MODELING BIO-NETWORKS AT MULTIPLE SCALES

By

SEAN T. MCQUADE

A dissertation submitted to the

Graduate School Camden

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Doctor of Philosophy

Graduate Program in Computational and Integrative Biology

Written under the direction of Professor Benedetto Piccoli

And approved by

—————————————————

Benedetto Piccoli

—————————————————

Jongmin Nam

—————————————————

Mahesh Nerurkar

—————————————————

Karim Azer

Camden, New Jersey

May 2019

ABSTRACT OF THE DISSERTATION

Modeling Bio-Networks in Multiple Scales

by SEAN MCQUADE

Dissertation Director:

Benedetto Piccoli

A network facilitates the description of selective interactions among the variables of a system. In this work, networks are used to depict selective interactions between molecules, cells, and agents. This research leverages the structure of a network to model biological systems using Ordinary Differential Equations.

The first bio-network investigated is a metabolic network (the nano scale). Metabolism can be captured as a directed graph of nodes and edges. The nodes represent bio-molecules or metabolites, and each edge corresponds to a chemical reaction in which the nodes at the tail of the edge are reactants and nodes at the head of the edge are products. The goal is to develop a methodology to accurately simulate large networks. This methodology has been named Linear-In-Flux-Expressions (briefly LIFE).

The second class of bio-network investigated is a cell lineage (the micro scale). This is a network of daughter cells from an embryo. The goal is to develop an analytical procedure which can be used on data regarding potential *cis*-regulatory modules(briefly CRMs) to determine which are active CRMs, as well as where (which cell in the organism) and when (which cell generation) a CRM was active. From this analysis we predict how perturbations of spatial activity will impact the data, and confirm predictions with simulation.

The third class of bio-network investigated is a collection of interacting agents (the macro scale). In opinion formation models, these agents often represent a multidimensional opinion held by an intelligent organism. The goal is to model the evolution of opinions over time as they are influenced by other opinions. Models of this type study the emergence of global patterns driven by local interactions. This work on opinion formation models has two aims: 1. construct a mathematical framework to define classical opinion formation models on the more complex state space of a general compact Riemannian manifold, and 2. investigate the effects of dynamics which govern how influential each agent is.

Acknowledgements

Thanks to the invaluable support from my family and adviser.

Contents

Introduction

Biological phenomena at different scales such as 1. The metabolism of an organism (the nano scale), 2. Cell proliferation and differentiation (the micro scale), and 3. Selective local interactions among intelligent animals leading to emergent behavior (the macro scale). The systems can all be studied by organizing the interacting variables according to a network in this thesis.

In this work we study graphs defined from a given network. Each chapter corresponds to a biological network of different scale. In the nano scale, the graph will indicate reactants and products of a chemical reaction; in the micro scale the graph shows a cell lineage; in the macro scale the graph indicates which agents influence one another. This object is a simple way of describing essential information for the selective interactions studied in this work.

Metabolic systems

Ordinary Differential Equations (briefly ODEs) are used to model metabolism. Additional parameters must be determined in order to study the system. The evolution of the system is determined by two vectors; the vector of fluxes and initial metabolite levels. Among these parameters are the speed of each reaction which we call the flux, and the initial metabolite levels. The flux is a vector of values, one for each edge, and the vector of initial metabolites along with the structure of the network will determine the evolution of the system. Researchers are increasingly implementing pharmacology simulators for early drug discovery in order to reduce costs and increased accuracy. These simulators have an arguably more valuable goal, which is to predict human response to treatment at the resolution of groups of patients with similar phenotype. The further step is the holy grail of personalized medicine; this is being able to establish results at the resolution of the individual patient.

With Quantitative Systems Pharmacology(QSP) and a new methodology called linear-in-flux-expressions, we can elucidate how important parameters of the model are correlated. This has the effect of fewer independent parameters and small deviation among similar virtual patients. This is also advantageous for modeling very large metabolic networks where data is incomplete.

Cis-regulatory modules

Cis-regulatory modules(briefly CRMs) are fragments of DNA which work as an on switch to activate a collection of genes in a cell at a particular time. Researchers study the process of gene regulation by observing the presence of CRMs in the genome of an organism. CRMs can tell us which genes work together to manifest different cell types and accomplish other essential biological processes. Generally CRMs activate genes that are relatively nearby on a strand of DNA, so they contain information about potential genes that they regulate. The traditional methods of studying CRMs are time consuming and tedious because one CRM is studied at a time. A high-throughput experiment has potential to analyze many CRMs at the same time. By introducing many DNA fragments (transfected DNA) that may be CRMs to an organism one can observe if any of those fragments produced RNA. By labeling these fragments, proper analysis allows us to study spatial and temporal CRM expression in one large experiment.

The structure of the cell lineage allows us to deduce the time and place of an incorporation event of a putative CRM. The time resolution is enough to provide the cell generation of the event, and the spatial resolution of this procedure is enough to yield a group of cells, or in some cases a single cell. The analytical procedure along with information contained in the cell lineage allows this deduction. The cell lineage network can be specified by a graph where the nodes at the head of the edge are daugter cells, and the cell group at the tail are parent cells. To refine the

deduction technique the cell lineage of the purple sea urchin was recreated in matlab, and a simulation was built to model a practical experiment. The simulation of the experiment could be run many times to observe how incorporation events in specific target cells (specified a priori) affected the outomes. Observing the results with different sets of target cells revealed the effects of different targets. In the cell lineage, the different growth speed of some cells was necessary to deduce spatial information. We used the cell lineage given by [1] to test our procedure.

Opinion formation models

In opinion formation models, there are many ways to define how an agent interacts with others. The interaction network of such a model informs us which agents experience dynamics from a given agent. This network is an especially useful when we discuss topological opinion models versus metric ones. In a topological model, an agent only interacts with a specified number of closest other agents, and in a metric model, an agent only interacts with other agents that are within a given distance.

Researchers study the emergence of a group's global behavior from local interactions. Global configurations such as consensus, alignment, and clustering of agents and the local interactions that generate these configurations common considerations in the field of opinion formation models. We discuss a framework to understand these interactions for models defined on compact Riemannian manifold. A Riemannian manifold is a more general state space than euclidean space, and there are challenges to defining consistent interaction dynamics. These challenges and approaches to overcome them are discussed in detail in sections 3.1 to 3.6.

The remaining sections of chapter three discuss metric opinion models for which the interaction weight for each agent may change with time. These weights correspond to an edge in the graph of the interaction network. We describe four different time varying models in detail, and show how they can be applied. We explain how their

behavior differs from classical opinion models, such as the well known Hegselmann-Krause model [2] and provide simulations to show how these models impact the chance and speed of consensus. Then, we explore emergent behavior that results from heterophilious dynamics.

Chapter 1

Linear-in-flux-expressions

1.1 Introduction to Linear-in-flux-expressions

Researchers and pharmaceutical companies have recognized the advantages of using in silico models to replace various practical experiments. In the case of drug development, their are at least two main advantages, 1. In silico models are much less expensive to perform than animal drug trials. These experiments are expensive in both time and money. 2. In silico models are becoming more advanced, and will become more accurate in predicting the human response to a treatment, possibly able to achieve predicting different responses of individuals based on patient physiology. Such personalized therapies are more easily classified as safe and effective for an invididual patient [3, 4, 5]. Quantitative systems pharmacology (QSP) is attempting to simulate a patients metabolism to predict the safety and effectiveness of potential drugs in early phase drug discovery. Parameters of the model are informed by a combination of different data sources such as clinical studies with human patients and animal models in the absence of human data. In a collaboration between the Center for Computational and Integrative Biology at Rutgers Camden and A French multinational pharmaceutical company, Sanofi, we have developed a new approach to

QSP models called Linear-In-Flux-Expressions (briefly LIFE). Using LIFE, one can determine how key parameters of the model are structurally related. Understanding the structural relationship between these parameters brings us closer to more rigorous parameter estimation and calibration of QSP models. This is especially important for very large metabolic networks becuase the large number of parameters relative to available data presents obstacles to traditional model fitting paradigms useful for smaller models [6, 4]. By including detail in the model about the behavior and kinetics of the biochemical pathways that will be directly targeted by the treatment, and connecting these pathways to clinical outcomes, QSP models are an important tool for identifying components of the biological background that will alter the effectiveness of a drug. In this way, QSP models can be used to investigate new targets for treatment, and suggest potential combination therapies that would improve treatment.

In this chapter we explain how to implement LIFE methodology on a given metabolic network. The network must be defined as a directed graph first where the nodes are metabolites and the edges are metabolic reactions. Each edge has a flux assigned to it which represents the speed of the chemical reaction. LIFE methodology is demonstrated on a portion of the human cholesterol metabolism network. Then we explain how to generate a population of virtual patients with a specified patient type and how this differs from traditional methods. Results about general networks are given which provides a deeper understanding of metabolic networks, such as the dimensionality of important subspaces of the flux space.

1.2   Methodology

The technique presented here is referred to as the "Linear-In-Flux-Expressions" method (briefly LIFE). The LIFE method begins by precisely defining a network of biochemical reactions (metabolism) as a graph, where the nodes are reactants/products of

metabolic reactions; the edge labels represent reaction rate constants. We describe a system of Ordinary Differential Equations (ODEs) which governs the quantities of biochemical molecules over time.

A general system of ODEs which governs the quantities of biochemical compounds (x) and fluxes (f) is written as

$$\frac{dx}{dt} = F(x, f),$$
$$\frac{df}{dt} = G(x, f).$$

The dynamics of the state of the compounds is a relatively fast process compared to the dynamics of fluxes in a biochemical system [9, 10]. We can assume $G(x, f) \approx 0$. In the LIFE method, we consider

$$\frac{dx}{dt} = S(x)f \quad x \in \mathbb{R}^n, f \in \mathbb{R}^m. \tag{1.1}$$

$S(x)$ is called the *Stoichiometric Matrix*. Note that the system is linear in fluxes, but $S(x)$ is not necessarily linear in x. The nullspace of $S(x)$ is denoted by $K(S(x))$ (or $K(x)$ for brevity), it is a subspace of the space of fluxes, and a mapping of the state to the corresponding nullspace can be written $x \in \mathbb{R}^n \to K(x) \in \mathbb{R}^m$.

$$K(x) = \{f : S(x)f = 0\}.$$

In Figure 1.3, a simple graph of a metabolic network is shown. In this network are six metabolites $X = x_i|\{i = 1, 2, ..., 6\}$. $\dot{x}_i$ indicates the time derivative of metabolite, x_i. While the system itself can be generally nonlinear, it is linear with respect to the flux, and we represent the system as a matrix multiplied by a vector of fluxes.

$$\dot{X} = \begin{pmatrix} 1 & 0 & 0 & -x_1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & -x_2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & -x_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & x_1 & x_2 & x_3 & -x_4 & -x_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & x_4 & 0 & -x_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_4 & x_5 & -x_6 \end{pmatrix} \cdot \begin{pmatrix} f_1 \\ f_2 \\ f_3 \\ f_4 \\ f_5 \\ f_6 \\ f_7 \\ f_8 \\ f_9 \\ f_{10} \end{pmatrix} \quad (1.2)$$

In Figure 1.3, the fluxes inside rectangles (f_1, f_2, f_3) represent constant source terms, whereas those in circles are the rates of first order reactions. Specifically, the amount of x_1 (in units of nmol) molecules increases at a rate of f_1 (in units of nmol/h). This gives us the first term on the right hand side (RHS) of $\dot{x}_1$. The other term in $\dot{x}_1$ is "$x_1 \cdot f_4$." This term represents the conversion of metabolite x_1 into x_4 at the rate f_4. All six equations $\dot{x}_1, ..., \dot{x}_6$ govern the dynamics of our example system. We write this system as:

$$\dot{X} = S(X) \cdot f$$

$S(X)$ is a matrix dependent on metabolite values, X. $S(X)$ is referred to as the "Stoichiometric matrix". Traditionally, this type of system is modeled differently, whereby the system is written $\dot{X} = S(f) \cdot X$, as in [7].

For this example, $S(X)$ is a 6×10 matrix. The entries of this matrix are either real numbers or algebraic expressions of variables (representing metabolite values), and f is a vector composed of all ten rate constants, called fluxes: $(f_1, ..., f_{10})$ (Fig. 1.3). A similar method for modeling biochemical networks is explained in [8].

We may write this system of six ODEs from our example matrix in (1.2). One advantage of writing our system this way is we can calculate the nullspace for large systems, as in [8]. We calculate the nullspace of $S(x)$ for figure 1.3 in equation (1.3).

$$
\hat{f} = a_1 \begin{bmatrix} x_6 \\ 0 \\ 0 \\ \frac{x_6}{x_1} \\ 0 \\ 0 \\ 0 \\ \frac{x_6}{x_4} \\ 0 \\ 1 \end{bmatrix} + a_2 \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{x_5}{x_4} \\ \frac{-x_5}{x_4} \\ 1 \\ 0 \end{bmatrix} + a_3 \begin{bmatrix} -x_3 \\ 0 \\ x_3 \\ \frac{-x_3}{x_1} \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} + a_4 \begin{bmatrix} -x_2 \\ x_2 \\ 0 \\ \frac{-x_2}{x_1} \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} a_1 x_6 - a_3 x_3 - a_4 x_2 \\ a_4 \cdot x_2 \\ a_3 \cdot x_3 \\ \frac{a_1 x_6}{x_1} - \frac{a_3 x_3}{x_1} - \frac{a_4 x_2}{x_1} \\ \mathbf{a}_4 \\ \mathbf{a}_3 \\ \frac{a_2 x_5}{x_4} \\ \frac{a_1 x_6}{x_4} - \frac{a_2 x_5}{x_4} \\ \mathbf{a}_2 \\ \mathbf{a}_1 \end{bmatrix} \tag{1.3}
$$

In equation (1.3) we see the nullspace for our matrix from the system shown in Figure 1.3. Note that there are four free variables, a_1, a_2, a_3, a_4, in this nullspace for any fixed set of metabolite levels, X. We call these free variables, "core parameters."

Let $\hat{f}_i$ be the ith element of $\hat{f}$

$$\hat{f}_5 = a_4,$$

$$\hat{f}_6 = a_3,$$

$$\hat{f}_9 = a_2, \tag{1.4}$$

$$\hat{f}_{10} = a_1.$$

Writing our system as (1.3) allows one to identify fluxes in the model that are independent, as well as fluxes that are dependent on other fluxes. The independent subset of fluxes becomes the reduced set of parameters. The system evolves according to the randomly sampled fluxes, and the evolution of metabolites values is called a trajectory. Fewer parameters will reduce the variation between trajectories, thus providing a more precise prediction of metabolite levels.

Remark: The basis (four vectors) shown in (1.3) is not unique, however, we view the nullspace of the system with respect to these fluxes.

With analysis of the nullspace of $S(x)$ in (1.1), we investigate fluxes that keep the system in equilibrium. Generally, biological systems are composed of subsystems which have a faster dynamics than the larger system, and it is reasonable to assume that the state x is not far from equilibrium [10].

We use generalized idea of a *directed graph*, where we allow inflows to a graph from a general source, and outflows from the graph to a general sink. We say *graph* for brevity in this chapter.

Definition 1 *The* indegree *of a node is the number of directed edges for which the node is the terminal vertex. The* outdegree *of a node is the number of directed edges for which the node is the initial vertex.*

Definition 2 *A* source *of a graph is a directed edge with a node representing a com-*

pound only at the terminal end; the initial vertex has outdegree 1, indegree 0 and is not represented in our system. This is equivalent to an exchange reaction entering the system [8]. A sink *of a graph is a directed edge with a node representing a compound only at the initial end. The terminal vertex has indegree 1 and outdegree 0, and is not represented in the system. This is equivalent to an exchange reaction leaving the system.*

Definition 3 *The stoichiometric matrix depends on a state variable x, and is denoted $S(x)$, (or S for brevity). $(S)_{ij} = s_{ij}$ can be defined from a graph. If the edge f_j has initial vertex x_i and terminal vertex x_k, then*

$$\begin{cases} s_{ij} = -x_i \\ s_{kj} = x_i. \end{cases}$$

If the edge f_j is a source with terminal vertex x_k,

$$s_{kj} = 1.$$

If the edge f_j is a sink with initial vertex x_i,

$$s_{ij} = -x_i.$$

Definition 4 *A* Weakly Connected Component *of a graph is a maximum subgraph such that an undirected path exists between every pair of nodes. A graph is weakly connected if there exists such a path between every pair of nodes.*

Definition 5 *The* grassmannian *$G(k, V)$ is the k-dimensional linear subspace of a space of dimension V.*

The nullspace of dimension d of a system is a subset of the $Grassmannian(d, m)$. We study the map $x \to K(x)$ as it relates to perturbations of stable systems.

Lemma 1 *Let x be the initial state for system (1.1), $f \in K(x)$. Assume that all eigenvalues of the jacobian matrix of the system at x have negative real part. Then there exists $\varepsilon > 0$ such that if $y = x + \delta$, $|\delta| < \varepsilon$, $y(\cdot)$ is the solution starting at y*

$$\lim_{t \to +\infty} y(t) \in K^{-1}(f).$$

Proof 1 *The assumption on the eigenvalues of the jacobian matrix imply the system is Lyapunov stable at x, see theorem 4.1.2 of [11], which implies $\lim_{t \to +\infty} S(y(t))f = 0$; we conclude $\lim_{t \to 0} y(t) \in K^{-1}(f)$.*

Lemma 1 motivates our investigation of $K^{-1}(f)$ and will determine candidate states to which a stable system will return after a perturbation.

Two main problems are investigated in this chapter.

- Problem 1: Define a methodology that reveals correlations among the fluxes of a system based on a choice of key independent fluxes.

- Problem 2: Investigate structural properties of metabolic networks, such as the intersections of the nullspace for different metabolic states, i. e. for $x \neq y$ determine the intersection of the nullspaces $K(x) \cap K(y)$.

- Problem 3: Given $x, f \in K(x)$ study $K^{-1}(f)$ to understand perturbations of $K(x)$.

By exploring the map $x \to K(x)$ we will characterize $K(x) \cap K(\tilde{x})$ for some perturbation of the state, $\tilde{x} \neq x$. We show that for a fixed state x, $K(x) \cap K(\tilde{x})$ can have any dimension depending on $\tilde{x}$. That is, there exists $\tilde{x}$ such that $dim(K(x) \cap K(\tilde{x})) = 1$, and for some other $\tilde{x}$, $dim(K(x) \cap K(\tilde{x})) = 2$, etc.

Traditional approach for virtual patient generation

To demonstrate the advantages of the LIFE method, we contrast our approach for generating virtual patients to a standard approach used in the literature. In both approaches, we begin with a parameterization of the model that has been calibrated to the average patient response. In the traditional approach, key fluxes that have been identified to prominently contribute to variability in the model are sampled to create a patient population. Typically, bounds are placed on the parameters of interest based on experiments or guidelines in the literature, and some type of optimization or weighting is used to narrow down which parameter values within these bounds are feasible based on how simulated trajectories compare to clinical data [12, 13, 14]. In this work, we aim to evaluate the performance of the LIFE method, rather than to validate the QSP model to which this method is applied, so we use a simpler traditional VP method for comparison. The traditional approach we implement does not utilize clinical data to tailor results, but generates all results proscribed by the model based on the parameter values given, and constrained by physiological knowledge of plausible output ranges. We sample key flux values from lognormal distributions, using the optimized parameter values as the mean, and the same sigma value (set to 0.25 in these simulations). Once the key flux values are chosen, the simulation is run to generate VP response.

LIFE method approach

We again start with the objective of introducing variability into the identified key fluxes of the model, but do this by directly varying the core fluxes in the nullspace which control the variability of the key fluxes (4). If multiple fluxes in the nullspace control variability for a key flux, we choose to vary the nullspace flux with the largest effect. In inducing variability in the nullspace flux parameters, our goal is to sample randomly from a lognormal distribution. To sample from a distribution that reflects

variability in the clinical data, our approach needs to envelope the observed variability, while remaining within known physiological bounds and maintaining positivity for all fluxes.

A sample flux s, is called "suitable" if it satisfies the inequality:

$$0.5 \times \bar{f} \le s \le 1.5 \times \bar{f} \tag{1.5}$$

where $\bar{f}$ is the average value of the flux we are sampling.

For a parameter, our goal is to find a standard deviation for our sampling distribution such that it will generate a "suitable" s with a 95% confidence interval. A standard deviation for a parameter is found when all fluxes dependent on the parameter generate "suitable" sample at least 95% of the time.

To evaluate the standard deviation of a parameter (a_1) we use a simplified "importance sampling." [15] We choose equally spaced values from the 2.5th percentile of the lognormal distribution to the 97.5th percentile. From these chosen values, we observe which fluxes, which are dependent on a_1, are "suitable" fluxes. This is depicted in Figure 1.4. If the total number of "suitable" fluxes dependent on a_1 is below 95%, we decrease our standard deviation, σ, effectively narrowing the sampling distribution; if the number of "suitable" fluxes is above 95%, we increase σ.

