Pearson, 2001), 3rd edn.

Khamis, A., A. Hussein and A. Elmogy, "Multi-robot task allocation: A review of the state-of-the-art", in "Cooperative Robots and Sensor Networks 2015", pp. 31–51 (Springer, 2015).

Khapalov, A. Y., *Controllability of partial differential equations governed by multiplicative controls* (Springer, 2010).

Khesin, B., P. Lee *et al.*, "A nonholonomic Moser theorem and optimal transport", Journal of Symplectic Geometry **7**, 4, 381–414 (2009).

Kingston, P. and M. Egerstedt, "Index-free multi-agent systems: An Eulerian approach", IFAC Proceedings Volumes **43**, 19, 215–220 (2010).

Kingston, P. and M. Egerstedt, "Distributed-infrastructure multi-robot routing using a 'h'elmholtz-'h'odge decomposition", in "2011 50th IEEE Conference on Decision and Control and European Control Conference", pp. 5281–5286 (IEEE, 2011).

Klamka, J., "Constrained controllability of nonlinear systems", Journal of Mathematical Analysis and Applications **201**, 2, 365–374 (1996).

Klavins, E., S. Burden and N. Napp, "Optimal rules for programmed stochastic self-assembly.", in "Robotics: Science and Systems", (Philadelphia, PA, 2006).

Krapivsky, P. L. and S. Redner, "Dynamics of majority rule in two-state interacting spin systems", Physical Review Letters **90**, 23, 238701 (2003).

Krishnan, V. and S. Martínez, "Distributed optimal transport for the deployment of swarms", in "2018 IEEE Conference on Decision and Control (CDC)", pp. 4583–4588 (IEEE, 2018).

Lasota, A. and M. Mackey, *Chaos, Fractals and Noise* (Springer-Verlag, New York, 1994).

Lasry, J.-M. and P.-L. Lions, "Mean field games", Japanese Journal of Mathematics **2**, 1, 229–260 (2007).

Lasserre, J. B., D. Henrion, C. Prieur and E. Trélat, "Nonlinear optimal control via occupation measures and LMI-relaxations", SIAM Journal on Control and Optimization **47**, 4, 1643–1666 (2008).

Lee, J. M., *Introduction to smooth manifolds* (Springer, 2001).

Leoni, G., *A first course in Sobolev spaces*, vol. 105 (American Mathematical Society, Providence, RI, 2009).

Lerman, K. and A. Galstyan, "Mathematical model of foraging in a group of robots: Effect of interference", Autonomous Robots **13**, 2, 127–141 (2002).

Lerman, K., A. Galstyan, A. Martinoli and A. Ijspeert, "A macroscopic analytical model of collaboration in distributed robotic systems", Artificial Life **7**, 4, 375–393 (2001).

Lerman, K., C. Jones, A. Galstyan and M. J. Matarić, "Analysis of dynamic task allocation in multi-robot systems", The International Journal of Robotics Research **25**, 3, 225–241 (2006).

Lerman, K., A. Martinoli and A. Galstyan, "A review of probabilistic macroscopic models for swarm robotic systems", in "International Workshop on Swarm Robotics", pp. 143–152 (Springer, 2004).

Li, H., C. Feng, H. Ehrhard, Y. Shen, B. Cobos, F. Zhang, K. Elamvazhuthi, S. Berman, M. Haberland and A. L. Bertozzi, "Decentralized stochastic control of robotic swarm density: Theory, simulation, and experiment", in "Intelligent Robots and Systems (IROS), 2017 IEEE/RSJ International Conference on", pp. 4341–4347 (IEEE, 2017).

Li, P. and S. T. Yau, "On the parabolic kernel of the Schrödinger operator", Acta Mathematica **156**, 1, 153–201 (1986).

Liu, Z., B. Wu and H. Lin, "A mean field game approach to swarming robots control", in "2018 Annual American Control Conference (ACC)", pp. 4293–4298 (IEEE, 2018).

Lunardi, A., *Analytic semigroups and optimal regularity in parabolic problems* (Springer Science & Business Media, 2012).