For example, in (1.3) the fluxes dependent on a_1 are $\hat{f}_1, \hat{f}_4, \hat{f}_8,$ and $\hat{f}_{10}$

For each VP, extra bounds may be needed to ensure that all fluxes are positive. The nine fluxes we sample determine other flux values to guarantee that our system has a steady state.

The standard deviations in table 1 do provide more than 90% confidence that all fluxes are between $0.5\bar{f}$ and $1.5\bar{f}$, however, for simple cases we add several strict bounds to be sure a flux isn't sampled in a way that will cause another flux to be negative (Supp Table 1).

We use this algorithm separately for each of nine fluxes that we intend to sample. This method generated the following standard errors shown as $\frac{\sigma}{\mu}$ in Table 1. Once we have determined the distributions of each of the core nullspace fluxes to use for our virtual patient population, we use the following steps to run simulations of the virtual patient population:

1. Sample each core nullspace flux value from its distribution, to generate values for these fluxes for each virtual patient.

2. Calculate values for all fluxes in the model, using the VP specific core flux values, and the constant values for the remaining fluxes in the nullspace and the steady state species values.

3. Run the simulation for each VP.

1.3 Results

We have developed the LIFE method to create virtual patient populations which are informed by the biological structure represented in the QSP model. Our approach has three main methodological advantages over the traditional approach (see Methods for details on both LIFE and traditional methods for VP generation). LIFE first allows us to identify a subset of parameters which represent variability along the pathways controlling mechanistic phenotypes of disease, rather than leaving the choice of parameters to be more empirically evaluated. Once parameters are identified, LIFE allows us to represent the relationships between parameters so that their interdependence is systematically taken into account, and we do not need to assume that parameters are independent. Finally, when a perturbation is introduced into the system (such as a drug), the LIFE method gives us a mechanistic way of understanding the impact of this perturbation on the network.

To compare the performance of our method to the traditional method used to simulate virtual patient response to therapy, we apply both approaches to a QSP model of cholesterol metabolism, which encompasses the processes of lipoprotein creation and transfer of cholesterol, the role of PCSK9 in regulating LDL-receptor levels, and the action of statin or PCSK9 inhibitor therapy in this system [16]. In the original development of the model, we identified 12 parameters which could be used to define two typical patient profiles (Table 2) – the first represents a patient treated effectively by statin therapy ('Statin Responder'), the second of which represents a patient who is poorly treated by statins ('Statin Non-responder'). While the creation of these two patients allowed us to gain insight into the most important individual mechanisms for statin response, these patient profiles represented extreme examples of response or non-response, and do not give us an understanding of how a representative population of patients with different underlying biological characteristics would fall between these extremes. It further did not allow us to predict which of these patient populations would be best treated by a different therapeutic intervention, such as PCSK9 inhibitors, which could be beneficial for patients on statin with elevated levels of PCSK9 [17]. We demonstrate here how the performance of the traditional vs. the LIFE virtual population methods compare in generating populations of statin responders and non-responders in terms of the three key areas described above: identification of key model parameters, representation of relationships between parameters, and prediction of the systemic effect of a perturbation to the model.

1.3.1 Identification of key mechanistic pathways

We represent the model at steady state as a network of "fluxes", representing the rates at which different biological processes are carried out. This representation allows us to see the interconnections between these processes and to calculate how a change in one rate can be compensated by changes in the kinetics of other processes to maintain the

same steady state biology. The placement of the identified typical patient parameters within the network representation of the cholesterol model is shown in Figure 4a. Within this network, all independent fluxes defining the nullspace are on the left side of the graph; fluxes past this left side of the graph are defined as a linear combination of the independent fluxes. Some subnetworks can be identified in this depiction, such as the cluster of fluxes controlling absorption of dietary cholesterol in the top left corner (f1-f9) and the cluster of fluxes related to the antibody (in green). Overall, one can see that there are many interconnections between different processes in the model and that fluxes chosen to define PD variability in virtual patients are located in different places throughout the network (colored in red, Fig. 4a).

Using LIFE, we found the 10 core independent parameters included in the nullspace of the system, controlling variability in the 12 identified typical patient fluxes (Table 2 and Fig. 4b). Some of the fluxes identified to define virtual patients were confirmed to be core parameters, such as those fluxes controlling reverse cholesterol transport (f34 and f35). In other instances, core parameters differed from the identified parameters, and may appear to be less evenly distributed across different parts of the network. This is because in subsections of the model without any core parameters, variability could be induced through interconnections to other parts of the model (Fig. 4b).

We also represent the pharmacokinetic parameters in the model as fluxes in part of the larger network (In green, Fig. 4a). The drug dose is included as a flux, and is connected only to the rate of drug absorption (Fig. 4c, f72 and f73, respectively). The other parameters defining the PK model for the anti-PCSK9 antibody are dependent on the steady state level of PCSK9, but form an interconnected network that is independent of the fluxes involved in the pharmacodynamics part of the model. We choose to vary the rate of MAb-PCSK9 elimination to simulate variability in patient processing of drug (indicated in the network as f80 in Fig. 4c) which will be propagated through to the two direct connections of this flux, f74 and f76.

1.3.2 LIFE approach results in less variation in metabolic trajectories

We induced variability in the key model parameters using the LIFE method or the traditional VP approach. Lognormal distributions were created for each of the identified parameters through both methods, but the distributions were noticeably different for some parameters (Fig. 5a). The LIFE method generates variability in the values of core parameters that govern the variability in the original 12 VP fluxes rather than the original parameters themselves(Table 2, Fig. 5b), which contributes to the different distributions seen in seven of the identified VP parameters which are not directly varied in LIFE (Fig. 5a, bottom row and last 3 plots of middle row). For the other five identified VP parameters, particularly for the bile cholesterol secretion, hepatic cholesterol production, and PCSK9 plasma clearance rates (Fig. 5b), distributions generated by LIFE are significantly narrower (10-fold lower in range) due to restrictions on what parameter values are physiologically relevant when parameter relationships are considered (distribution parameters given in Table 1 and further constraints imposed to ensure positivity of all fluxes given in Supp. Table 1). This is of key importance in creating virtual populations which distinguish the statin responder and non-responder virtual populations. Because of the very limited range for the bile cholesterol secretion rate and hepatic cholesterol production rate generated by the LIFE method, we can clearly distinguish that the distinct values for these parameters define the main difference between statin responders and non-responders across the identified fluxes (Fig. 5c, top left plots). This distinction is strongly reduced in the traditional VP method because of the uniformly wide distributions generated for all parameters (Fig. 5d). Using the traditional approach, the distribution of bile cholesterol secretion rates is nearly identical between statin responders and non-responders, and the intersection between distributions generated for hepatic cholesterol production, and hepatic and peripheral LDLR production rates

is noticeably larger in the traditional method. Comparing the virtual population parameterizations generated by these two methods (Fig. 5c-d), we can predict that the distinct responses of statin responder and non-responder subpopulations will not be well preserved using the traditional approach.

In order to gain insight from these parameterizations of the model into how each patient will react to a PCSK9 inhibitor therapy, we can look at parameters altered in the production and clearance pathways that are most important in controlling response to the new therapy (Fig. 6). Using the traditional individual fluxes approach, we see that some of the identified parameters lie in the PCSK9 and LDL pathways, including PCSK9 plasma clearance, PCSK9 intracellular production, PCSK9-LDLR dissociation, and HDL to LDL exchange rate, and will likely generate some variability in response to this new therapy. In the LIFE method, many of the core parameters used to control variability in the identified parameters are in the PCSK9/LDL pathways themselves, such as PCSK9-LDLR peripheral uptake (Fig 6a, bottom right plot). In addition to these parameters, we see that variability has been propagated to all other parameters involved in PCSK9 production and clearance in the model using the LIFE method (Fig. 6a), suggesting that PCSK9 biology is highly involved in statin response and that there is likely to be a variability in response to PCSK9 inhibitor therapy as well (Fig 5). Using LIFE, there is variability generated for five out of eight parameters controlling LDL flux, whereas, only HDL to LDL exchange rate is varied in the traditional approach (Fig. 6b).

In the pharmacokinetic portion of the model, the same flux was utilized in both the traditional and LIFE VP approaches, the MAb-PCSK9 elimination rate(Fig 6c). However, in the LIFE method, variability was generated in two additional fluxes of the PK model due to connections across the network, representing relationships between parameters (Fig. 6c). The implementation of the TMDD model in the LIFE framework and propagation of variability through parameters that will affect both the

model PK (through MAb) and PD (through PCSK9) shows that the LIFE method can be used to represent relationships not only within the PD vs. PK compartment, but to represent covariance of parameters across these categories.

1.3.3 LIFE allows us to differentiate variability in patient response from variability in baseline values

While the traditional approach generates patient profiles with different baseline levels of lipoproteins and PCSK9 which can then be weighted or optimized based on the baseline distribution of patients in clinical trials, the LIFE method allows us to create a population of patients with the same baseline levels of lipoproteins and PCSK9 but with fundamentally different biology, based on the structure of the network represented in the model itself. We used both the traditional and LIFE approach, for both statin responder and statin non-responder patients to simulate population response to PCSK9 inhibitor treatment (Fig. 7). Subcutaneous administration of a 150 mg dose of the PCSK9 inhibitor was simulated every 2 weeks from baseline to 90 weeks. At this point, outputs of the LIFE method appeared to be in a steady state (on drug), but there were some continuing downward trends in the VLDL levels of virtual patients created by the traditional approach. Overall trends predicted by both approaches where similar; MAb therapy had an effect on all virtual patients, but of varying degrees. However, there were some key differences in the simulated outputs from virtual populations created under the two approaches.

The traditional approach consistently inflated variability predicted for lipoprotein outputs over the LIFE methodology. This is readily apparent by comparing predictions of HDL using the LIFE approach (Fig. 7d,l) and the traditional approach (Fig. 7h,p). The LIFE approach predicts a very small level of variability in HDL throughout treatment, whereas the traditional approach creates a virtual population where the standard deviation of the distribution is 10% of the mean value. This is

likely due to the variability added to parameters controlling transfer of cholesterol from HDL to LDL and VLDL through the RCT pathway (Fig. 5a-b, Fig. 6b). In the traditional approach, this variability is not compensated for by changes to any other parameters in the model and has a disproportionate effect on HDL at baseline, which contributes to the wide HDL variability throughout the simulation. In contrast, the LIFE approach imposes variability across RCT related parameters to balance out the HDL level (Fig. 6b). Larger variability can also be seen in the populations predicted for LDL and VLDL using the traditional approach (Fig. 7f,g,n,o). The final standard deviation across VP simulations of LDL is 4.6 times larger (58.56 v. 12.62) using the traditional approach compared to the LIFE approach, for the statin-responder population, and the standard deviation of VLDL is 3.9 times larger (3.17 v. 0.82). Again, this is a combination of variability at baseline and in response to drug. More virtual patients created this way develop unrealistic levels of LDL over time compared to the LIFE approach: 0 statin responder and 2 statin non-responder VPs had final LDL levels > 300 mg/dL using the LIFE approach compared to 38 and 13 VPs using the first approach.

Overall, the LIFE approach allows us to better distinguish between the statin responder and non-responder populations once a drug is given. The populations created by the traditional approach have a lot of variability in their response to drug, to the point where these two populations cannot be clearly distinguished. HDL and VLDL response seems to be nearly the same in both virtual patient populations, and the range of LDL levels predicted for the statin non-responder population by the traditional approach seems to correspond to the LDL in a subset of the statin-responder population (Fig. 7n,o,p vs. Fig. 7f,g,h). Using the LIFE methodology, the population responses can be more clearly distinguished for statin responder and non-responder populations. LDL and VLDL mean response is observably higher for statin responders after MAb treatment, and there is little overlap in the ranges of

responses for both virtual patients (Fig. 7b,c vs Fig. 7j,k). As we noted, the overlap between parameter distributions in statin responders and non-responders generated by the traditional approach makes the parameterizations of these virtual populations less distinguishable, and thus generates patient responses that are not distinct (Fig. 5d vs. Fig. 5c).

LDLR expression level is key to the efficacy of PCSK9 inhibitor therapy. By examining LDLR levels at the cell surface and degradation of LDLR in both approaches, we can get insight into the mechanism of action of the therapy. We find that the statin non-responder population gains a higher level of LDLR at the cell surface with treatment than the statin-responder population(Figure 8a,d,g,j), in accordance with the better response of this population to PCSK9 inhibitor therapy. In examining the degradation of LDLR in endosomes when it is alone or in complex with PCSK9 (Fig. 8b-c,e-f,h-i,k-l), we find that there is a similar range of LDLR degraded independently of PCSK9 for both virtual patient populations, but that there is more variability in the amount of LDLR degraded in complex with PCSK9 over time for statin non-responders using both approaches. For both populations again, more variability is present in simulations using the traditional approach than in simulations from LIFE.

Finally, we can see the contribution of PK v. PD variability throughout the system using the LIFE method. We ran simulations of the model for 90 weeks of bi-weekly treatment with a 150 mg dose of the antibody. We compared outputs from simulations which included only PD variability to outputs where both PK and PD variability was used (Fig. 9) and found that the additional influence of PK variability was evident in some but not all outputs. Because the pharmacokinetics of the antibody are represented as a TMDD model and are dependent on the levels of PCSK9 in plasma, there is some variability in the MAb-PCSK9 compound in the simulation where only PD variability is used (Fig. 9a,d). This variability is noticeably enhanced when the PK variability is added (Fig. 9g,j). Other MAb outputs from

the model, such as free MAb in the plasma, did not noticeably change when PK variability was added (data not shown). Similarly for lipoproteins, the effect of the included PK was not always noticeable, but in PCSK9 we noted an increase in the variability of predicted levels when PK was added (Fig. 9h,k,i,l compared to 9b,e,c,f). Moreover, as we have consistently showed, the simulations of the traditional VP approach have much larger variability compared to the LIFE approach in both PK and PD simulations, inflating the contribution of PD variability on PK outputs like MAb-PCSK9 concentration (Fig. 9d), and making it more difficult to distinguish variability in response to treatment from variability in the original population.

1.3.4 Structural investigation of metabolic networks

Lemma 2 *Let $S \in M_{n \times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Denote by s_i the ith row of S. Then we have,*

$$\alpha_1 = \alpha_2 = \cdots = \alpha_n \iff \sum_{i=1}^{n} \alpha_i s_i = \vec{0}.$$

Proof 2 *Because G has no sources and sinks S will have exactly two nonzero elements in each column. This is because each column represents a flow from one node to another.*

$\impliedby$) Fix a column j and let a, b be the rows with nonzero entries. Consider a linear combination of the rows of S such that

$$\sum_{i=1}^{n} \alpha_i s_i = \vec{0}. \tag{1.6}$$

Recall from definition (3) we have $s_{a,j} = -s_{b,j}$. Because a, b are the only nonzero

entries in column j, the jth entry of $\sum_{i=1}^{n} \alpha_i s_i$ satisfies

$$\alpha_a s_{a,j} + \alpha_b s_{b,j} = 0 \implies \alpha_a = \alpha_b.$$

Because G is weakly connected there exists a path between any pair of nodes. Select two arbitrary nodes in the graph G and label them v, v'. Let W be the path between v and v' and label the nodes on the path W as $v = v_1, v_2, \ldots, v_{p-1}, v_p = v'$. Let j_i be the edge connecting v_i and v_{i+1}. Then for any i, the j_ith column satisfies

$$\begin{cases} s_{i,j_i} = -s_{i+1,j_i} \\[2mm] s_{k,j_i} = 0, \qquad for\ k \neq i, i+1. \end{cases}$$

Assume (1.6) then,

$$\alpha_i s_{i,j_i} + \alpha_{i+1} s_{i+1,j_i} = 0 \implies \alpha_i = \alpha_{i+1}.$$

Because j_i can represent any edge on path W, we have $\alpha_1 = \alpha_2, \alpha_2 = \alpha_3, \ldots, \alpha_{p-1} = \alpha_p \implies \alpha_1 = \alpha_2 = \cdots = \alpha_p$. Because v, v' were arbitrary nodes,

$$\sum_{i=1}^{n} \alpha_i s_i = \vec{0} \implies \alpha_1 = \alpha_2 = \cdots = \alpha_n. \tag{1.7}$$

$\implies$) *We assume that $\alpha_1 = \alpha_2 = \cdots = \alpha_n$. As before, fix a column j and let a, b be the rows with nonzero entries. From definition (3) we have $s_{a,j} = -s_{b,j}$. Now consider the jth column of $\sum_{i=1}^{n} \alpha_i s_i$,*

$$\sum_{i=1}^{n} \alpha_i s_{i,j} \tag{1.8}$$

which simplifies to

$$\alpha_a s_{a,j} + \alpha_b s_{b,j} = \alpha_a s_{a,j} - \alpha_b s_{a,j} = (\alpha_a - \alpha_b)s_{a,j} = 0$$

This is true for each column, which gives us $\sum_{i=1}^{n} \alpha_i s_i = \vec{0}$.

Proposition 1 *Let $S \in M_{n \times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Then we have,*

$$Rank(S) = n - 1.$$

Proof 3 *The ($\implies$) of lemma 2 implies $Rank(S) < n$.*

Now we show that $Rank(S) \geq n - 1$. Consider the submatrix S^* constructed by removing the nth row from S. Then for s_i^* the ith row of S^*,

$$\sum_{i=1}^{n-1} \alpha_i s_i^* = \left(\sum_{i=1}^{n} \alpha_i s_i \right)_{\Big|_{\alpha_n = 0}}. \tag{1.9}$$

by (1.9) and lemma 2 It follows that

$$\sum_{i=1}^{n} \alpha_i s_i^* = \vec{0} \implies \alpha_i = 0 \text{ for all } i \in \{1, \ldots, n-1\}.$$

Therefore, $Rank(S^*) = n - 1 \implies n - 1 \leq Rank(S) < n$ and so $Rank(S) = n - 1$.

Proposition 2 *Let $S \in M_{n \times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with at least one source and no sinks. Then we have,*

$$Rank(S) = n.$$

Proof 4 *First we show that for a graph G with a single source and no sinks, that for S, the stoichiometric matrix for G, $Rank(S) = n$. Let the source be called f_{m+1}*

and the terminal vertex of f_{m+1} be called x_1. Let G^* be the subgraph of G without the source, and S^* be the matrix for G^*. S^* is a submatrix of S excluding the column containing the source. We have $rank(S^*) = n - 1$. We can use elementary row operations to change a row in S^* without changing the rank of S^*. We replace the first row of S^* with $\sum_{i=1}^{n} s_i^* = \vec{0}$ and call this new matrix S^1, $rank(S^1) = rank(S^*) = n-1$. Similarly, if we append a column of zeros to the right side of S_1^* the rank will not change. We call the matrix with the added column S^2, $rank(S^2) = rank(S^1) = rank(S^*) = n - 1$. S^2 is S with the first row of S set to $\vec{0}$. Now we replace the first row of S^2 with $(s_{1,1}, s_{1,2}, \ldots, s_{1,m-1}, s_{1,m} = 1)$ which gives us S. Because the first row is independent to all others:

$$rank(S) = rank(S^2) + 1 = n.$$

For ease of proof the graph contained no sinks. However, adding sinks to the graph will not change the rank of the S. This is because S is already full rank and adding a sink will append a new column to S. A graph with sources and no sinks is not realistic as it will have continuous accumulation of metabolites.

Proposition 3 *Let S be the stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with no sources or sinks. Consider the kernel of S, $K(x)$, and assume that $\tilde{x} = cx$ for some $c \in \mathbb{R}$. Then we have, $K(x) = K(\tilde{x})$.*

Proof 5 *Let $\hat{f} \in K(x)$, and $\check{f} \in K(\tilde{x})$ then*

$$\begin{aligned}
S(x)\hat{f} = 0 &\implies cS(x)\hat{f} = S(\tilde{x})\hat{f} = 0 \\
S(\tilde{x})\check{f} = 0 &\implies cS(x)\check{f} = 0.
\end{aligned} \tag{1.10}$$

Proposition 4 *Let S be a stoichiometric matrix for a graph containing a directed path along three nodes, and the middle node has only one incoming and one outgoing*

edge. For a state x and different state $\tilde{x}$, if $\tilde{x} \neq cx$ for $c \in \mathbb{R}$ then

$$K(x) \cap K(\tilde{x}) = \{\vec{0}\}. \tag{1.11}$$

Proof 6 G has directed path along three nodes, initial node x_1, middle node x_2, terminal node x_3; call edge connecting x_1 to x_2 as f_1 the other edge is f_2. Then the second row of S is $s_2 = (-x_1, x_2, 0, \ldots, 0)$ and

$$S(x)f = 0 \implies f_1 x_1 = f_2 x_2 \implies f_1 = f_2 \frac{x_2}{x_1}$$

$$S(\tilde{x})f = 0 \implies f_1 \tilde{x}_1 = f_2 \tilde{x}_2 \implies f_1 = f_2 \frac{\tilde{x}_2}{\tilde{x}_1}$$

$$f \in K(x) \cap K(\tilde{x}), f \neq \{\vec{0}\} \implies f_2 \frac{x_2}{x_1} = f_2 \frac{\tilde{x}_2}{\tilde{x}_1} \implies \tilde{x} = cx.$$

Proposition 5 Let $S \in M_{n \times m}$, $n < m$, be a stoichiometric matrix and G the associated directed graph. Assume G to be weakly connected with one source and no sinks. Let $S^* \in M_{n \times m-1}$, be a submatrix of S where the source is removed. (WLOG let the source in G be represented by the last column of S). Consider the kernels of S and S^*, $K(x)$ and $K^*(x)$ respectively and let B^* be a basis of $K^*(x)$. Let B be the collection of vectors such that each $b \in B$ is equal to a $b^* \in B^*$ with a 0 appended as the last entry for each vector. Then B is a basis for $K(x)$.

Proof 7 We prove:

1. for $b \in B, Sb = \vec{0}$ and so $b \in K(x)$.

2. B is an independent set with number of elements equal to dimension of $K(x)$.

Let e be an $n \times 1$ column vector containing a single 1 and the other entries 0.

$$S_{n \times (m+1)}b = \left(S^* | e \right) \begin{pmatrix} b^* \\ \text{-----} \\ 0 \end{pmatrix}.$$

Let b_j be the jth entry of vector b. For A_i, the ith entry of the solution to Sb.

$$A_i = \sum_{j=1}^{m} S_{ij}b_j = \sum_{j=1}^{m-1} S_{ij}^* b_j^* + S_{i\,m} \cdot 0 = 0.$$

Appending a 0 to each vector of a linearly independent set gives an linearly independent set. From propositions 1 and 2 we know that

$$Rank(S^*(x)) = n - 1, Rank(S(x)) = n.$$

The dimension of each kernel is the same $(dim(K^*(x)) = (m-1) - (n-1)$ and $dim(K(x)) = m - n$. The cardinality of $B^* =$ cardinality of $B = dim(K(x)) = dim(K^*(x))$ because B is a basis and both kernels have the same dimension (though the dimension of their ambient space differs), we conclude that B is a basis for $K(x)$.

Here we show a complete solution to problem 1. In this section we explore the kernel of an example network. The initial state of the kernel will be characterized, and the intersection of this kernel with the kernels of perturbed metabolic states will be analyzed.
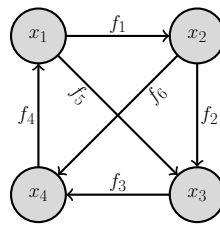

Figure 1.1: A directed graph representing a biochemical system.

$S(x)$ is the stoichiometric matrix associated to the graph in Fig. 1.1

$$S(x) = \begin{pmatrix} -x_1 & 0 & 0 & x_4 & -x_1 & 0 \\ x_1 & -x_2 & 0 & 0 & 0 & -x_2 \\ 0 & x_2 & -x_3 & 0 & x_1 & 0 \\ 0 & 0 & x_3 & -x_4 & 0 & x_2 \end{pmatrix}.$$

From proposition 1 we have $rank(S) = 3$, which implies the dimension of the kernel is 3. The basis for the kernel is

$$(v_1|v_2|v_3)^T = \begin{pmatrix} 0 & -1 & -\frac{x_2}{x_3} & 0 & 0 & 1 \\ -1 & -\frac{x_1}{x_2} & 0 & 0 & 1 & 0 \\ \frac{x_4}{x_1} & \frac{x_4}{x_2} & \frac{x_4}{x_2} & 1 & 0 & 0 \end{pmatrix}.$$

A perturbation of $x \to \tilde{x}$ will alter the basis vectors and thus change the kernel $K(x) \to K(\tilde{x})$. $K(\tilde{x})$ may have some non trivial intersection with $K(x)$. Any flux in the perturbed kernel can be represented by the perturbed basis vectors. For all $f \in K(\tilde{x}) : f = \tilde{\lambda}_1 \tilde{v}_1 + \tilde{\lambda}_2 \tilde{v}_2 + \tilde{\lambda}_3 \tilde{v}_3$ where each $\tilde{v}_i$ represents a perturbed basis vector and each $\tilde{\lambda}_i \in \mathbb{R}$.

A flux $f \in K(x) \cap K(\tilde{x})$ can be found as a solution to the following equation:

$$\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = \tilde{\lambda}_1 \tilde{v}_1 + \tilde{\lambda}_2 \tilde{v}_2 + \tilde{\lambda}_3 \tilde{v}_3.$$

Comparing the equation by components, we have the conditions that must be satisfied for any flux in the intersection.

$$\lambda_1 = \tilde{\lambda}_1, \quad \lambda_2 = \tilde{\lambda}_2, \quad \lambda_3 = \tilde{\lambda}_3 \tag{1.12}$$

$$\begin{cases} \lambda_1 \frac{x_4}{x_1} - \lambda_2 &= \tilde{\lambda}_1 \frac{\tilde{x}_4}{\tilde{x}_1} - \tilde{\lambda}_2 \\[2mm] \lambda_1 \frac{x_4}{x_2} - \lambda_2 \frac{x_1}{x_2} - \lambda_3 &= \tilde{\lambda}_1 \frac{\tilde{x}_4}{\tilde{x}_2} - \tilde{\lambda}_2 \frac{\tilde{x}_1}{\tilde{x}_2} - \tilde{\lambda}_3 \\[2mm] \lambda_1 \frac{x_4}{x_3} - \lambda_3 \frac{x_2}{x_3} &= \tilde{\lambda}_1 \frac{\tilde{x}_4}{\tilde{x}_3} - \tilde{\lambda}_3 \frac{\tilde{x}_2}{\tilde{x}_3}. \end{cases} \tag{1.13}$$

With (1.12), we simplify system (1.13) to

$$\frac{x_4}{x_1} = \frac{\tilde{x}_4}{\tilde{x}_1} \tag{1.14}$$

$$\lambda_1 \left(\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2} \right) = \lambda_2 \left(\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2} \right) \tag{1.15}$$

$$\lambda_1 \left(\frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} \right) = \lambda_3 \left(\frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} \right). \tag{1.16}$$

Depending on which of the conditions are met the dimension of the intersection $(dim(K(x) \cap K(\tilde{x})))$ can be determined. Different perturbations of x will be considered that satisfy only some of these conditions. The following cases ((I) through (V)) show results specific to which condition are satisfied.

(I) Let $\tilde{x}$ be a perturbation such that (1.14) is not satisfied, the intersection will be trivial and $dim(K(x) \cap K(\tilde{x})) = 0$.

(II) Let $\tilde{x}$ be a perturbation which satisfies (1.14) and

$$\left(\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2} \right) \left(\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2} \right) \neq 0, \left(\frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} \right) \left(\frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} \right) \neq 0.$$

This allows equations (1.15) and (1.16) to be arranged in the following manner.

$$\lambda_2 = \lambda_1 \frac{\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2}}{\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2}}, \qquad \lambda_3 = \lambda_1 \frac{\frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3}}{\frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3}}.$$

This shows a relationship where both λ_2 and λ_3 depend on λ_1 and the metabolites x_i. λ_1 is the only free variable and so $dim(K(x) \cap K(\tilde{x})) = 1$.

(III) Let $\tilde{x}$ satisfy (1.14). And also let $\left(\frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} \right) \left(\frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} \right) \neq 0$, $\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2} = 0$, $\frac{x_1}{x_2} -$

$\frac{\tilde{x}_1}{\tilde{x}_2} = 0$. Then (1.15) is satisfied regardless of the value of λ_2. λ_2 is a free variable in addition to λ_1 while λ_3 is still dependent on the state $\tilde{x}$ and λ_1. With two free variables $dim(K(x) \cap K(\tilde{x})) = 2$.

(IV) Let $\tilde{x}$ satisfy (1.14) and also let

$$\frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} = 0, \frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} = 0 \tag{1.17}$$

$$\left(\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2}\right)\left(\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2}\right) \neq 0. \tag{1.18}$$

Upon further inspection, however, we find that (1.14) and (1.17) implies $\left(\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2}\right) = \left(\frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2}\right) = 0$ which contradicts (1.18). Thus the perturbation given by case (IV) doesn't exist.

(V) Let $\tilde{x}$ satisfy (1.14). And let

$$\frac{x_4}{x_2} - \frac{\tilde{x}_4}{\tilde{x}_2} = \frac{x_1}{x_2} - \frac{\tilde{x}_1}{\tilde{x}_2} = \frac{x_4}{x_3} - \frac{\tilde{x}_4}{\tilde{x}_3} = \frac{x_2}{x_3} - \frac{\tilde{x}_2}{\tilde{x}_3} = 0.$$

Equations (1.14), (1.15) and (1.16) are satisfied for any value of λ_1, λ_2 and λ_3. With three free variables $dim(K(x) \cap K(\tilde{x})) = 3$. Fig. 1.2 shows states $\tilde{x}$ for which $K(x) \cap K(\tilde{x})$ is non trivial. The reference state x is shown, and the entire space represents other states $\tilde{x}$ such that equation (1.14) is satisfied.

This work exploits the linearity of some metabolic systems with respect to fluxes. Namely, we are able to see systematic relationships among fluxes at equilibrium as opposed to treating all fluxes as independent. Propositions concerning the rank of our stoichiometric matrix are presented, from which the dimension of the kernel may be easily deducted.

1.4 Discussion

We present a virtual patient methodology which leverages the QSP model structure to derive relationships among parameters and uses these relationships to create virtual patient populations, thus systematically taking into account the interdependence of parameters in the model. We show a method for sampling parameter values, which creates distributions for each parameter of interest that maximizes variability in the parameter space while constraining fluxes in the nullspace to produce biologically reasonable values for all fluxes in the model. This will ensure that the majority of simulations will produce results within the physiological range. The LIFE method propagates variability in key model parameters to other fluxes in the model, generating more variability in pathways of interest. These pathways can be analyzed with a view to a new therapeutic intervention, to predict the range of responses expected to a different type of treatment. Our methodology generates virtual patients with tighter ranges of variability around model outputs than the traditional method of virtual patient creation, which uniformly predicts large variance in all model outputs. In this way, our method better maintains separation between distinct patient populations and allows us to see clearly what outputs will differ across virtual patients in response to treatment and which should not. We show that variability in both PK and PD parameters can be incorporated using this methodology, so that interdependence of PK and PD parameters can be taken into account as well. This representation will allow us to systematically analyze whether patient response is due to properties of the drug that can be altered to make treatment more effective, or whether there is an underlying biological cause of non-responsiveness. Finally, it is important to note that virtual patients generated by the LIFE methodology can have the same baseline values, but different responses to drug, because drug response is dictated by the underlying processes which give rise to this state. By enabling us to systematically identify and analyze the factors controlling the different responses of

patients with similar baseline clinical measurements to treatment, the LIFE method enables us to better understand what clinical data is needed to accurately identify drug responders.

A virtual patient consists of a set of specific parameter values which represent different kinetic rates in a patient's biology. To determine how the parameter values should vary across virtual patients, different approaches are used. Generally, a typical value or an upper and lower bound are set for each parameter based on what is known in the literature, and each parameter is sampled from a normal or uniform distribution within these bounds [13, 14, 12]. From this initial set of virtual patients, a subset is selected by comparing model predictions for these virtual patients to clinical baseline or time course data of key outputs, and using a weighting scheme such as prevalence weighting [18] to select patients which replicate the distribution of this data [13, 14, 12, 19]. Thus, the distributions of parameters can be examined after the virtual population is defined, but there is no fundamental understanding of what the distribution and covariance of parameters should be. Using the LIFE methodology, we sample parameters in a directed way from a lognormal distribution, and calculate the feasibility of the resulting parameter set by considering the values of other parameters generated from the core fluxes. We use a cost function which considers feasibility of the parameter set generated rather than the feasibility of the baseline patient values, because we know that the baseline values will be equal to the steady state values that we put into the LIFE method. Using this procedure, we demonstrated how the LIFE method can be useful for discerning the differential impact of key parameter values and treatment responses in distinct virtual patient populations.

The LIFE method allows us to systematically include the structural relationship of parameters in the model by representing the relationships between the parameters at steady state. This representation tells us which parameters can be independently

varied, and how these independent fluxes will propagate variability to other parts of the model through these relationships. This way of representing interdependence in the model is an advantage over current VP methodologies where this interdependence must be empirically explored for sets of parameters [14, 13], or potentially estimated from the dataset [19]. The LIFE method allows the interdependence to be an intrinsic part of the parameter calculation. As shown in Figures 7,8, and 9, this results in outputs which have more realistic variability. This also helps to reduce the number of parameters in the model which can be used to generate variability in a VP response, better informing the initial choice of parameters to define a VP. Finally, it gives us mechanistic insight into how clinical steady state levels will change based on different parameter values or on treatment with a new drug before we run the model.

Extensive work has been done in the field of population pharmacokinetics to model inter-individual variability of parameters in the model by fitting the variance and covariance of pharmacokinetic and simple pharmacodynamic parameters to clinical data [20]. Many virtual patient cohorts developed from QSP models are created by varying pharmacodynamics parameters affecting intrinsic patient biology, without considering pharmacokinetic variability [19, 13]. Work by Gadkar, et al. [12] simulates variability in both pharmacodynamic and pharmacokinetic parameters by using the prevalence weighting method, but covariance between these parameters is not mentioned. The LIFE methodology is a novel approach to enable modeling of variability in pharmacokinetic and mechanistically-specific pharmacodynamic parameters, where interdependence among these parameters is handled in the same coherent way as interdependence among pharmacodynamics parameters. This approach more accurately represents the different factors that can alter patient response. It is especially important for Target-Mediated Drug Distribution models, such as the one used for the anti-PCSK9 antibody in this work, because the clearance and transport of the drug is inextricably linked to the target.

Virtual patient methods in the literature focus on generating patients with a range of initial output values that are similar to the clinical range and distribution for each output [14, 12]. This approach seeks to replicate the range of patients who enter into a trial, with different baseline levels across patients for key parameters. The LIFE method can be used to create population of subjects with a distribution of baseline measurements. However, the LIFE method was developed with the knowledge that patients with the *same* baseline levels may have vastly *different* responses to drug, meaning that the LDL and HDL of a patient at baseline is not enough to predict the efficacy of treatment for this patient. The LIFE method enables us to advance mechanistic hypotheses on the factors driving this variable drug response. We can create populations with the same baseline values but different biology to analyze the effect of biological and pharmacokinetic variability on treatment outcome. Previous work by Hosseini et al. [13] shows some analysis of virtual patients with different parameterizations but a similar time course of disease progression, which speaks to this idea. Our methodology gives the user an analytical representation of this set of patients and an understanding of how they relate, so that we can efficiently generate a population of these virtual patients by utilizing the flux relationship structure rather than numerically determining individual patient profiles empirically. This is of key importance in identifying responder sub-populations so that we can better target trial populations to include patients who have a higher probability of being treated by drug because of their underlying biology. This approach can also be connected to a disease progression model, to help us predict how patients with the same clinical endpoint levels in the short term will have different long term outcomes.

We have demonstrated here the ability of the LIFE method to create virtual patient populations and the advantages that it has over existing virtual population development methods, including more restricted variability in key identified parameters, enhanced variability in other parameters in relevant pathways, and more applicable

range of therapeutic outcomes when both PK and PD parameters are varied. Our ultimate goal is to use this method to match clinical data for different patient subgroups, and to hypothesize mechanistic characteristics likely to lead a patient to best respond to therapy. Existing methods for virtual patient generation have been used to this effect [13, 12, 19], and we would like to see what additional information can come from our understanding of the model as a system of connected fluxes. Our method can give insight into key parameters to use in optimizing parameter values and in generating virtual patients. In the QSP model presented here, from an original set of 101 parameters, our method generates 49 core fluxes which are independently responsible for variability in the model and can be used for virtual patient creation. We would like to further develop our method to determine from these 49 parameters, which are most important to vary, and how many parameters are necessary to encompass all variability seen in the clinical data. We believe that the LIFE method can make a significant contribution towards improving the efficiency and robustness of existing virtual patient methods and ultimately allowing for more comprehensive QSP simulators that advance mechanistic hypotheses ascertaining a drug's mechanism of action, and disease mechanisms commensurate with drug mechanism of action, linking the right drug to the right patient.

Figure Legends

Figure 1.3 Example metabolic network. Each node represents a biomolecule, or metabolite and each edge represents a rate constant, or flux. From this graph we obtain a system of Ordinary Differential Equations that govern metabolite levels over time.

Figure 1.4 An example of "importance sampling." The blue curve is a sampling distribution of a flux f. To estimate the range of fluxes resulting from a sample, we calculate fluxes dependent on f using the red tick marks, which are equally spaced from 2.5 percentile to the 97.5 percentile.

Figure 4 From the original set of identified virtual patient-defining parameters, LIFE method determines a core set of parameters which control this variability in the model. The QSP model of cholesterol synthesis can be represented as a network of fluxes (f 's). The nullspace defines a connectivity matrix between all reaction rates (f 's) of the model and the core f's controlling the variability in the model. Fig 4(a) highlights (in red) the 12 reaction rates originally identified to define virtual patients within this network context. Fig 4(b) shows core fluxes (in red) used to control variability in network approach. The TMDD model of anti-PCSK9 inhibitor is incorporated into the model (green nodes)and is shown in isolation in Fig 4(c). One flux was used to create variability in the PK portion of the model (shown in red).

Figure 5 Both LIFE and traditional VP methods generate variability around parameters originally identified to define VPs. However, the LIFE method also creates variability in other key model parameters, and it generates different parameter distributions for identified parameters. Parameter distributions were generated either by individually varying identified VP-defining fluxes (blue,cyan) or by using LIFE

method (red,yellow) to vary core fluxes controlling variability in the network and to calculate values for parameters dependent on these fluxes. Plots show overlaid histograms of the parameter value distributions generated with both approaches for (a) identified VP-defining parameters, defining the statin responder population, and (b) core parameters controlling variability in the model for the statin responder population. In (c-d), identified VP-defining parameter distributions are shown for (c) the LIFE method, using parameter values for a typical statin responder (red) vs. non-responder (yellow) patient and for (d) the traditional method, using parameter values for a typical statin responder (cyan) vs. non-responder (blue) patient. Unstable or unphysiological simulations (with final LDL > 300 mg/dL) were excluded from plots.

Figure 6 The different VP approaches generate different levels of variability in key pathways of interest. Virtual populations based on the statin responder profile were created using the LIFE (red) and traditional (cyan) VP methods. Variability generated in (a). parameters controlling synthesis and clearance of PCSK9, (b) parameters controlling synthesis and clearance of LDL, and (c) PK-related parameters is shown. Unstable or unphysiological simulations (w final LDL > 300 mg/dL) were excluded from plots.

Figure 7 LIFE Methodology minimizes variability by taking covariance between parameters into account and creates populations with the same baseline values but different drug responses. Simulations were run to simulate 90 weeks of Q2W dosing with 150 mg anti-PCSK9 antibody. Predicted outputs are shown for Statin responder (a-h) and non-responder (i-p) populations, created by fluxes using the traditional or by using the LIFE network approach. Plots show the mean (in blue) +/- standard deviation(dotted line) of the virtual patient responses generated from each simulation of PCSK9 (a,e,i,m), LDL(b,f,j,n), VLDL(c,g,k,o), and HDL(d,h,l,p). Unstable or unphysiological simulations (with final LDL > 300 mg/dL) were excluded from plots.

Figure 8 Model predictions about LDLR again show increased variability predicted from simulations of a virtual population created using the traditional approach compared to the LIFE method. Simulations were run to simulate 90 weeks of Q2W dosing with 150 mg anti-PCSK9 antibody. Predicted outputs are shown for Statin responder (a-f) and non-responder (g-l) populations, created with the LIFE approach (a-c, g-i) or the traditional approach (d-f, j-l). Plots of hepatic LDLR at the cell surface (a,d,g,j) show the mean (in blue) +/- standard deviation of the virtual patient responses generated from each simulation. Plots of the amount of LDLR degraded in the endosome per hour when isolated (b,e,h,k), or in complex with PCSK9 (c,f,i,l) show time courses for each VP. Unstable or unphysiological simulations (with final LDL > 300 mg/dL) were excluded from plots.