Maas, J., "Gradient flows of the entropy for finite Markov chains", Journal of Functional Analysis **261**, 8, 2250–2292 (2011).

Markowich, P. A. and C. Villani, "On the trend to equilibrium for the Fokker-Planck equation: an interplay between physics and functional analysis", in "Physics and Functional Analysis, Matematica Contemporanea (SBM) 19", (1999).

Mather, T. W. and M. A. Hsieh, "Distributed robot ensemble control for deployment to multiple sites", Proceedings of Robotics: Science and Systems VII (2011).

Mather, T. W. and M. A. Hsieh, "Synthesis and analysis of distributed ensemble control strategies for allocation to multiple tasks", Robotica **32**, 02, 177–192 (2014).

Matthey, L., S. Berman and V. Kumar, "Stochastic strategies for a swarm robotic assembly system", in "Robotics and Automation, 2009. ICRA'09. IEEE International Conference on", pp. 1953–1958 (IEEE, 2009).

Mayya, S., P. Pierpaoli, G. Nair and M. Egerstedt, "Localization in densely packed swarms using interrobot collisions as a sensing modality", IEEE Transactions on Robotics **35**, 1, 21–34 (2019).

Maz'ya, V., "On the boundedness of first derivatives for solutions to the Neumann–Laplace problem in a convex domain", Journal of Mathematical Sciences **159**, 1, 104–112 (2009).

Mermoud, G., L. Matthey, W. C. Evans and A. Martinoli, "Aggregation-mediated collective perception and action in a group of miniature robots", in "Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 2-Volume 2", pp. 599–606 (International Foundation for Autonomous Agents and Multiagent Systems, 2010).

Mesbahi, M. and M. Egerstedt, *Graph Theoretic Methods in Multiagent Networks* (Princeton University Press, 2010).

Mesquita, A. R. and J. P. Hespanha, "Jump control of probability densities with applications to autonomous vehicle motion", IEEE Transactions on Automatic Control **57**, 10, 2588–2598 (2012).

Mesquita, A. R., J. P. Hespanha and K. Åström, "Optimotaxis: A stochastic multi-agent optimization procedure with point measurements", in "International Workshop on Hybrid Systems: Computation and control", pp. 358–371 (Springer, 2008).

Meyer-Nieberg, P., *Banach lattices* (Springer Science & Business Media, 2012).

Mielke, A., "Geodesic convexity of the relative entropy in reversible Markov chains", Calculus of Variations and Partial Differential Equations pp. 1–31 (2013).

Milutinovic, D. and P. Lima, "Modeling and optimal centralized control of a large-size robotic population", IEEE Transactions on Robotics **22**, 6, 1280–1285 (2006).

Milutinovic, D. L. and P. U. Lima, *Cells and robots: modeling and control of large-size agent populations* (New York, 2007).

Minc, H., *Nonnegative Matrices* (Wiley-Interscience, 1988).

Murray, R. M. and S. S. Sastry, "Nonholonomic motion planning: Steering using sinusoids", IEEE Transactions on Automatic Control **38**, 5, 700–716 (1993).

Napp, N., S. Burden and E. Klavins, "Setpoint regulation for stochastically interacting robots", Autonomous Robots **30**, 1, 57–71 (2011).

Nhieu, D.-M., "The neumann problem for sub-Laplacians on Carnot groups and the extension theorem for Sobolev spaces", Annali di Matematica Pura ed Applicata **180**, 1, 1–25 (2001).

O'Donoghue, B., E. Chu, N. Parikh and S. Boyd, "Operator splitting for conic optimization via homogeneous self-dual embedding", arXiv preprint arXiv:1312.3039 (2013).

Oh, K.-K., M.-C. Park and H.-S. Ahn, "A survey of multi-agent formation control", Automatica **53**, 424–440 (2015).

Ouhabaz, E.-M., *Analysis of heat equations on domains (LMS-31)* (Princeton University Press, 2009).

Pedersen, G. K., *Analysis Now*, vol. 118 (Springer Science & Business Media, 2012).

Pilipenko, A., *An Introduction to Stochastic Differential Equations with Reflection*, vol. 1 (Universitätsverlag Potsdam, 2014).