Figure 9 Key model outputs are influenced by both PK and PD variability. Simulations were run after using either the traditional or the LIFE approach to generate parameter distributions for use in the model. Simulations were based on the typical statin responder profile. Parameters varied included PD parameters only (a-f) or PD and PK parameters (g-l). The simulation was run over a period of 90 weeks. 2.5 stable cycles Q2W Mab dosing are shown, for simulated levels of Mab-PCSK9in plasma (a,d,g,j) and PCSK9 in plasma (b-c,e-f,h-i,k-l) . The mean (blue) +/- standard deviation (dotted black lines) are shown in plots (a-b,d-e,g-h, j-k), while individual virtual patient predictions are overlaid on the same plot in (c,f,i,l). Unstable or unphysiological simulations (with final LDL > 300 mg/dL) were excluded from plots.

1.5 Tables

Standard error ($\frac{\sigma}{\mu}$, $\quad \sigma$ = standard deviation, μ = mean) of flux sampling distributions for each VP class.

- f_1 = rate constant for the trafficking of newly synthesized LDLR to the surface of peripheral cells.

- f_2 = rate constant for the trafficking of newly synthesized LDLR to the surface of hepatocytes.

- f_3 = rate constant for the clearance of PCSK9 from the plasma.

- f_4 = PC9 LDLR internalization rate in peripheral cells.

- f_5 = rate constant for the exchange of cholesterol from HDL to LDL.

- f_6 = rate constant for the exchange of cholesterol from HDL to VLDL.

- f_7 = PCSK9 LDLR association rate in plasma.

- f_8 = rate of cholesterol production by hepatocytes.

- f_9 = rate constant for secretion of biliary cholesterol into the GI tract plus bileacid cholesterol secretion rate k.

- f_{10} = PCSK9 intra cellular to plasma release rate in hepatocytes. f_{10} is dependent on the sample flux f_3.

Table 1	f_1	f_2	f_3	f_4	f_5	f_6
VP0	1.2545	0.7855	0.2020	0.0466	152.1765	17.5696
VP1	1.3230	0.8745	0.2910	0.0365	97.6538	12.2185
VP2	1.2310	0.6585	0.2900	0.0365	97.6538	12.2185
VP3	1.3170	0.5575	1.0420	0.0361	97.6538	12.2185
VP4	1.4870	1.1310	0.0117	0.0370	97.6538	12.2185

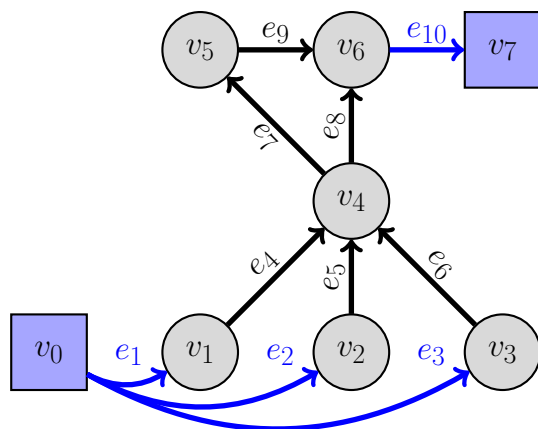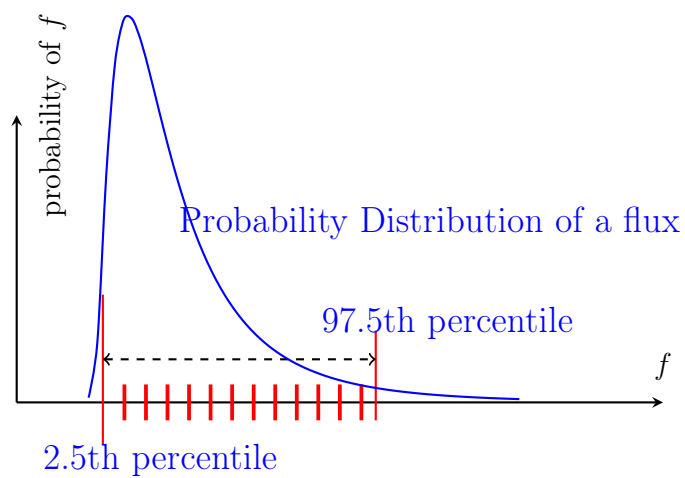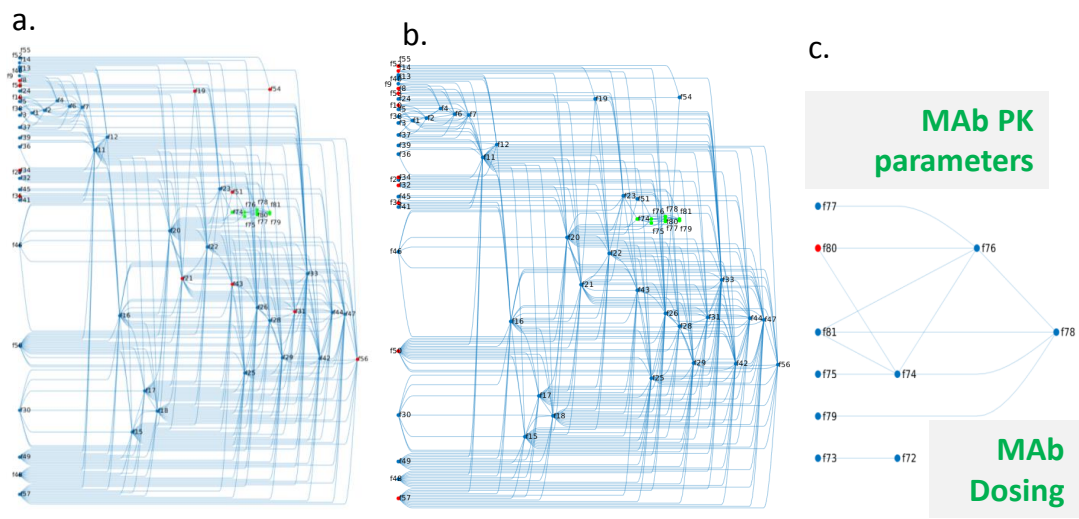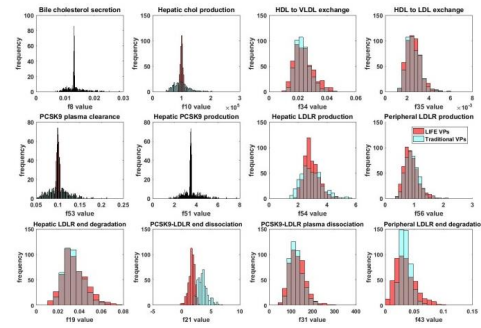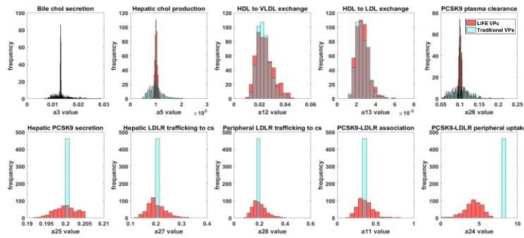| Table 1 cont'd	f_7	f_8	f_9
VP0	0.7536	3.4829×10^{-6}	2.9733
VP1	0.7536	4.4700×10^{-7}	0.6412
VP2	0.7536	7.4850×10^{-6}	0.5211
VP3	0.7389	4.3671×10^{-6}	0.6154
VP4	0.7608	6.4475×10^{-6}	3.7740

Figure 1.3

Figure 1.4

Figure 4.

Figure 5.

a. **Identified VP parameters: LIFE vs. traditional**
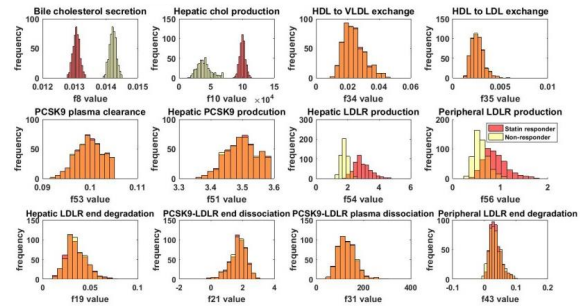

b. **Core VP parameters: LIFE vs. traditional**

c. **Statin responder vs. non-responder parameters: LIFE approach**

d. **Statin responder vs. non-responder parameters: traditional approach**
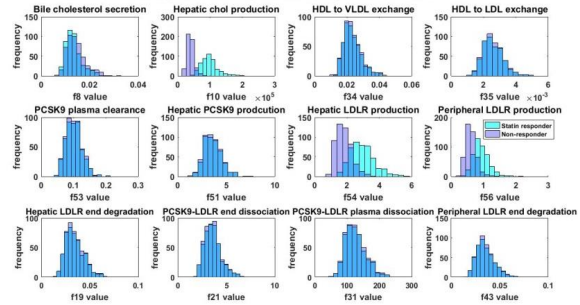
Figure 6.

a.

PCSK9 related parameters:
LIFE vs. traditional
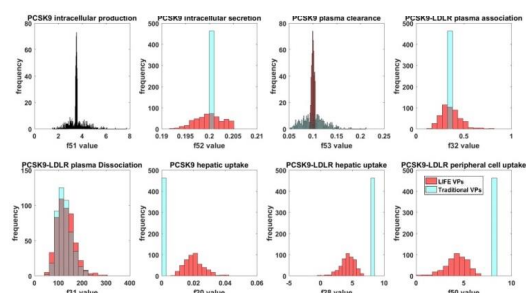
b.

LDL related parameters: :
LIFE vs. traditional
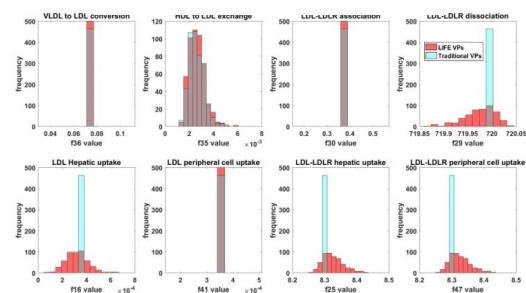
c.

PK parameters :
LIFE vs. traditional

Figure 7.

Statin Responder | **Statin Non-Responder**

LIFE method | Traditional method | LIFE method | Traditional method

Figure 8.

| Statin Responder | | Statin Non-Responder | |
| LIFE method | Traditional method | LIFE method | Traditional method |

Figure 9.

PD parameters varied

LIFE method   ### Traditional method

PK & PD parameters varied

LIFE method   ### Traditional method
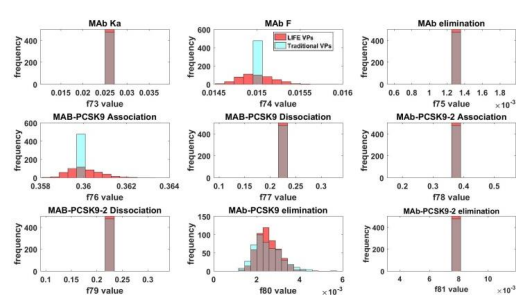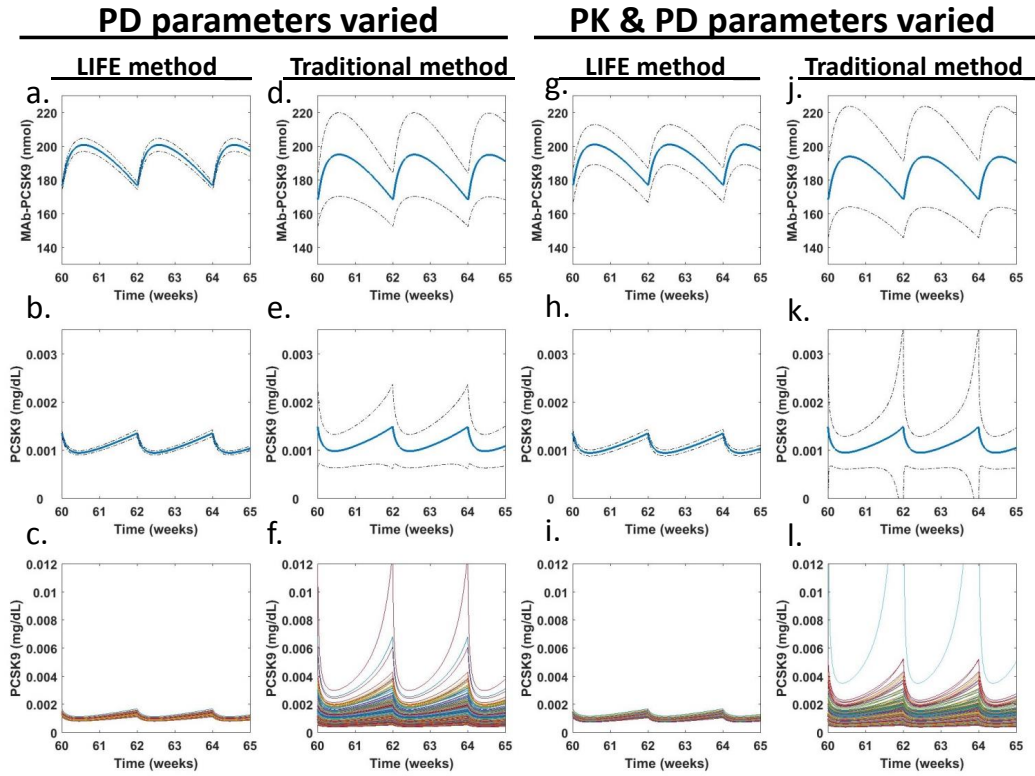
Supplementary Table 1

We have two general constraints on sampled parameters. Each VP will have additional constraints. The description of fluxes $f_1, \ldots, f_9$ are found in Table 1; $\bar{f}_i$ is the average value of flux f_i.

1.) We will allow f_4 only to decrease with the sampling algorithm:

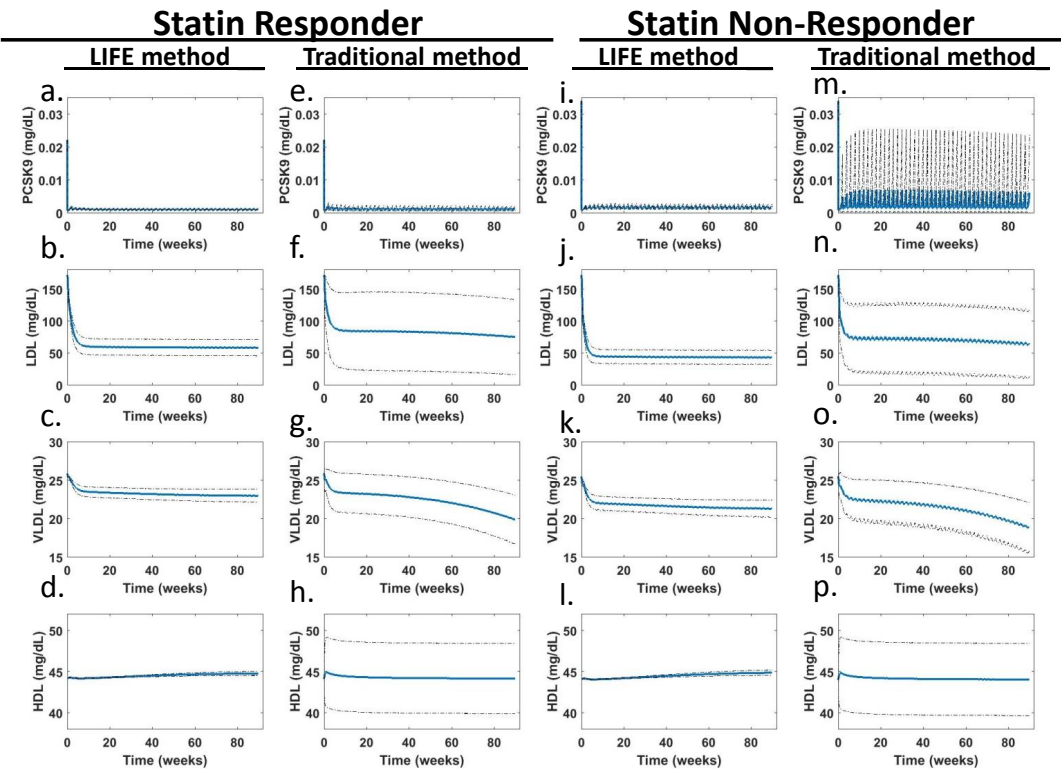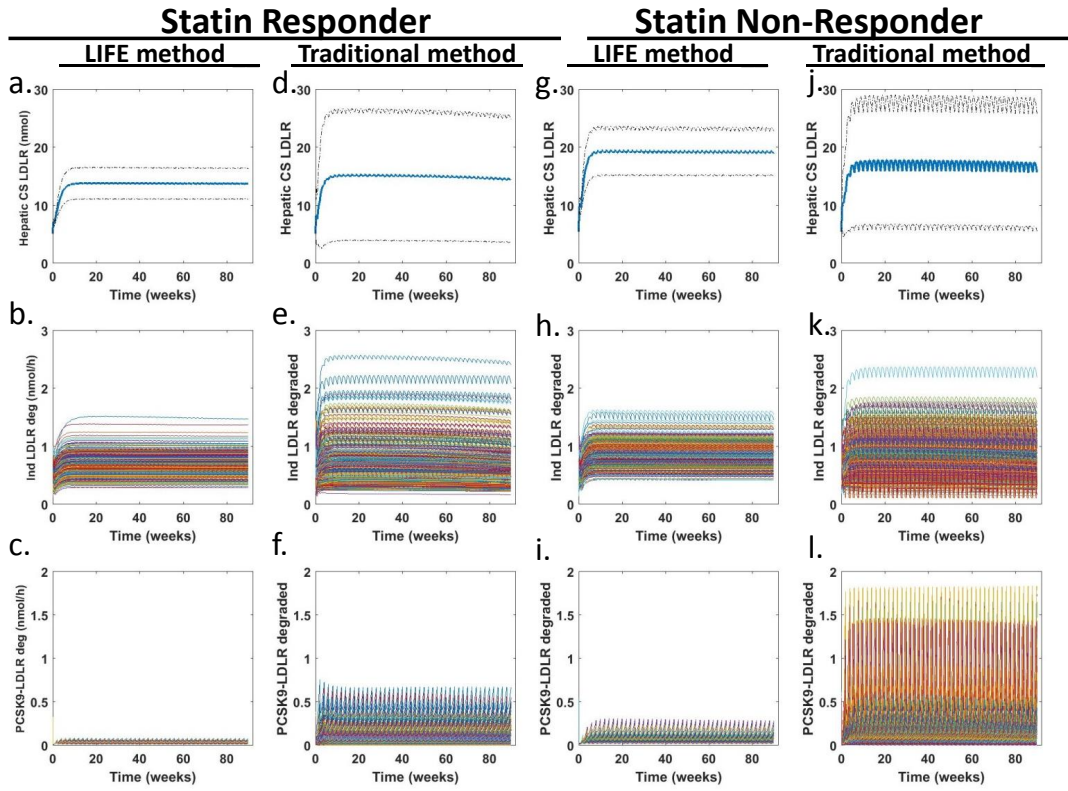$$\begin{cases} f_4 & \text{sampled from lognormal} \\ f_4 = 8.3, & \text{if sampled value is } > 8.3 \end{cases}$$

2.) For all virtual patients, for a given sample, f_3 and f_{10} will be coupled in the following way:

$$f_3 - \bar{f}_3 = \varepsilon_3, \text{ then let, } f_{10} = \bar{f}_{10} + \varepsilon_3.$$

This is to say that the amount that $a9$ deviates from $\bar{f}_3$ is precisely the amount that f_{10} will deviate from $\bar{f}_{10}$.

For VP0

3.) We apply a strict bound to f_1. $\bar{f}_1 = 0.2$

$$\begin{cases} f_1 & \text{sampled from lognormal} \\ f_1 = 0.0985, & \text{if sampled value is } < 0.0985 \end{cases}$$

4.) We apply a strict bound to f_2. $\bar{f}_2 = 0.2$

$$\begin{cases} f_2 & \text{sampled from lognormal} \\ f_2 = 0.1000, & \text{if sampled value is } < 0.1000 \end{cases}$$

5.) We apply a strict bound to f_2. $\bar{f}_2 = 0.1$

$$
\begin{cases}
f_3 & \text{sampled from lognormal} \\
f_3 = 0.1060, & \text{if sampled value is } > 0.1060
\end{cases}
$$

6.) We apply a strict bound to f_6. $\bar{f}_6 = 0.0158$

$$
\begin{cases}
f_6 & \text{sampled from lognormal} \\
f_6 = 0.0784, & \text{if sampled value is } > 0.0784
\end{cases}
$$

7.) We apply a strict bound to f_8. $\bar{f}_8 = 70,000$

$$
\begin{cases}
f_8 & \text{sampled from lognormal} \\
f_8 = 1.0870e + 05, & \text{if sampled value is } > 1.0870e + 05
\end{cases}
$$

8.) We apply a strict bound to f_9. $\bar{f}_9 = 0.0150$

$$
\begin{cases}
f_9 & \text{sampled from lognormal} \\
f_9 = 0.0147, & \text{if sampled value is } < 0.0147
\end{cases}
$$

For VP1

3.) We apply a strict bound to f_1. $\bar{f}_1 = 0.2$

$$
\begin{cases}
f_1 & \text{sampled from lognormal} \\
f_1 = 0.0896, & \text{if sampled value is } < 0.0896
\end{cases}
$$

4.) We apply a strict bound to $a8$. $\bar{f}_2 = 0.2$

$$\begin{cases} f_2 & \text{sampled from lognormal} \\ f_2 = 0.0906, & \text{if sampled value is } < 0.0906 \end{cases}$$

5.) We apply a strict bound to f_3. $\bar{f}_3 = 0.1$

$$\begin{cases} f_3 & \text{sampled from lognormal} \\ f_3 = 0.1048, & \text{if sampled value is } > 0.1048 \end{cases}$$

6.) We apply a strict bound to f_6. $\bar{f}_6 = 0.0238$

$$\begin{cases} f_6 & \text{sampled from lognormal} \\ f_6 = 0.0985, & \text{if sampled value is } > 0.0985 \end{cases}$$

7) We apply a strict bound to f_8. $\bar{f}_8 = 100,000$

$$\begin{cases} f_8 & \text{sampled from lognormal} \\ f_8 = 1.6483e + 05, & \text{if sampled value is } > 1.6483e + 05 \end{cases}$$

8.) We apply a strict bound to f_9. $\bar{f}_9 = 0.0131$

$$\begin{cases} f_9 & \text{sampled from lognormal} \\ f_9 = 0.0128, & \text{if sampled value is } < 0.0128 \end{cases}$$

For VP2

3.) We apply a strict bound to f_1. $\bar{f}_1 = 0.2$

$$
\begin{cases}
f_1 & \text{sampled from lognormal} \\
f_1 = 0.1097, & \text{if sampled value is } < 0.1097
\end{cases}
$$

4.) We apply a strict bound to f_2. $\bar{f}_2 = 0.2$

$$
\begin{cases}
f_2 & \text{sampled from lognormal} \\
f_2 = 0.1117, & \text{if sampled value is } < 0.1117
\end{cases}
$$

5.) We apply a strict bound to f_3. $\bar{f}_3 = 0.1$

$$
\begin{cases}
f_3 & \text{sampled from lognormal} \\
f_3 = 0.1048, & \text{if sampled value is } > 0.1048
\end{cases}
$$

6.) We apply a strict bound to f_6. $\bar{f}_6 = 0.0238$

$$
\begin{cases}
f_6 & \text{sampled from lognormal} \\
f_6 = 0.0977, & \text{if sampled value is } > 0.0977
\end{cases}
$$

7) We apply a strict bound to f_8. $\bar{f}_8 = 40,000$

$$
\begin{cases}
f_8 & \text{sampled from lognormal} \\
f_8 = 6.9189e + 04, & \text{if sampled value is } > 6.9189e + 04
\end{cases}
$$

8.) We apply a strict bound to f_9. $\bar{f}_9 = 0.0142$

$$\begin{cases} f_9 & \text{sampled from lognormal} \\ f_9 = 0.0139, & \text{if sampled value is} < 0.0139 \end{cases}$$

For VP3

3.) We apply a strict bound to f_1. $\bar{f}_1 = 0.2$

$$\begin{cases} f_1 & \text{sampled from lognormal} \\ f_1 = 0.1245, & \text{if sampled value is} < 0.1245 \end{cases}$$

4.) We apply a strict bound to f_2. $\bar{f}_2 = 0.2$

$$\begin{cases} f_2 & \text{sampled from lognormal} \\ f_2 = 0.1250, & \text{if sampled value is} < 0.1250 \end{cases}$$

5.) We apply a strict bound to f_3. $\bar{f}_3 = 0.05$

$$\begin{cases} f_3 & \text{sampled from lognormal} \\ f_3 = 0.0545, & \text{if sampled value is} > 0.0545 \end{cases}$$

6.) We apply a strict bound to f_6. $\bar{f}_6 = 0.0238$

$$\begin{cases} f_6 & \text{sampled from lognormal} \\ f_6 = 0.0985, & \text{if sampled value is} > 0.0985 \end{cases}$$

7) We apply a strict bound to f_8. $\bar{f}_8 = 70,000$

$$\begin{cases} f_8 & \text{sampled from lognormal} \\ f_8 = 1.3501e + 05, & \text{if sampled value is} > 1.3501e + 05 \end{cases}$$

8.) We apply a strict bound to f_9. $\bar{f}_9 = 0.0117$

$$\begin{cases} f_9 & \text{sampled from lognormal} \\ f_9 = 0.0114, & \text{if sampled value is } < 0.0114 \end{cases}$$

For VP4

3.) We apply a strict bound to f_1. $\bar{f}_1 = 0.2$

$$\begin{cases} f_1 & \text{sampled from lognormal} \\ f_1 = 0.0405, & \text{if sampled value is } < 0.0405 \end{cases}$$

4.) We apply a strict bound to f_2. $\bar{f}_2 = 0.2$

$$\begin{cases} f_2 & \text{sampled from lognormal} \\ f_2 = 0.0450, & \text{if sampled value is } < 0.0450 \end{cases}$$

5.) We apply a strict bound to f_3. $\bar{f}_3 = 0.3$

$$\begin{cases} f_3 & \text{sampled from lognormal} \\ f_3 = 0.2983, & \text{if sampled value is } < 0.2983 \end{cases}$$

6.) We apply a strict bound to f_6. $\bar{f}_6 = 0.0238$

$$\begin{cases} f_6 & \text{sampled from lognormal} \\ f_6 = 0.0958, & \text{if sampled value is } > 0.0958 \end{cases}$$

7) We apply a strict bound to f_8. $\bar{f}_8 = 40,000$

$$\begin{cases} f_8 & \text{sampled from lognormal} \\ f_8 = 5.9096e + 04, & \text{if sampled value is} > 5.9096e + 04 \end{cases}$$

8.) We apply a strict bound to f_9. $\bar{f}_9 = 0.0177$

$$\begin{cases} f_9 & \text{sampled from lognormal} \\ f_9 = 0.0173, & \text{if sampled value is} < 0.0173 \end{cases}$$
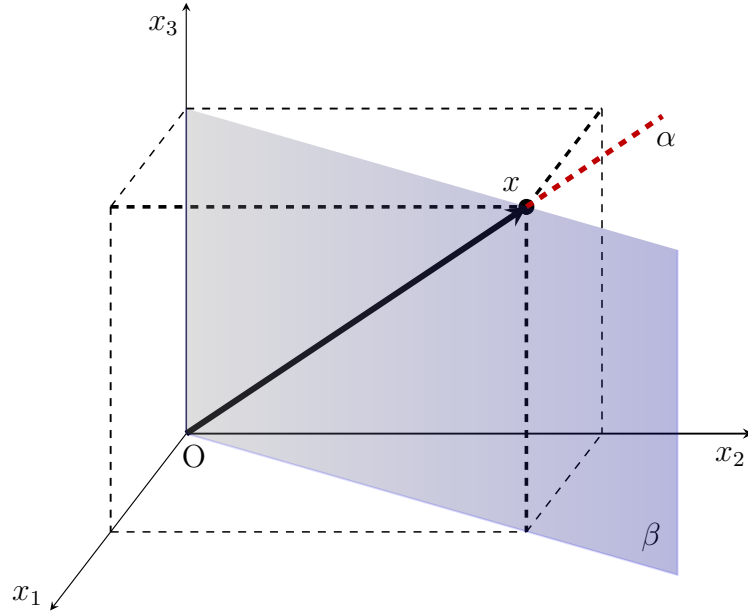
Figure 1.2: A 3-D representation of the metabolic state space which highlights states $\tilde{x}$ with nontrivial intersections with the kernel of the initial state (represented by x). A three dimensional representation is appropriate because (1.14) implies that x_1 and x_4 are proportional. The line α represents case (IV) where $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ are proportional to x_1, x_2, x_3 respectively. States $\tilde{x}$ on this line will have kernels such that $dimension(K(x) \cap K(\tilde{x})) = 3$. The plane β represents case (III) where only $\tilde{x}_1, \tilde{x}_2$ are proportional to x_1, x_2.

Chapter 2

Cis-regulatory module expression profile

2.1 Introduction to CRM analysis

Cis-regulatory modules (CRMs) control gene expression in an organism. CRMs may activate and deactivate over the organism's lifetime, and may be active or inactive across different cell types, or even individual cells in the organism. Using the purple sea urchin as a model organism, we have constructed a procedure to detect the spatial and temporal expression in the an organism in which putative CRMs fused to reporter "barcodes" have been introduced. Moreover the analysis is tested by a simulation programmed in Matlab. The code simulates the cell growth of the purple sea urchin and the random integration of DNA fragments into it's genome. Comprehension of genetic regulatory networks requires understanding of many regulatory genes, their target genes, and CRMs that control the interactions of regulatory genes and their targets [21]. These three components comprise gene regulatory networks. Studying CRMs has advantages over the other components, specifically 1. CRMs often control genes that are physically close by on a strand of DNA which facilitates locating these

target genes and 2. cell differentiation results from large groups of genes working together that are often activated with shared CRMs. That is, CRMs play a critical role in development and can uncover genetic mechanisms that result in a fully formed adult organism.

Cis-regulatory analysis measures temporal and spatial activities of CRMs in organisms that grow from normal and perturbed embryos. The conventional approach for a cis-regulatory analysis follows:

1. Build a reporter construct that contains a wild-type or mutated CRM, a core promoter that can bind RNA polymerase II, a reporter gene such as green fluorescent protein (GFP), and a core poly-(A)denylation signal.

2. Inject reporter constructs into cells or embryos.

3. Examine expression of the reporter gene.

4. Compare reporter expression with gene expression patterns to build a cis-regulatory model for gene expression control [22].

While this approach has been effective for the current understanding of cis-regulatory mechanisms, it is relatively slow compared to the progress in genomics for measuring gene expression patterns. Because many eukaryotic genes are controlled by multiple CRMs, experimentally studying one CRM at a time and exhausting all possibilities is impractical. This approach alone would lead to incomplete gene regulatory models. It follows that a high through-put method to analyze thousands of CRMs would help fill in the gaps of knowledge left by approaches relegated to a single CRM at a time.

Regarding the advantages of high-throuput analysis, several new assay methods for CRMs have been developed [23, 24, 25, 26, 27, 28, 29]. The new approaches take advantage of the immense diversity of DNA oligomers to barcode and track many CRMs in parallel. However, application of these methods has been limited to quantitative measurement of CRM activities in cells or embryos.

Here, using sea urchin embryos, we describe a procedure to analyze a different high-throughput and highly scalable experiment. It is called Multiplex and Mosaic Observation of Spatial Information encoded in Cis-regulatory modules (MMOSAIC). This method simultaneously measures both quantitative and spatial activities of CRMs. MMOSAIC is based on two well-known observations in a variety of model systems:

- Stochastic and mosaic incorporation of linear reporter constructs into only one cell in an early embryo.

- Unequal clonal replication of the incorporated reporter constructs depending on cell lineages during embryogenesis [30, 31, 32].

The level of reporter expression in a mosaic embryo is determined by the combination of intrinsic activity of a given CRM as well as specific cells that harbor the construct before the time of measurement. Since a large sample size neutralizes the effect of random mosaic DNA incorporation, the hypothesis is that the quantitative profiles of single-embryo resolution reporter expressions measured in a sufficiently large number of embryos is determined by spatial activity of a given CRM. Using a new single-embryo resolution reporter assay method, it is shown that the quantitative profile of single-embryo resolution reporter expressions measured in a large number of mosaic embryos can be used for spatial cis-regulatory analysis. To understand the proposed analytical procedure, the underlying experiment that is analyzed must be elaborated on.

2.2 Experimental set up

2.2.1 Preparing reporter constructs

The size of CRMs used in this study ranges from 351 bp to 2716 bp, and the majority of the CRMs are 2 kb-long. Reporter constructs were generated in two steps:

- Fuse individual putative CRMs with unique identifier (ID) secquences. These are referred to as CRM::ID constructs.

- The CRM::ID construct receives an additional 'extreme' barcode, which is a random 25-mer.

The 'extreme' barcode is added to the with one cycle of Polymerase Chain Reaction (PCR). To generate a large number of CRM::ID constructs, a pre-barcoded library of empty reporter vectors that already contain a promoter fused to a green fluorescent protein (GFP) pre-labeled with approximately 100 million random "barcodes" (25 base pairs long). The abundance of CRM constructs allows a very high probability that a sample contains no duplicate barcodes. Our analysis will assume that no duplicate barcodes will become integrated into the growing embryo.

2.2.2 Incorporation of barcoded reporter constructs

Equal amounts of pooled CRM::ID are introduced to an embryo of an organism as described in [23]. The expected number of unique barcodes delivered per putative CRM ≥ 1500 in the entire pool of embryos to be injected. The many differently barcodes per CRM leads to many experiments done simultaneously to see how the CRM can be expressed in the organism.

After the CRM::ID constructs are injected into the embryo, there is a chance that a construct will become incorporated into the DNA of the organism. From the time of this "incorporation event" onward, the CRM can be expressed as if it were part of

the original organism's genome. It is expected that such an incorporation event may happen only during mitosis (when the organism's DNA is exposed to the cytoplasm), and it will happen in one specific daughter cell. If the CRM::ID is contained in the DNA of a cell where the CRM will be expressed, transcription of the putative CRM and the attached barcode is expected to occur. Once the DNA is incorporated into a cell, it will duplicated during mitosis, and all daughter cells will contain the barcoded CRM.

2.2.3 Rank ordered profiles

From the integration events of CRM::ID constructs, the CRM and barcode will be copied to the next generation of cells. Therefore, to know in which cell generation a barcoded CRM was incorporated, it is a simple matter of measuring the relative copies of barcoded DNA in the organism after it grows. The cell lineage becomes relevant here, because not all cell types grow at the same rate in our example organism, the purple sea urchin. Ignoring this detail for a moment, it is easy to see that an integration event which occurred in the second to last cell generation will result in twice as much barcoded DNA that was incorporated, and an integration event that occurs two cell generations before the end of the experiment would lead to four times as much barcoded DNA (the amount of DNA was doubled twice, see Fig. 2.4).

Just as we can use analysis to determine when the incorporated DNA was integrated into the organism, we may also understand in which cells incorporated CRMs were expressed. The putative CRMs will be incorporated randomly into different cells. Randomly, a CRM will become incorporated in cells. If one of these expresses this CRM, there will be RNA transcribed from the CRM and the reporter. These copies of RNA for a barcode are called expressed barcodes. By observing the ratio of expressed barcodes to incorporated barcodes and considering the different speeds that cells grow and divide in the cell lineage, one can eliminate possible cells where

expression could not occur. With enough trials we can pinpoint which cell group and which cell generation expression of a barcode occurred. We now know that the CRM attached to the barcode is active in these cells.

To organize our data of expression over incorporation, we plot all barcodes along the x axis with their expression ratio (expressed barcodes to incorporated barcodes) on y axis so that as the highest ratios are plotted on the left of the x axis with descending ratios to the right. If there are any instances of no expression for the given CRM, these will be plotted to the far right on the x-axis. This plot is called a rank ordered profile (ROP). With a very large number of incorporation events, resulting ROPs contain the information needed to deduce which cell types and which cell generations a CRM was expressed.

2.3   Simulation Details

A simulation of the experiment described in section 2.2 was constructed in MAT-LAB. With the simulation, one can investigate An in silico version of the previously described experiment whereby one could specify the cells which would express a CRM. With this tool, we observe what the ROPs should look like when pre-selected cells drive the expression of a CRM. The MATLAB program simulates many random incorporation events according to the uniform probability over all cells which would undergo mitosis in each cell generation. The program then generates ROPs according to simulated experiments. This gave us a framework to test CRMs active in specific target cells will affect ROPs. The framework also let us test how the ROPs change if the expressed cells where spatially or temporally perturbed.
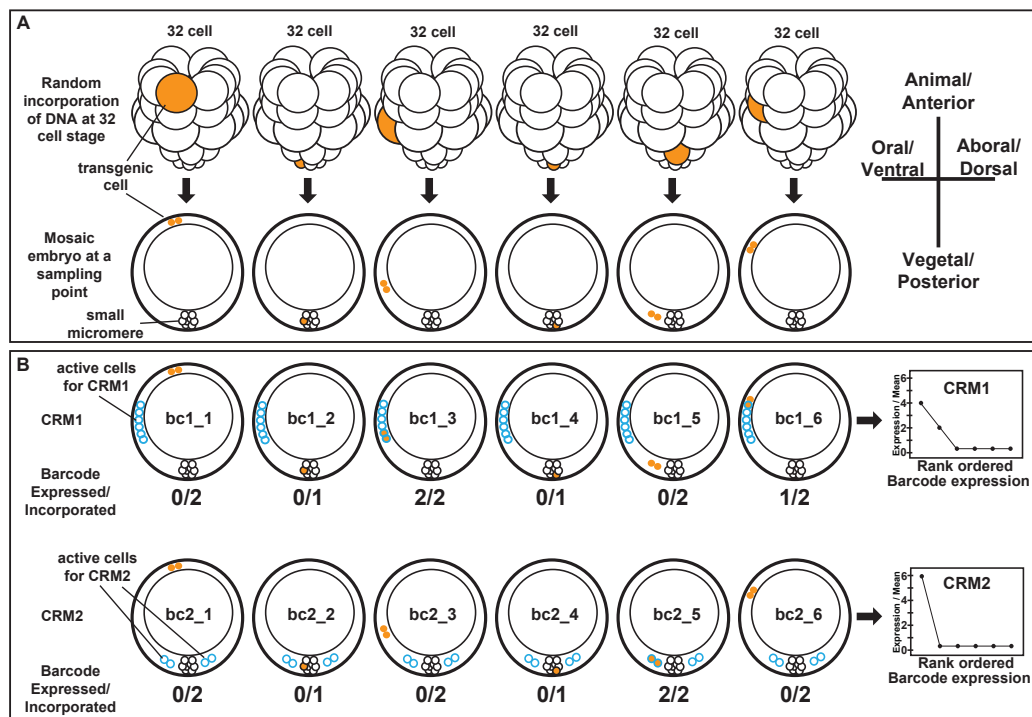
Figure 2.1: figure showing hypothetical example of random incorporation events leading to CRMs being incorporated in cells which sometimes express the CRM, and sometimes not. From many different barcodes fused to the same CRM we can construct a rank ordered profile which tells us how often a specific CRM is expressed and the ratio of expression to incorporation.

2.3.1   Reproducing the cell lineage

The cell lineage reproduced by my program was of a purple sea urchin, referenced from the book, "Development of Sea Urchins, Ascidians, and other invertebrate Deuterostomes: Experimental Approaches," by Charlse A. Ettensohn, Gary M. Wessel, and Gregory A. Wray [1] shown in Fig. 2.2.

The purple sea urchin embryo does not grow larger until later in the life cycle beyond the cell generations with which our experiment is concerned. During early generations, when the cell divides, it is by way of cell cleavage, where the two daughter cells are half the size of the parent cell. The cell cleavage can occur in three different ways:

1. A horizontal cleavage of a group of cells results in two groups with the same distribution of DNA in each group.

2. A vertical cleavage will result in two groups with potentially different DNA distributions.

3. An oblique cleavage is a combination of horizontal and vertical cell cleavages. Horizontal cell cleavage and oblique cell cleavage are shown in Fig. 2.3.

As a simplifying assumption which sidesteps the inherent ambiguity of oblique cleavage, all oblique cleavages in the simulation were written as horizontal cell cleavages. A cell group (represented by a node in Fig. 2.2) in the simulation is a vector of zeros, and the length of the vector doubles when the group undergoes a cleavage.

2.3.2   Simulated incorporation of barcoded CRMs

In the actual experiment, many incorporation events all with different barcodes will happen in a single embryo, and there are many embryos in the experiment. However

we assume that these incorporation events are independent, and so my in silico representation will simulate one barcoded CRM incorporation event at a time. In other words, any CRM behaves like any other with the exception of the cells which drive expression.

The simulation begins with a single number '0'. When this cell undergoes mitosis, the next generation will be represented by two 0's. The '0's double according to the cell lineage of the organism. With this scheme, one can represent a growing embryo. During the growth of the embryo, there is a random chance for a barcoded CRM to become incorporated into the DNA of a cell. For an incorporation event the specific cell in a given generation is chosen uniformly, however the generation in which it occurs is not uniform, and the simulation replicates this probability distribution for incorporation events (shown in Fig. 2.4).