Porretta, A., "On the planning problem for the mean field games system", Dynamic Games and Applications **4**, 2, 231–256 (2014).

Pratt, S. C., "Quorum sensing by encounter rates in the ant Temnothorax albipennis", Behavioral Ecology **16**, 2, 488–496 (2005).

Prorok, A., N. Corell and A. Martinoli, "Multi-level spatial modeling for stochastic distributed robotic systems", The International Journal of Robotics Research **30**, 5, 574–589 (2011).

Prorok, A., M. A. Hsieh and V. Kumar, "The impact of diversity on optimal control policies for heterogeneous robot swarms", IEEE Transactions on Robotics **33**, 2, 346–358 (2017).

Prorok, A. and V. Kumar, "A macroscopic privacy model for heterogeneous robot swarms", in "International Conference on Swarm Intelligence", pp. 15–27 (Springer, 2016).

Puterman, M. L., *Markov decision processes: discrete stochastic dynamic programming* (John Wiley & Sons, 2014).

Ramachandran, R. K., K. **Elamvazhuthi** and S. Berman, "An optimal control approach to mapping GPS-denied environments using a stochastic robotic swarm", in "Robotics Research", pp. 477–493 (Springer, 2018).

Reina, A., G. Valentini, C. Fernández-Oto, M. Dorigo and V. Trianni, "A design pattern for decentralised decision making", PLoS One **10**, 10, e0140950 (2015).

Rifford, L., *Sub-Riemannian geometry and optimal transport* (Springer Science & Business Media, 2014).

Robin, C. and S. Lacroix, "Multi-robot target detection and tracking: taxonomy and survey", Autonomous Robots **40**, 4, 729–760 (2016).

Robinson, D., *Elliptic Operators and Lie Groups*, vol. 100 (Clarendon Press, 1991).

Roth, G. and W. H. Sandholm, "Stochastic approximations with constant step size and differential inclusions", SIAM Journal on Control and Optimization **51**, 1, 525–555 (2013).

Roubíček, T., *Nonlinear partial differential equations with applications*, vol. 153 (Springer Science & Business Media, 2013).

Sanders, J. A., F. Verhulst and J. Murdock, *Averaging methods in nonlinear dynamical systems*, vol. 59 (Springer, 2007).

Schmüdgen, K., *Unbounded self-adjoint operators on Hilbert space*, vol. 265 (Springer Science & Business Media, 2012).

Seeja, G., A. Selvakumar and V. B. Hency, "A survey on swarm robotic modeling, analysis and hardware architecture", Procedia Computer Science **133**, 478–485 (2018).

Solomon, J., R. Rustamov, L. Guibas and A. Butscher, "Continuous-flow graph transportation distances", arXiv preprint arXiv:1603.06927 (2016).

Stroock, D. W., "Logarithmic Sobolev inequalities for Gibbs states", in "Dirichlet forms", pp. 194–228 (Springer, 1993).

Sun, Z., S. S. Ge and T. H. Lee, "Controllability and reachability criteria for switched linear systems", Automatica **38**, 5, 775–786 (2002).

Sussmann, H. J., "A general theorem on local controllability", SIAM Journal on Control and Optimization **25**, 1, 158–194 (1987).

Talay, D., "Numerical solution of stochastic differential equations", (1994).

Tembine, H., "Energy-constrained mean field games in wireless networks", Strategic Behavior and the Environment **4**, 1, 99–123 (2014).

Elamvazhuthi, K., C. Adams and S. Berman, "Coverage and field estimation on bounded domains by diffusive swarms", in "Decision and Control (CDC), 2016 IEEE 55th Conference on", pp. 2867–2874 (IEEE, 2016).

Elamvazhuthi, K. and S. Berman, "Nonlinear generalizations of diffusion-based coverage by robotic swarms", in "2018 IEEE Conference on Decision and Control (CDC)", pp. 1341–1346 (IEEE, 2018).

Elamvazhuthi, K., S. Biswal and S. Berman, "Mean-field stabilization of robotic swarms to probability distributions with disconnected supports", in "American Control Conference (ACC), 2018", pp. 885–892 (IEEE, 2018a).