A randomly generated number determines the generation in which a barcoded CRM becomes incorporated. After the generation is randomly chosen, every cell in this generation has an equal chance to be the cell in which the CRM incorporates. Each cell group is represented in the simulation as a vector of zeros, and when an incorporation event happens in a cell, the 0 representing this cell becomes a 1. Daughter cells of all cells designated by a '1' are also designated as a '1' which indicate the cells containing the barcode in all cell generations.

2.4 Results

2.4.1 Constructing a library of ROPs for known CRMs

Perhaps the most obvious result from this work is that one can catalog the rank ordered expression profiles of known CRMs. When a new putative CRM is verified to be active in an organism, the ROP can be compared to the catalog, and if the ROP

matches that of another CRM, then we know that our putative CRM is active in the same cells (see Fig. 2.7). In the case that a ROP doesn't match exactly then we still estimate information about the cells where the putative CRM is active.

To test the similarity of ROPs derived from CRMs which are expressed in similar or different cell groups of our simulation we define distance between ROPs for CRM_1 and CRM_2 as:

$$D_{\text{CRM}_1,\text{CRM}_2} = \frac{1}{N} \sum_{n=1}^{N} \frac{|B_{\text{CRM}_1,n} - B_{\text{CRM}_2,n}|}{\frac{\left(B_{\text{CRM}_1,n} + B_{\text{CRM}_2,n}\right)}{2}} \tag{2.1}$$

where $B_{\text{CRM}_1,n}$ and $B_{\text{CRM}_2,n}$ are the nth ranked barcode expression for two CRMs respectively. This is the absolute difference of expression for the nth ranked barcode normalized to the mean expression of the two.

Simulations were run that calculated the distance between ROPs from a CRM active in the Nodal gene, Nodal_5P and Nodal_INT. The distance was small compared to the distance between the ROP corresponding to a CRM active in the Nodal gene and one active in the Delta gene (shown in Fig. 2.5).

2.4.2 High-throughput *cis*-regulatory analysis at single-embryo resolution

The new quantitative method for single-embryo resolution *cis*-regulatory analysis begins by delivering each uniquely barcoded reporter construct into only one embryo during the entire experiment. Many embryos are pooled with mosaic DNA incorporation (hereafter called mosaic embryos), transcripts of a unique barcode originate from one embryo. Note that a single embryo can contain many different reporter constructs, as long as each barcode is unique. The practical experiment requires new

'extreme' barcoding technology, with which one can add many billions of unique random 25-mers (N25) to established constructs that already contain unique identifier (ID) sequences (Fig. 2A). Delivering a small sampling of the reporter constructs into embryos means that probabilistically each unique barcode is introduced into only one embryo.

To measure reporter expression and to correct for DNA copy number at a desired stage, total DNA and total RNA is extracted from the same lysate of embryos (Fig. 2C). PCR with a pair of universal primers selectively amplifies the entire set of barcodes from total genomic DNA and total cDNA, and next-generation sequencing is used for counting the relative copy numbers of barcodes. Because the sequence reads also contain IDs specific to individual CRMs, each N25 barcode can be computationally assigned to a CRM. Relative expression level of a unique barcode in a single embryo (B) is computed by the following equation.

$$B = \frac{E}{I} \tag{2.2}$$

where E is the number of sequence reads originated from expressed unique barcode and I the number of sequence reads from incorporated unique barcode. Note that I is proportional to the number of cells that carry the barcode within a single embryo, as each barcode is present in only one embryo. One can also estimate the traditional and averaged relative CRM activity (A) measured by reporter expressions using the following equation.

$$A = \frac{(E_1 + E_2 + E_3 + \ldots + E_N)}{(I_1 + I_2 + I_3 + \ldots + I_N)} \tag{2.3}$$

where E_N and I_N respectively are the numbers of sequence reads from expressed and incorporated Nth unique barcode, and N the total number of unique barcodes driven by a given CRM. Because the scales of B and A also depend on the number of sequence reads, we further normalize the data to the background activity measured by using

a co-injected inactive DNA fragment. The advantages of background normalization are reduction of variations among samples and elimination of the need for measuring absolute levels [23, 24]. Nodal_3P fragment is used to measure background activity, because there was no detectable positive or negative activity observed with the fragment [33].

2.4.3 Temporal DNA profile

At the end of the last cell generation to be analyzed, the organism's cells are lysed and we numerically examing relative numbers of sequence reads from the lysate. In the underlying experiment, the copy number of a barcode cannot exceed the number of cells in the organism (a barcode found in every cell would signify an incorporation even occurring in the very first cell before mitosis, and this is the maximum example). The relative amount of sequence reads of the lysate will reflect the generation in which the incorporation event occurred, as well as the cell fate of the first cell to harbor the barcode. Referencing the purple sea urchin cell lineage map (Fig. 2.2) , one can see that the number of cells after each cell generation increases by a factor close to 2. After the first cell division there are 2 cells, then 4, 8, 16, 32 ,60 ,96, 104, 216, 424 (this is a correction to the typo "408" written on the original picture), 808. The factor is not exactly 2 because of the different speeds at which some cell types divide. One can find the average copy number of an incorporated barcode is $\frac{400}{c_n}$ where c_n is the total number of cells in the embryo at the time of the incorporation event. The number of sequence reads are proportional to the number of copies of a barcode. One can bin the number of reads by which cell generation the incorporation event happened in. As you can see in Fig. 2.2 after the fifth cell division, the occurrence of mitosis becomes less synchronous across cell groups.

2.4.4 Spatial DNA profile

For a known ROP for a given CRM, one can anticipate the effect of perturbations to the spatial regions of cells which drive expression. Simulations were run where the region of cells driving CRM expression was perturbed, and P-values are given for each comparison. The CRM responses have not been confirmed by independent experiments.
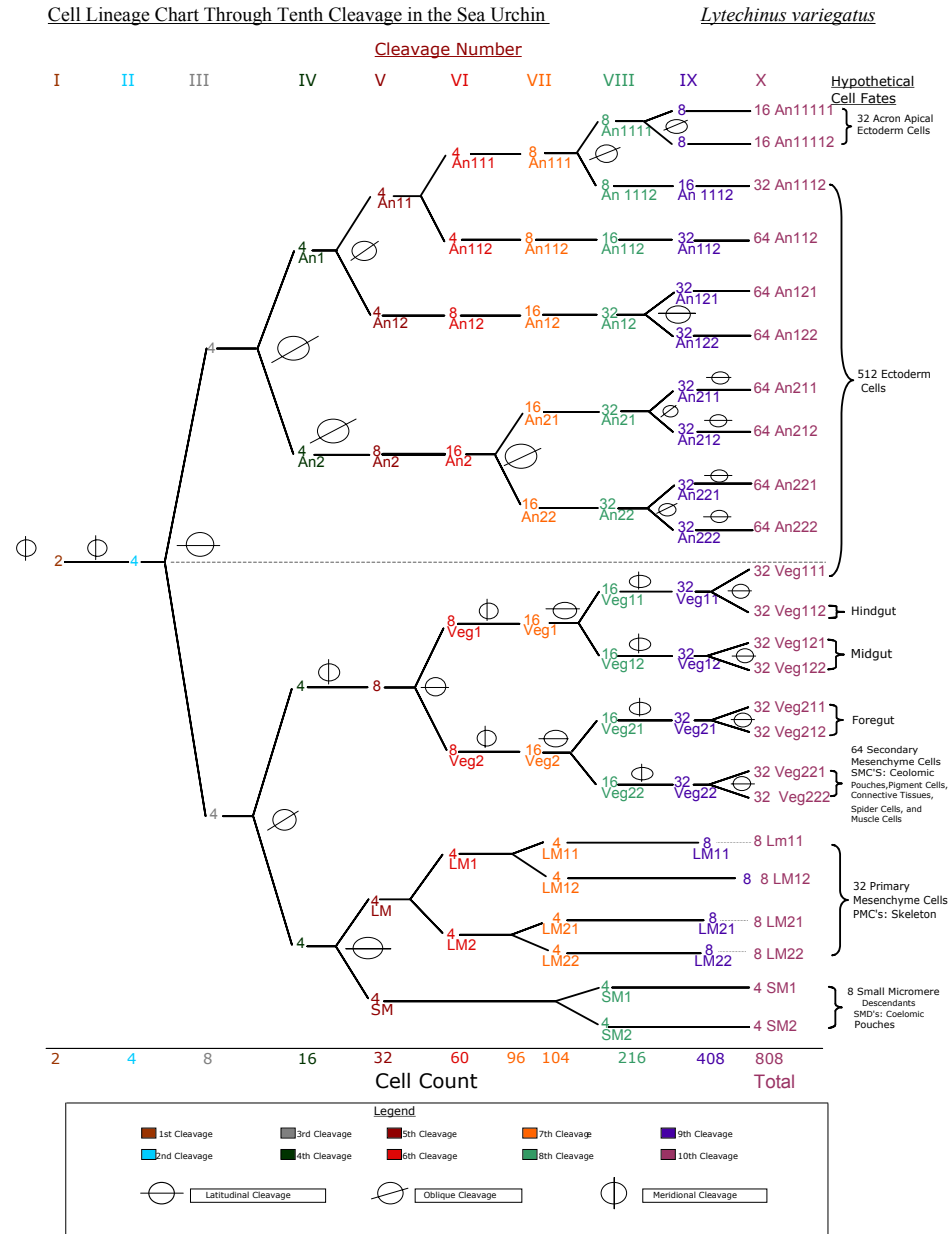
Figure 2.2: The cell lineage of the purple sea urchin as given by [1]. The nodes of the graph indicate a group of cells, and the numbers indicate the number of cells in the group. Some groups are labeled as a particular cell type. In the cell count just below the cell lineage, the count of 408 should be 424, the original picture was kept for citation purposes.
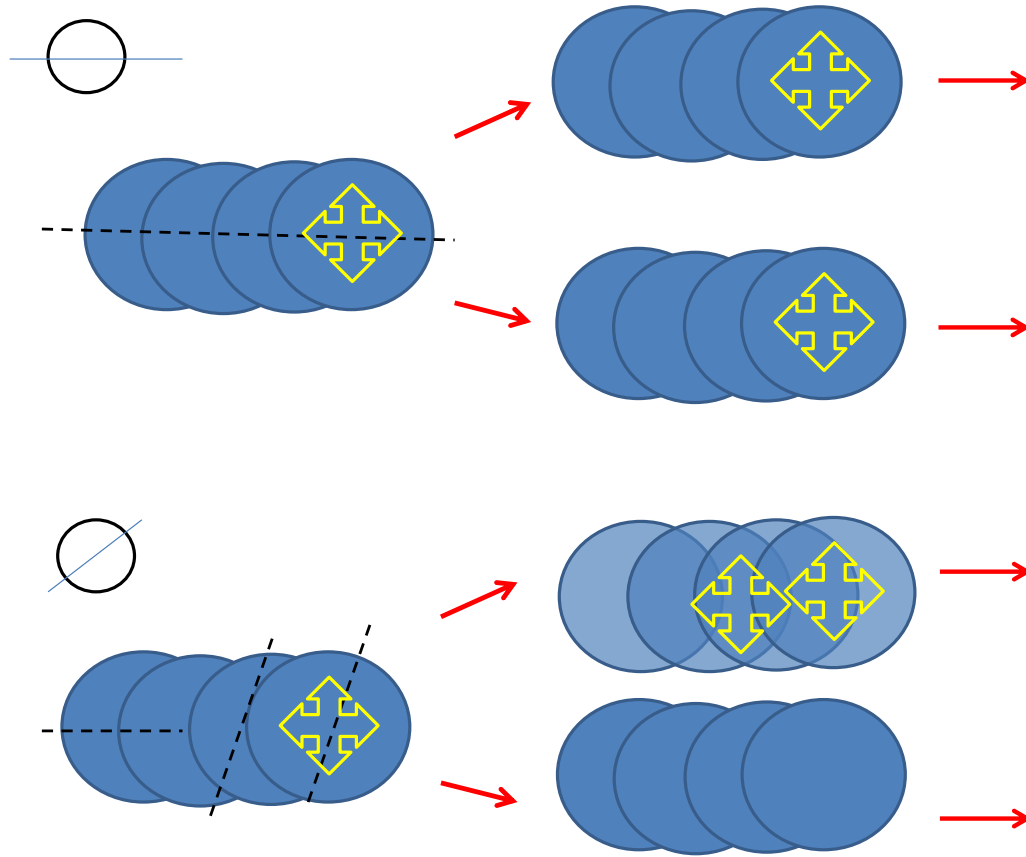
Figure 2.3: When cells cleave, they may do so in different ways. The cell with the yellow cross contains a barcoded CRM. The top shows a horizontal cleavage (identified by a circle with a horizontal line). The cartoon depicts how the two groups of daugter cells will cleave from the parent cells. The bottom shows an oblique cleavage (identified by a circle with a diagonal line) which is a combination of horizontal and vertical cleavages. The cell with the cross is cleaved vertically, and so both daughter cells end up in the same cell group.
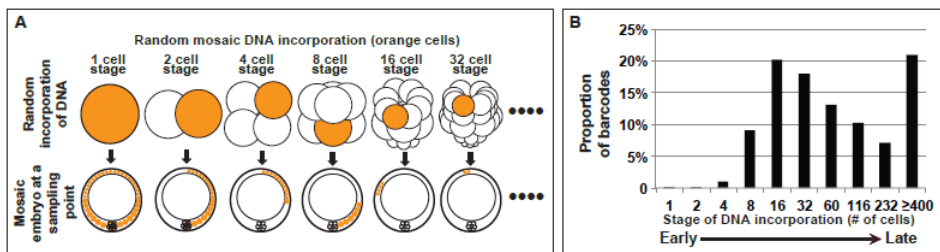

Figure 2.4: (A) When a cell is randomly incorporated, the stage of incorporation will determine how many copies of barcoded CRMs there are. (B) The distribution describing the empiricle frequency of incorporation events shows that incorporation events are more likely to happen in some cell generations than others.
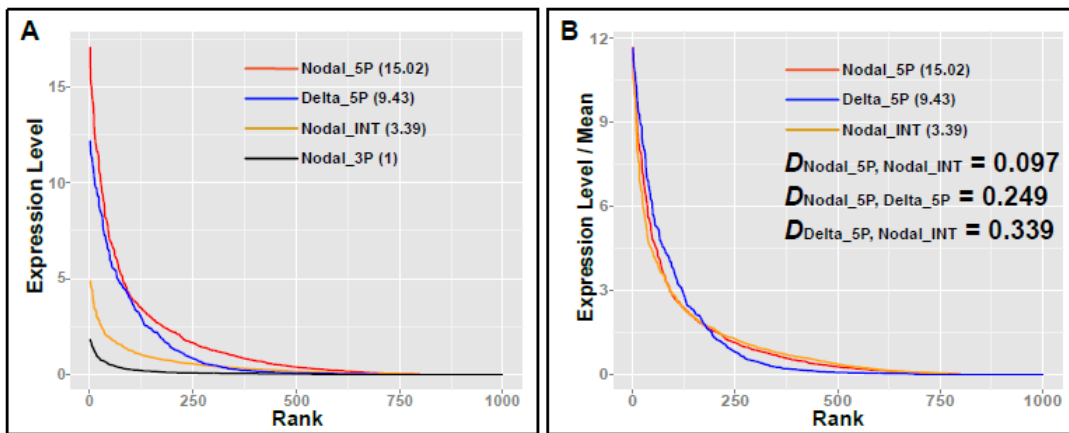
Figure 2.5: (A) We compare serveral ROPs for similarity. ROPs driven by cells in the same cell groups appear similar to each other. (B) The distance between two ROPs is calculated with equation (2.1) in order to quantify their similarity.

Figure 2.6: Figure depicting the ability to reference a catalog of ROPs which provides information on newly discovered CRMs

Figure 2.7: (A) Expected changes in ROPs when region of activity expands or contracts. (B) Delta_5P expression and Delta_5P expanded by nodal MOE. (C) Nodal_5P expression and expanded expression. (D) Nodal_INT expression and expanded expression.

Chapter 3

Opinion formation models

3.1 Introduction to opinion formation models on Riemannian manifolds

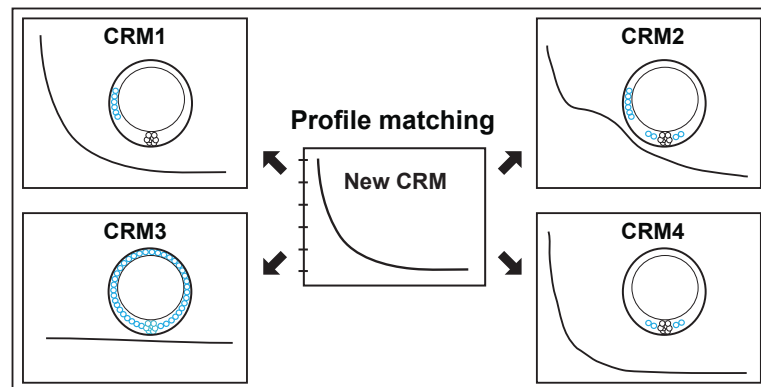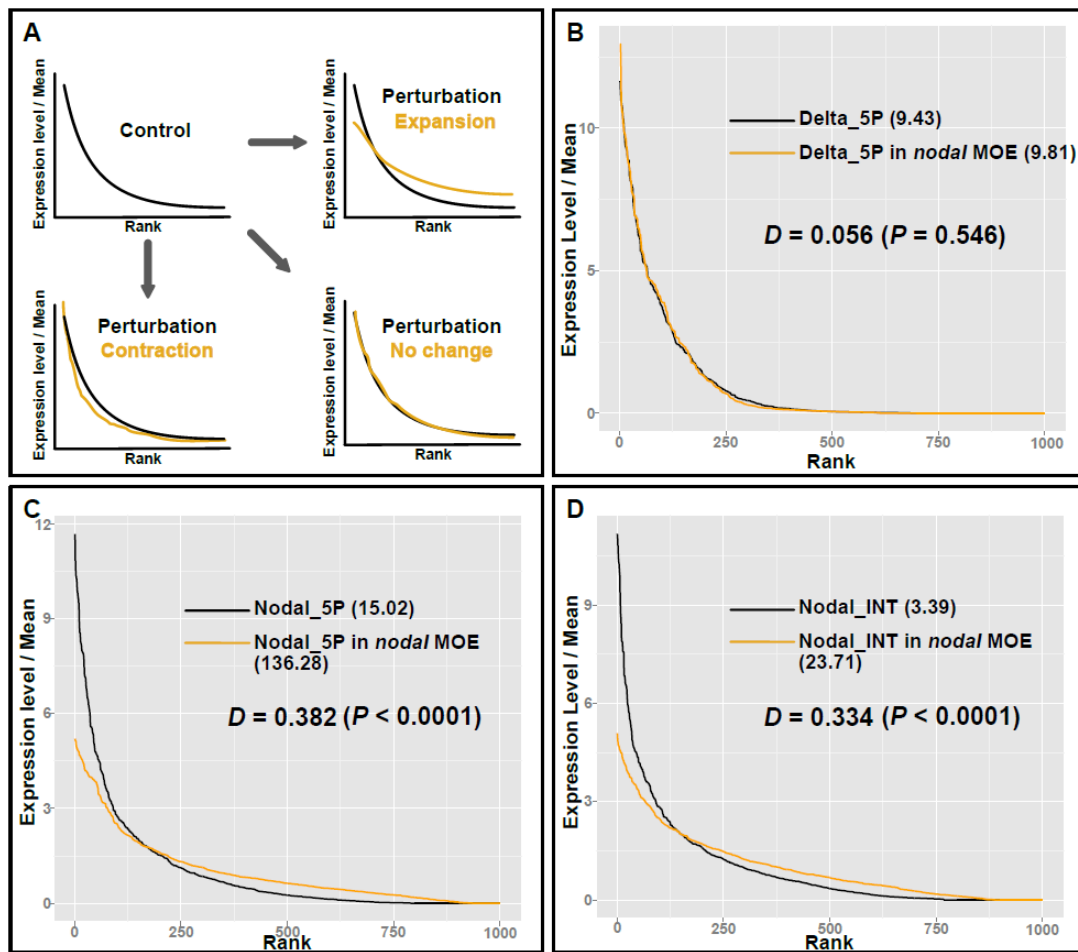The emergence of a group's global behavior from local interactions among individual agents is called emergent behavior. When local rules imply global patterns in a population, we are observing a phenomenon called *self-organization*. Traditionally, researchers focus on understanding the complex rules of interacting opinions which lead to certain global configurations, such as classic *consensus*, *alignment*, *clustering*, or the less studied *dancing equilibrium* [34]. Many different types of opinion models have been studied, such as bounded-confidence models like the one proposed by Hegselmann and Krause, the radius of interaction determines the clustering of the system [2]. Motsch and Tadmor studied the influence of the shape of the interaction potential on the convergence to consensus of the Hegselmann-Krause system [35]. Ha, Ha and Kim looked at the Cucker-Smale second-order alignment model and provided a condition on the interaction potential ensuring convergence of the system to alignment [36]. Cristiani, Frasca and Piccoli studied the effect of anisotropic interactions on the behavior of the group [37].