Elamvazhuthi, K. and P. Grover, "Optimal transport over nonlinear systems via infinitesimal generators on graphs", Journal of Computational Dynamics **5**, 1&2, 1–32 (2018).

Elamvazhuthi, K., P. Grover and S. Berman, "Optimal transport over deterministic discrete-time nonlinear systems using stochastic feedback laws", IEEE Control Systems Letters **3**, 1, 168–173 (2018b).

Elamvazhuthi, K., M. Kawski, S. Biswal, V. Deshmukh and S. Berman, "Mean-field controllability and decentralized stabilization of Markov chains", in "Decision and Control (CDC), 2017 IEEE 56th Annual Conference on", pp. 3131–3137 (IEEE, 2017a).

Elamvazhuthi, K., H. Kuiper and S. Berman, "Controllability to equilibria of the 1-d Fokker-Planck equation with zero-flux boundary condition", in "IEEE Conference on Decision and Control (CDC)", pp. 2485–2491 (IEEE, 2017b).

Elamvazhuthi, K., H. Kuiper and S. Berman, "PDE-based optimization for stochastic mapping and coverage strategies using robotic ensembles", Automatica **95**, 356–367 (2018c).

Elamvazhuthi, K., H. Kuiper, M. Kawski and S. Berman, "Bilinear controllability of a class of advection-diffusion-reaction systems", IEEE Transactions on Automatic Control **64**, 6, 2282–2297 (2019).

Topaz, C. M., A. L. Bertozzi and M. A. Lewis, "A nonlocal continuum model for biological aggregation", Bulletin of Mathematical Biology **68**, 7, 1601 (2006).

Tröltzsch, F., *Optimal control of partial differential equations: theory, methods, and applications*, vol. 112 (American Mathematical Society, 2010).

Ulam, S. M., *Problems in modern mathematics* (Courier Corporation, 2004).

Vaidya, U., P. G. Mehta and U. V. Shanbhag, "Nonlinear stabilization via control Lyapunov measure", IEEE Transactions on Automatic Control **55**, 6, 1314–1328 (2010).

Valentini, G., *Achieving Consensus in Robot Swarms: Design and Analysis of Strategies for the best-of-n Problem*, vol. 706 (Springer, 2017).

Valentini, G., E. Ferrante and M. Dorigo, "The best-of-n problem in robot swarms: Formalization, state of the art, and novel perspectives", Frontiers in Robotics and AI **4**, 9 (2017).

Varopoulos, N. T., L. Saloff-Coste and T. Coulhon, *Analysis and geometry on groups*, vol. 100 (Cambridge university press, 2008).

Vázquez, J. L., *The porous medium equation: mathematical theory* (Oxford University Press, 2007).

Villani, C., *Topics in optimal transportation*, no. 58 (American Mathematical Society, 2003).

Villani, C., *Optimal Transport: Old and New*, vol. 338 (Springer Science & Business Media, 2008).

Wilson, S., T. P. Pavlic, G. P. Kumar, A. Buffin, S. C. Pratt and S. Berman, "Design of ant-inspired stochastic control policies for collective transport by robotic swarms", Swarm Intelligence **8**, 4, 303–327 (2014).

Yin, G. and C. Zhu, *Hybrid switching diffusions: properties and applications*, vol. 63 (Springer New York, 2010).

Young, L. C., *Lecture on the Calculus of Variations and Optimal Control Theory*, vol. 304 (American Mathematical Soc., 1980).

Yu, G., J. Hu, C. Zhang, L. Zhuang and J. Song, "Short-term traffic flow forecasting based on Markov chain model", in "Intelligent Vehicles Symposium, 2003. Proceedings. IEEE", pp. 208–212 (IEEE, 2003).

Zhang, F., A. L. Bertozzi, K. **Elamvazhuthi** and S. Berman, "Performance bounds on spatial coverage tasks by stochastic robotic swarms", IEEE Transactions on Automatic Control **63**, 6, 1563–1578 (2018).

Ziemer, W. P., *Weakly differentiable functions: Sobolev spaces and functions of bounded variation*, vol. 120 (Springer Science & Business Media, 2012).