The dynamics of an opinion model depend on the state-space [38] and interaction network [2]. Models on the Euclidean space in one dimension (for opinion dynamics) or in two or three dimensions (with applications to groups of animals or robots) have been extensively studied and are well understood. However, such models are locally linear, which may limit the ability to capture more complex phenomena and represent reality [39]. In this vein, the Kuramoto model on the circle $\mathbb{S}^1$ addresses the problem of synchronizing a large number of oscillators [40, 41]. There exist numerous applications to this model [42, 43, 44, 45]. Similarly, applications to automobiles or satellite coordination have motivated the development of models on special orthogonal groups [46, 47]: satellites evolve on $SO(3)$ while automobiles evolve on $SE(2)$ or $SE(3)$. A nonlinear model of opinion formation on the sphere was also developed in [34].

The present work defines a general model of opinion dynamics on a compact Riemannian manifold. We investigate how the manifold on which the model is defined, i.e. the state space, affects the global configurations resulting from opinion dynamics. These are the first steps to constructing a mathematical framework for the theory of opinion dynamics on general Riemannian manifolds.

Perhaps the most apparent difficulty in defining opinion dynamics on a general Riemannian manifold is the possible lack of unique geodesics between two points. Using the Riemannian distance, an agent will move towards a point by along this geodesic, which are well defined only locally. This problem arises even after we assume that one can overcome the vast complexity of computing geodesics, even on a relatively simple manifold such as the torus [48]. A method to bypass this issue is to consider the embedding of the manifold into a Euclidean space. Each agent's velocity is defined by projection of the other agents' influence onto the tangent space at that point. This is the choice made in [34].

Other than the mentioned practical convenience, there is theoretical reasoning

behind choosing to define the model in terms of geodesic distance versus a distance involving an embedding using projections through euclidean space. When evolving along the geodesics of the manifold, one assumes that each agent has a global understanding of the manifold's geometry and is able to choose the shortest path among all possible ones. This method imbues each agent with a comprehensive knowledge of the state space, whereas the projection-based approach restricts each agent's spatial understanding to its own point of view, that is the "view" of the manifold from its tangent space.

These two specific approaches for a generalized model are explored. The first method, Approach A, uses projections in the Euclidean space in which the manifold is embedded. The second method, Approach B, uses only geodesics defined on the manifold to define strength and direction of interaction. Then the properties of the interaction matrix that lead to specific types of equilibria are investigated. Simulations and examples compare the two methods. Dancing equilibria for Approach B are shown (dancing equilibria were studied for Approach A in [34]).

The sphere and torus are used as example manifolds to evaluate these approaches. Specifically, simulations are given for dynamics on $\mathbb{S}^1, \mathbb{S}^2$, and $\mathbb{T}^2$. These examples allow one to directly compare the two approaches, and see if an approach is more practical for a given manifold. The influence of the manifold's geometry on the dynamics is examined by observing the dynamics resulting from the same interaction matrix in $\mathbb{S}^2$, $\mathbb{T}^2$ and $\mathbb{R}^2$.

Opinion dynamics trajectories can resemble solutions to the $n-$body problem. Opinion dynamics trajectories along such orbits are referred to as "Social Choreography." This occurs when agents are driven along orbits which either are periodic, or have a periodic feature, and that may be shared by multiple agents. A simple example of Social Choreography in $\mathbb{R}^2$ is shown to not hold on $\mathbb{S}^2$ or $\mathbb{T}^2$, see Figures 3.10 and 3.11. For $\mathbb{R}^2$ initial conditions and properties of the interaction matrix which

give rise to Social Choreography are investigated.

Another interesting feature of opinion models the ability to describe phenomena where agents are not equally influential. In this type of model, it is not only the distance separating agents that governs the influence of one agent over another, but also an a priori decision about which agents are more influential. We can introduce such heterogeneity of influence by prescribing a weight to an agent. A higher(lower) weight means that the agent is more(less) influential on all other agents.

The second half of this chapter considers a framework for which, in addition to the pairwise interaction function, a time-varying weight of influence is attributed to each agent. This weight allows us to model a dynamic, social hierarchy within a population, where the most influential agents (i.e. those with the larger weights) may not be known initially. The dynamics of the weights can be prescribed in a number of ways. We present different models coupling the opinions' dynamics and the weights' dynamics. We refer to Piccoli and Rossi [49] for a recent model with time-evolving masses for pedestrian dynamics.

3.2   Choice of the model

This work will primarily discuss two approaches to define opinion dynamics on a Riemannian manifold. Let M be a Riemannian manifold. Let $N \in \mathbb{N}$ represent the number of agents with opinions evolving on M. We denote by $x := (x_i)_{i \in \{1,...,N\}} \in M^N$ the set of opinions. For each $i \in \{1,...,N\}$, $\dot{x}_i \in T_{x_i} M$. The opinions x_i evolve according to the following general dynamics:

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} \Psi(d(x_i, x_j)) \nu_{ij} \qquad (3.1)$$

where

- $a_{ij} \in \mathbb{R}$ is the interaction coefficient of the pair of agents i and j,

- $\Psi : \mathbb{R} \to \mathbb{R}$ is the interaction potential,

- $d(\cdot, \cdot) : M \times M \to \mathbb{R}^+$ represents the distance between opinions,

- $\nu_{ij} \in T_{x_i} M$ is a unit vector giving the direction of the influence of j over i.

Each of these terms is further specified in the following.

3.2.1 Approaches

The evolution of each agent's opinion depends on the opinions of all other agents, with influences weighted by the interaction coefficients a_{ij}. More specifically, an agent x_j's influence on x_i is determined by two elements: the direction of influence $\nu_{ij} \in T_{x_i} M$ and the magnitude of influence $\Psi(d(x_i, x_j)) \in \mathbb{R}^+$. We propose and study two different approaches for the choices of d and ν_{ij}. Approach A uses the embedding of M in $\mathbb{R}^n$ to define $d(x_i, x_j)$, whereas Approach B is intrinsic to M, with distance and direction of influence based on geodesics.

Approach A Assume that M of dimension m is embedded in a Euclidean space $\mathbb{R}^n$, with $n \geq m$. Agent x_j acts on agent x_i via a projection onto $T_{x_i} M \subset \mathbb{R}^m$. Now considering points $(x_i, x_j) \in M^2$ as points of $\mathbb{R}^n$, the difference $x_j - x_i$ is a vector of $\mathbb{R}^n$. Given a vector subspace Y of $\mathbb{R}^n$, we denote by $\Pi_Y y$ the projection of $y \in \mathbb{R}^n$ onto $Y \subset \mathbb{R}^n$ and define $d_P(\cdot, \cdot)$ as follows:

$$d_P(x_i, x_j) = \|\Pi_{T_{x_i} M}(x_j - x_i)\| \tag{3.2}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$. The same projection also defines the direction of influence of x_j on x_i:

$$\nu_{ij} = \begin{cases} \frac{\Pi_{T_{x_i}M}(x_j - x_i)}{\|\Pi_{T_{x_i}M}(x_j - x_i)\|} & \text{if } \Pi_{T_{x_i}M}(x_j - x_i) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

With the specific choice $\Psi \equiv \text{Id}$, system (3.1)-(3.2)-(3.3) becomes:

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} \Pi_{T_{x_i}M}(x_j - x_i). \tag{3.4}$$

This is the approach used in [34], applied to the sphere $\mathbb{S}^2$.

Notice that the magnitude of influence, $d_P(x_i, x_j)$, is symmetric for the sphere in the sense that $d_P(x_i, x_j) = d_P(x_j, x_i)$, but not symmetric for a general Riemannian manifold (see Figure 3.1). However, it is a continuous function defined for all pairs of points $(x_i, x_j) \in M^2$. The originality of this approach is that the influence of x_j on x_i is not related to a notion of distance between the points. The use of the projection of $x_j - x_i$ onto $T_{x_i}M$ reflects the concept of "local visibility." For the situation of two agents evolving on a one dimensional manifold, if $x_j - x_i \perp T_{x_i}M$, then a local displacement of x_i does not affect the distance between the points $\|x_i - x_j\|$. Indeed, a first order Taylor expansion gives: $x_i(\varepsilon) = x_i(0) + \varepsilon \dot{x}_i(0) + o(\varepsilon)$.

Supposing that x_j is fixed, we have:

$$\begin{aligned} \|x_i(\varepsilon) - x_j\|^2 &= \langle x_i(\varepsilon) - x_j, x_i(\varepsilon) - x_j \rangle \\ &= \langle x_i(0) - x_j, x_i(0) - x_j \rangle + 2\varepsilon \langle \dot{x}_i(0), x_i(0) - x_j \rangle + o(\varepsilon) \end{aligned} \tag{3.5}$$

so if $x_j - x_i(0) \perp T_{x_i(0)}M$, then $\|x_i(\varepsilon) - x_j\|^2 = \|x_i(0) - x_j\|^2 + o(\varepsilon)$. Hence if x_i only has local visibility, all directions of displacement seem equivalent (at first order), which justifies the influence of x_j over x_i to be zero if their difference is orthogonal
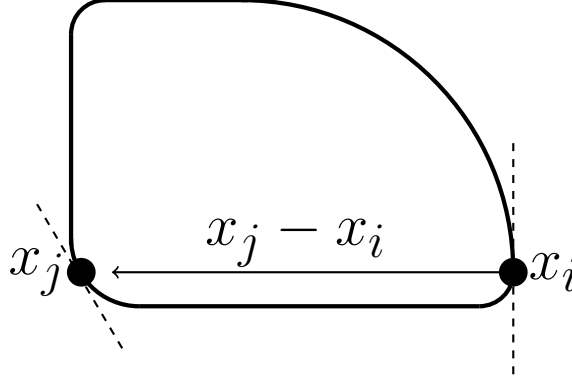
Figure 3.1: An example of a manifold M such that $d_p(x_i, x_j) \neq d_p(x_j, x_i)$, Using system (3.4), an agent is subject to "local visibility", and movement of x_i along $T_{x_i}M$ (dashed line through x_i) will not bring x_i closer to x_j in this local sense.

to the tangent space of M at x_i. This is illustrated in Figure 3.1.

Approach B This second approach defines d and ν_{ij} using the manifold M itself, and does not require any reference to the space in which M is immersed. This would make Approach B a natural way to define system dynamics, however the complete knowledge of the geodesics between any two points on the manifold must be knowable by the agents which may be unrealistic. Furthermore, the geometry of the manifold may introduce difficulties to the uniqueness of ν_{ij}, particularly at the cut-locus of a point.

Definition 6 *The **cut locus** of a point $q \in M$ is the set of points $\mathcal{CL}(q) \subset M$ for which there are multiple geodesics between q and $p \in \mathcal{CL}(q)$ (see also [50]).*

Let $\gamma_{ij} : [0, 1] \to M$ denote a geodesic connecting x_i to x_j, $\gamma_{ij}(0) = x_i$ and $\gamma_{ij}(1) = x_j$. We then define the distance between x_j and x_i as the length of a geodesic, i.e. denoting by $g_y : T_yM \times T_yM \to \mathbb{R}^+$ the Riemannian metric at point $y \in M$,

$$d_G(x_i, x_j) = \int_0^1 \sqrt{g_{\gamma_{ij}(s)}(\dot{\gamma}_{ij}(s), \dot{\gamma}_{ij}(s))} ds. \tag{3.6}$$

The direction of influence is determined by the same geodesic:

$$
\nu_{ij} = \begin{cases} 0 & \text{if } x_j = x_i \text{ or if } x_j \in \mathcal{CL}(x_i) \\[2ex] \dfrac{\dot{\gamma}_{ij}(0)}{\sqrt{g_{x_i}(\dot{\gamma}_{ij}(0), \dot{\gamma}_{ij}(0))}} & \text{otherwise.} \end{cases} \tag{3.7}
$$

Unlike in Approach A, the magnitude of influence is a symmetric function: $d_G(x_i, x_j)$ $= d_G(x_j, x_i)$. Furthermore, this approach ensures that the magnitude of influence of one agent on another is a function of the exact Riemannian distance between the agents.

Interaction networks. In finite-dimensional systems such as system (3.1), the set of interacting agents can be described by vertices of a graph. A directed edge exists from a vertex i to a vertex j if and only if $a_{ij} \neq 0$. The system depends on the interaction network, and likewise, if the coefficients a_{ij} are chosen to be functions of the state, the interaction network may change as a result of the dynamics. Two main types of interaction networks have been proposed in the literature: metric interactions and topological interactions. If interactions between agents occur only locally, only the neighbors of agent i influence agent i. Metric interactions define the set of neighbors of agent i, given a radius $r > 0$, as

$$
\mathcal{S}_i^r(x) = \{j \in \{1, \ldots, N\}, d(x_i, x_j) \leq r\}, \tag{3.8}
$$

where $d(\cdot, \cdot)$ can represent either the projection or the geodesic distance, as specified in each of the two approaches described above (see equations (3.2) and (3.6)). The other main type of interactions specifies that an agent is influenced by only its k closest neighbors. We call these topological interactions [38]. We define the relative separation between two agents as $\alpha_{ij} = \mathbf{card}\{k : d(x_i, x_k) \leq d(x_i, x_j)\}$, The set of neighbors of agent i is then defined as the set of its k closest neighbors, i.e. for a
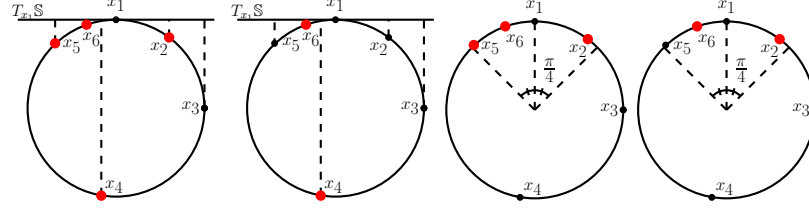
Figure 3.2: The set of agents that influence x_1 depends on how the interaction network is defined. In (a) and (b) the dashed lines show the projection of agents onto the tangent space of x_i, $(T_{x_i}\mathbb{S}^1)$. The agents depicted in red with larger dots influence x_1. With the same configuration on $\mathbb{S}^1$, four combinations are possible (approach {A,B} type {Metric, Topological}). Each combination implies x_1 interacts with a different set of agents.

given $k \in \mathbb{N}$,

$$\mathcal{S}_i^k(x) = \{j \in \{1, \ldots, N\}, \alpha_{ij} \leq k\}. \tag{3.9}$$

Figures 3.2 and 3.3 illustrate differences between the metric and topological networks for the specific example of $\mathbb{S}^1$, with each of the approaches A and B.
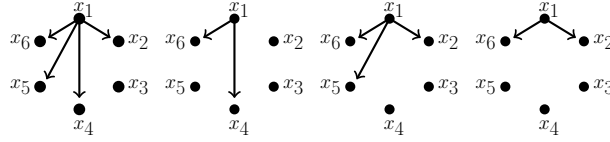
Figure 3.3: The agent x_1 is influenced by different agents depending on how the interaction network is defined. These networks may change as the dynamics move the agents on $\mathbb{S}^1$. Each agent $x_j, j \in \{1, \ldots, 6\}$ will have a network describing which other agents influence x_j. The interaction networks corresponding to systems from Figure 3.2.

Resolution of discontinuities. The definitions of ν_{ij} for approaches A and B (given by equations (3.3) and (3.7)) allow discontinuities of ν_{ij} at certain points. Thus, one must impose conditions on the interaction potential $\Psi \in \mathcal{C}^0(\mathbb{R}^+, \mathbb{R}^+)$, in order to ensure the continuity of the right-hand side of the system (3.1), and hence the existence and uniqueness of a solution. Table 3.1 lists the discontinuities of ν_{ij} and gives necessary conditions on Ψ to ensure the continuity of $\Psi(d(x_i, x_j))\nu_{ij}$.

Firstly, notice that in both approaches, ν_{ij} is discontinuous at the point $x_i = x_j$.

Indeed, if $x_i = x_j$, $\nu_{ij} = 0$, whereas almost everywhere else, $\|\nu_{ij}\| = 1$. To ensure the continuity of $\Psi(d(x_i, x_j))\nu_{ij}$ at this point, we impose the following condition:

$$\Psi(0) = 0. \tag{3.10}$$

In Approach A, we created a discontinuity of ν_{ij} at the points $x_j \in \mathcal{N}(x_i)$, where we denote by $\mathcal{N}(q)$ the set $\mathcal{N}(q) := \{q \in M \mid \Pi_{T_p M}(q - p) = 0\}$. For convenience of notation, we will use interchangeably the notations $\mathcal{N}(x_i)$ and $\mathcal{N}_i$. More specifically, we have $\lim_{x_j \to \mathcal{N}_i} \|\nu_{ij}\| = 1$ but $\|\nu_{ij}\| = 0$ if $x_j \in \mathcal{N}_i$ (see also Table 3.1). However, from the definition of d_P (see equation (3.2)), we have $\lim_{x_j \to \mathcal{N}_i} d_P(x_i, x_j) = 0$ and $d(x_i, x_j) = 0$ for $x_j \in \mathcal{N}_i$. Hence a sufficient condition for $\Psi(d(x_i, x_j))\nu_{ij}$ to be continuous is again:

$$\Psi(0) = 0. \tag{3.11}$$

In Approach B, there is a discontinuity for $x_j \in \mathcal{CL}(x_i)$. Denoting by $B^{\mathrm{geo}}(p, \rho)$ the geodesic ball of center p and radius ρ, we require the following condition on the influence function Ψ:

$$\Psi(d) = 0 \text{ for all } d \geq \epsilon \tag{3.12}$$

where $\epsilon := \inf\{\rho > 0 \mid \forall p \in M, B^{\mathrm{geo}}(p, \rho) \cap \mathcal{CL}(p) = \emptyset\}$. This distance ϵ, also known as injectivity radius, is known to exist and be greater than 0 for any compact Riemannian manifold (see [50]).

Approach	A	B	A and B
Critical points	$x_j \in \mathcal{N}_i$	$x_j \in \mathcal{CL}(x_i)$	$x_j = x_i$
Discontinuities	$\lim\limits_{x_j \to \mathcal{N}_i} \|\nu_{ij}\| = 1$	$\lim\limits_{x_j \to \mathcal{CL}(x_i)} \|\nu_{ij}\| = 1$	$\lim\limits_{x_j \to x_i} \|\nu_{ij}\| = 1$
	$\|\nu_{ij}\| = 0$ for $x_j \in \mathcal{N}_i$	$\|\nu_{ij}\| = 0$ for $x_j \in \mathcal{CL}(x_i)$	$\|\nu_{ii}\| = 0$
Condition on Ψ	$\Psi(0) = 0$	$\Psi(d) = 0$ for all $d \geq \epsilon$	$\Psi(0) = 0$

Table 3.1: Possible discontinuities of the right-hand side of (3.1). The bottom row of the table show conditions for Ψ so that the system is continuous.

Notice that in the case of the geodesics approach (Approach B), the condition $\Psi(d) = 0$ for all $d \geq \epsilon$ is incompatible with the use of the topological network (3.9). Indeed, if agent j is among the k closest neighbors of agent i, the topological network would require: $a_{ij} \neq 0$. However, the interaction between i and j would be canceled if $d_G(x_i, x_j) > \epsilon$. On the other hand, the metric interaction network as defined by (3.8) is compatible with Approach A, and with Approach B if the interaction radius is smaller than the injectivity radius: $r \leq \epsilon$. For simplicity purposes, in the rest of this paper, we will consider that the interaction coefficients a_{ij} are constant, thus not requiring the need to differentiate between metric and topological networks. While models with constant interaction coefficients are our focus here, these models are quite restrictive, and exclude all models with dynamic interactions.

3.2.2   Definitions and general results

Definition 7 *The configuration $x_1 = ... = x_N$ is called **consensus**. On the sphere, $\mathbb{S}^n$, A configuration such that, for every $j \in \{2, \ldots, N\}$, either $x_j = x_1$ or $x_j = -x_1$, which is not a concensus is called **antipodal equilibrium**.*

Proposition 6 *The consensus configuration is an equilibrium for system (3.1).*

Proof 8 *In both approaches A and B, if $x_i = x_j$, then $\nu_{ij} = 0$. Hence if $x_1 = ... = x_N$, then for all $i \in \{1, \ldots, N\}$, $\dot{x}_i = 0$.*

Proposition 7 *Let $N > d + 1$. Then for every $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N) \in M^N$, there exists a square matrix $A = (a_{ij})_{i,j \in \{1,\ldots,N\}}$ such that $\bar{x}$ is an equilibrium for system (3.1).*

Proof 9 *The configuration $\bar{x} = (\bar{x}_1, \ldots, \bar{x}_N)$ is an equilibrium if and only if*

$$\frac{d}{dt}\bar{x}_i = \sum_{j=1}^{N} a_{ij} \Psi(d(\bar{x}_i, \bar{x}_j)) \nu_{ij} = 0.$$

This is a system of at most Nd equations in the $N^2 - N$ unknowns a_{ij}, $i \neq j$, notice that $\Psi(d(x_i, x_i))\nu_{ii} = 0$, and diagonal values of A do not change the system. So if $N > d + 1$ there exists a nontrivial choice of the interaction coefficients for which $\bar{x}$ is an equilibrium.

Definition 8 *The kinetic energy of System* (3.1)-(3.2)-(3.3) *is the quantity*

$$E_P(t) := \frac{1}{2} \sum_{i=1}^{N} \|\dot{x}_i(t)\|^2. \tag{3.13}$$

The kinetic energy of System (3.1)-(3.6)-(3.7) *is the quantity*

$$E_G(t) := \frac{1}{2} \sum_{i=1}^{N} g_{x_i}(\dot{x}_i(t), \dot{x}_i(t)). \tag{3.14}$$

Proposition 8 *Let M be a general Riemannian compact manifold. Consider the dynamics given by projection onto the tangent space (Approach A) given by* (3.4). *If the interaction matrix $A = (a_{ij})_{i,j \in \{1,...,N\}^2}$ is symmetric, then*

$$\lim_{t \to \infty} E_P(t) = 0. \tag{3.15}$$

Proof 10 *Let $F(t) = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ij} \|x_i - x_j\|^2$. Using the symmetry of A, we prove that*

$$\frac{d}{dt} F(t) = 4E_P(t). \tag{3.16}$$

Indeed, notice that

$$\nabla_{x_i} \left(\sum_{j=1}^{N} a_{ij} \|x_i - x_j\|^2 \right) = 2\Pi_{T_{x_i}M} \sum_{j=1}^{N} a_{ij}(x_j - x_i) = 2\dot{x}_i.$$

Then we compute

$$
\begin{aligned}
\frac{d}{dt}F(t) &= \sum_{k=1}^{N}\langle\nabla_{x_k}\frac{1}{2}\sum_{i,j=1}^{N}a_{ij}(\|x_i-x_j\|^2),\dot{x}_k\rangle \\
&= \sum_{k=1}^{N}\langle\nabla_{x_k}\left[\frac{1}{2}\sum_{i=1}^{N}a_{ik}(\|x_i-x_k\|^2)+\frac{1}{2}\sum_{j=1}^{N}a_{kj}(\|x_k-x_j\|^2)\right],\dot{x}_k\rangle \\
&= \sum_{k=1}^{N}\langle2\nabla_{x_k}\frac{1}{2}\sum_{i=1}^{N}a_{ik}(\|x_i-x_k\|^2),\dot{x}_k\rangle = \sum_{k=1}^{N}\langle2\Pi_{T_{x_k}M}\sum_{j=1}^{N}a_{kj}(x_j-x_k),\dot{x}_k\rangle \\
&= 2\sum_{k=1}^{N}\|\dot{x}_k\|^2 = 4E_P(t).
\end{aligned}
$$

$$(3.17)$$

where the third equality uses the property: $a_{ij}=a_{ji}$ for all i,j.

Since $E_P(t)\geq 0$, $F(t)$ is a non-decreasing function. Moreover $F(t)$ and $\frac{d^2}{dt^2}F(t)$ are bounded, since M is a compact manifold. Hence $F(t)$ converges as $t\to\infty$.

Now, $\int_0^{+\infty}E_P<\infty$ and there exists $c>0$ such that $\sup_t\frac{d}{dt}E_P(t)<c$. By contradiction, assume $\limsup E_P=\alpha>0$, then there exists a sequence (t_n) such that $t_{n+1}>t_n+\frac{\alpha}{2c}$ and $E_P(t_n)\geq\frac{\alpha}{2}$. Then $\int_0^{\infty}E_P>\sum_n\frac{\alpha}{c}\cdot\frac{\alpha}{2}\cdot\frac{1}{2}=+\infty$. Hence $\limsup_t E_P(t)=0$.

This shows that $\frac{d}{dt}F(t)\to 0$ when $t\to\infty$, which implies that $\lim_{t\to\infty}E_P(t)=0$.

Remark 1 *Propositions 7 and 11 are generalizations of results proven for the case $M=\mathbb{S}^2$ in [34].*

Remark 2 *Proposition 11 assumes that $\Psi\equiv\mathrm{Id}$ which creates a discontinuity for Approach B (see Table 3.1). A result is shown for the more restricted case of $M=\mathbb{S}^2$ and $\Psi(\cdot)\equiv\sin(\cdot)$, (Corollary 1).*

Definition 9 *Let x solve the differential equation (3.1). A dancing equilibrium is a configuration in which for all pairs of agents (i,j), the distance $d_P(x_i,x_j)$ (in Approach*

A) or $d_G(x_i, x_j)$ (in Approach B) is constant. In the context of a system of oscillators, this equilibrium is also known as a phase-locked state or entrainment state [51].

Remark 3 *This definition is a generalization of the concept of dancing equilibrium described in [34].*

Remark 4 *It follows immediately from definition 9 that the kinetic energy of a system in dancing equilibrium is constant.*

3.3 Analysis and simulations on $\mathbb{S}^1$

3.3.1 Models

We study both approaches A and B in the case $M = \mathbb{S}^1$, i.e. for the one-dimensional sphere embedded in $\mathbb{R}^2$. Let $(\theta_i)_{i \in \{1,...,N\}} \in [0, 2\pi]^N$ such that for all $i \in \{1, ..., N\}$, $x_i = (\cos \theta_i, \sin \theta_i)^T$.

Approach A The projection onto an agent's tangent space can be rewritten as:

$$
\begin{aligned}
\Pi_{T_{x_i}} \sum_{j=1}^N a_{ij}(x_j - x_i) &= \sum_{j=1}^N a_{ij} \left\langle \begin{pmatrix} \cos \theta_j \\ \sin \theta_j \end{pmatrix} - \begin{pmatrix} \cos \theta_i \\ \sin \theta_i \end{pmatrix}, \begin{pmatrix} -\sin \theta_i \\ \cos \theta_i \end{pmatrix} \right\rangle \begin{pmatrix} -\sin \theta_i \\ \cos \theta_i \end{pmatrix} \\
&= \sum_{j=1}^N a_{ij}(-\sin \theta_i \cos \theta_j + \sin \theta_j \cos \theta_i) \begin{pmatrix} -\sin \theta_i \\ \cos \theta_i \end{pmatrix} \\
&= \sum_{j=1}^N a_{ij} \sin(\theta_j - \theta_i) \begin{pmatrix} -\sin \theta_i \\ \cos \theta_i \end{pmatrix}.
\end{aligned}
$$

$$(3.18)$$

So System (3.1)-(3.2)-(3.3) becomes:

for all $i \in \{1, \ldots, N\}$,

$$\dot{\theta}_i \begin{pmatrix} -\sin\theta_i \\ \cos\theta_i \end{pmatrix} = \sum_{j=1}^{N} a_{ij} \Psi(|\sin(\theta_j - \theta_i)|) \operatorname{sgn}(\sin(\theta_j - \theta_i)) \begin{pmatrix} -\sin\theta_i \\ \cos\theta_i \end{pmatrix} \qquad (3.19)$$

where $\operatorname{sgn}(\cdot)$ is the sign function defined by:

$$\text{for all } x \in \mathbb{R}, \quad \operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0. \end{cases} \qquad (3.20)$$

We can then specify:

$$\text{for all } (i,j) \in \{1, \ldots, N\}^2, \quad d_P(x_i, x_j) = |\sin(\theta_j - \theta_i)|, \quad \nu_{ij}^P = \operatorname{sgn}(\sin(\theta_j - \theta_i)). \qquad (3.21)$$

This gives the system of scalar equations:

$$\text{for all } i \in \{1, \ldots, N\}, \quad \dot{\theta}_i = \sum_{j=1}^{N} a_{ij} \Psi(|\sin(\theta_j - \theta_i)|) \operatorname{sgn}(\sin(\theta_j - \theta_i)). \qquad (3.22)$$

In particular, in the case $\Psi \equiv \operatorname{Id}$, the system becomes the Kuramoto model [40].

$$\text{for all } i \in \{1, \ldots, N\}, \quad \dot{\theta}_i = \sum_{j=1}^{N} a_{ij} \sin(\theta_j - \theta_i). \qquad (3.23)$$

Approach B For $M = \mathbb{S}^1$, the geodesics distance d_G and the vector ν_{ij}^G are given by:

$$d_G(x_i, x_j) = \arccos(\cos(\theta_j - \theta_i)) \quad , \quad \nu_{ij}^G = \operatorname{sgn}(\sin(\theta_j - \theta_i)). \qquad (3.24)$$

System (3.1)-(3.6)-(3.7) is written:

$$\text{for all } i \in \{1, \ldots, N\}, \quad \dot{\theta}_i = \sum_{j=1}^{N} a_{ij} \Psi(\arccos(\cos(\theta_j - \theta_i))) \operatorname{sgn}(\sin(\theta_j - \theta_i)). \quad (3.25)$$

In order for the system to be well defined, the interaction function Ψ must satisfy the conditions given in Table 3.1. Notice that the injectivity radius is constant over $\mathbb{S}^1$, with $\epsilon = \pi$. Possible choices involve choosing Ψ from a family of function defined as follows:

$$\Psi^a(d) = \begin{cases} \frac{1}{a}d & \text{for } d \leq a \\[2mm] \frac{d-\pi}{a-\pi} & \text{for } d > a \end{cases} \quad (3.26)$$

where $a \in (0, \pi)$ (see Figure 3.8).

Another possible choice is: $\Psi : x \mapsto \sin(x)$. Notice that for the specific choices $\Psi = \operatorname{Id}$ for Approach A and $\Psi : x \mapsto \sin(x)$ for Approach B, the two approaches A and B are equivalent.

3.3.2 Analysis

We first examine the different equilibria for both approaches.

Theorem 1 *Consider Approach A, System (3.22). Let $N \in \mathbb{N}$ be even. Suppose that for all $i \in \{1, \ldots, N\}$ for all $j \in \{1, \ldots \frac{N}{2}\}$, $a_{ij} = a_{i(j+\frac{N}{2})}$. Then any configuration that is centrally symmetric, i.e.*

$$\text{for all } j \in \{1, \ldots, \frac{N}{2}\}, \ \theta_{j+\frac{N}{2}} = \theta_j + \pi$$

is an equilibrium.

Proof 11 *Using the hypotheses from Theorem 1, we can easily compute:*

$$\dot{\theta}_i = \sum_{j=1}^{N} a_{ij} \Psi(\|\sin(\theta_j - \theta_i)\|) \operatorname{sgn}(\sin(\theta_j - \theta_i))$$

$$= \sum_{j=1}^{N/2} [a_{ij} \Psi(\|\sin(\theta_j - \theta_i)\|) \operatorname{sgn}(\sin(\theta_j - \theta_i)) +$$

$$a_{i(j+\frac{N}{2})} \Psi(\|\sin(\theta_{j+\frac{N}{2}} - \theta_i)\|) \operatorname{sgn}(\sin(\theta_{j+\frac{N}{2}} - \theta_i))]$$

$$= \sum_{j=1}^{N/2} [a_{ij} \Psi(\|\sin(\theta_j - \theta_i)\|) \operatorname{sgn}(\sin(\theta_j - \theta_i)) +$$

$$a_{ij} \Psi(\|\sin(\theta_j + \pi - \theta_i)\|) \operatorname{sgn}(\sin(\theta_j + \pi - \theta_i))] = 0.$$

Interestingly, Theorem 1 is not applicable to Approach B. We illustrate the different behaviors of the two systems by studying the specific example of four agents initially in a rectangular configuration. According to Theorem 1, this configuration is an equilibrium for Approach A, independently of the choice of interaction function Ψ. However, one can easily prove that in the geodesics-based Approach B, with $N = 4$ and the choice $\Psi := \Psi^a$ with $a = \frac{3\pi}{4}$, the only equilibrium for which all agents have pairwise distinct positions is obtained by a regular polygon, i.e. all agents are evenly spaced out on the circle. This is illustrated by numerical simulations shown in Figure 3.4.

This highlights the fundamentally different behaviors of the systems (3.1)-(3.2)-(3.3) and (3.1)-(3.6)-(3.7) in the case $M = \mathbb{S}^1$.

In both approaches A and B, conditions on the interaction matrix A can be found such that the system forms a dancing equilibrium (see Definiton 9).

Theorem 2 *Consider the dynamics on $\mathbb{S}^1$ given by:*

$$\text{for all } i \in \{1, \ldots, N\}, \quad \dot{\theta}_i = \sum_{j=1}^{N} a_{ij} \Psi(d(x_i, x_j)) \nu_{ij} \tag{3.27}$$
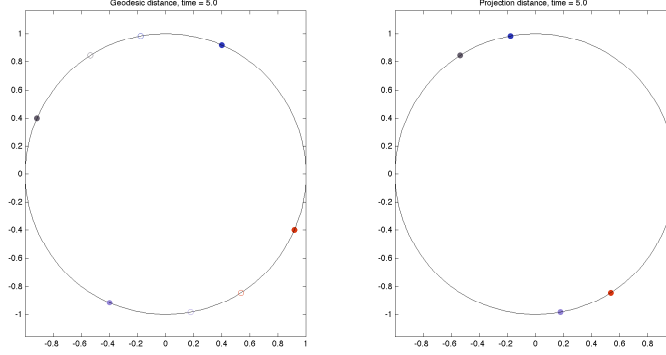
Figure 3.4: Initial (empty circles) and final positions (filled circles) of 4 agents initially on the vertices of a rectangle with Approach B (left) and Approach A (right), with $A = \mathbb{1}$, $\Psi \equiv \mathrm{Id}$ (Approach A) and $\Psi = \Psi^{3\pi/4}$(Approach B) (see equation (3.26)). Notice that with Approach A, initial and final positions are identical since any rectangle configuration is an equilibrium. However, with Approach B, the system reaches a square configuration, the only possible equilibrium with pairwise distinct positions.

where $d(\cdot, \cdot)$ and ν are given either by Approach A (3.21) or Approach B (3.24). Let $C \in \mathbb{R}$ and suppose that for all $i \in \{1, \dots, N\}$,

$$a_{ij} = \begin{cases} \frac{C}{\Psi(d(x_i(0), x_j(0)))} \nu_{ij} & \text{if } \Psi(d(x_i(0), x_j(0))) \neq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{3.28}$$

Then the system is in a dancing equilibrium.

Proof 12 *If the interaction matrix satisfies (3.28), then at $t = 0$,*

$$\text{for all } i \in \{1, \dots, N\}, \quad \dot{\theta}_i(0) = \sum_{j=1}^{N} C = CN$$

so for all $(i, j) \in \{1, \dots, N\}^2$, $\dot{\theta}_i(0) - \dot{\theta}_j(0) = 0$. Then $d(x_i, x_j)$ does not change in time, and (3.28) holds for all time.

Numerical simulations show the evolution of the system (3.27) with condition (3.28) for the projection or the geodesic distance, see Figures 3.5 and 3.6.
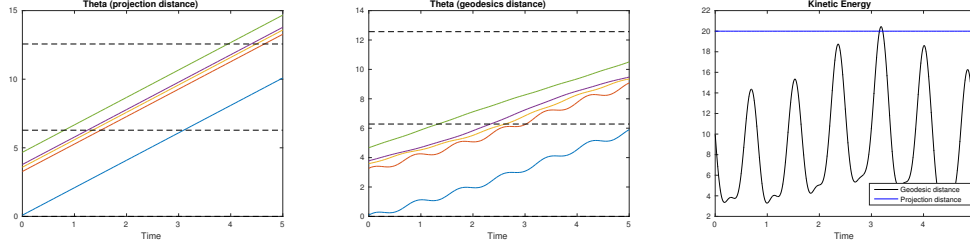
Figure 3.5: Evolution of the system (3.27) with Approach A (left) Approach B (center) when the interaction matrix satisfies condition (3.28) for the projection distance. Right: Kinetic energy.
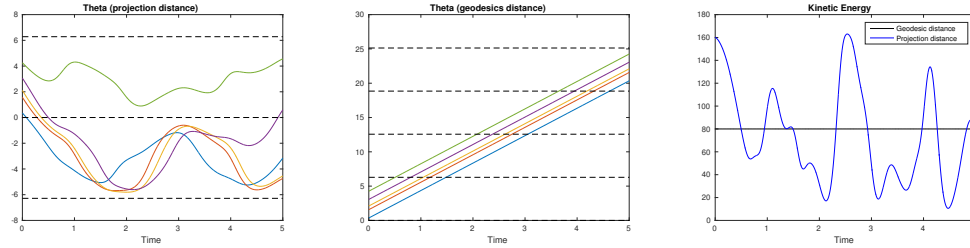

Figure 3.6: Evolution of the system (3.27) with Approach A (left) Approach B (center) when the interaction matrix satisfies condition (3.28) for the geodesic distance. Right: Kinetic energy.

3.4   Analysis and simulations on $\mathbb{S}^2$

3.4.1   Models

We study both approaches A and B for $M = \mathbb{S}^2$, i.e. for a two dimensional sphere embedded in $\mathbb{R}^3$. We use spherical coordinates: let $(\theta_i)_{i \in \{1,...,N\}} \in [0, 2\pi]^N$, and $(\phi_i)_{i \in \{1,...,N\}} \in [0, \pi]^N$ such that for all $i \in \{1, ..., N\}$,

$$x_i = (\cos\theta\sin\phi, \sin\theta\sin\phi, \cos\phi)^T.$$

Choice of influence function   We choose an influence function $\Psi(d)$ between two agents x_i and x_j so that the right-hand side of the system is continuous, the discontinuities are shown in Table 3.1. For Approach B, the only point in $\mathcal{CL}(x_i)$ for a given x_i is the antipodal point (this is an end point of a diameter for which x_i is the other end point.) As in the case of $\mathbb{S}^1$, for Approach B, we choose a function Ψ from a

family of functions of the form Ψ^a , see equation (3.26) (choices of Ψ are shown in Figure 3.8).

Approach A   On $\mathbb{S}^2$, the derivative for system (3.1)-(3.2)-(3.3) with $\Psi \equiv \text{Id}$ reduces to the sum of all projections onto the tangent space of agent x_i, weighted by the corresponding interaction term a_{ij}. This is rewritten as:

$$\Pi_{T_{x_i}} \sum_{j=1}^{N} a_{ij}(x_j - x_i) = \Pi_{T_{x_i}} \sum_{j=1}^{N} a_{ij}(x_j) = \sum_{j=1}^{N} a_{ij}(x_j - \langle x_j, x_i \rangle x_i) =$$

$$\sum_{j=1}^{N} a_{ij} \left(\begin{pmatrix} \cos\theta_j \sin\phi_j \\ \sin\theta_j \sin\phi_j \\ \cos\phi_j \end{pmatrix} - \left\langle \begin{pmatrix} \cos\theta_j \sin\phi_j \\ \sin\theta_j \sin\phi_j \\ \cos\phi_j \end{pmatrix} , \begin{pmatrix} \cos\theta_i \sin\phi_i \\ \sin\theta_i \sin\phi_i \\ \cos\phi_i \end{pmatrix} \right\rangle \begin{pmatrix} \cos\theta_i \sin\phi_i \\ \sin\theta_i \sin\phi_i \\ \cos\phi_i \end{pmatrix} \right).$$

Approach B   The geodesic distance $d_G(x_i, x_j)$ from (3.6) between two points x_i, and x_j on $\mathbb{S}^2$ is given by:

$$d_G(x_i, x_j) = 2 \arcsin\left(\frac{\|x_i - x_j\|}{2} \right),$$

and the direction toward x_j from x_i is

$$\nu_{ij} = \frac{x_j - \langle x_j, x_i \rangle x_i}{\|x_j - \langle x_j, x_i \rangle x_i\|},$$

where $\| \cdot \|$ is the standard norm in $\mathbb{R}^3$.

Noticing that for $\mathbb{S}^2$, Approach A with $\Psi \equiv \text{Id}$ is equivalent to Approach B with $\Psi \equiv \sin$, we extend the results of Proposition 11.

Corollary 1 *Consider the dynamics given by geodesic distance (Approach B) on $\mathbb{S}^2$, system (3.1)-(3.6)-(3.7), and let $\Psi \equiv \sin$. If the interaction matrix $A = (a_{ij})_{i,j \in \{1,...,N\}^2}$ is symmetric, then*

$$\lim_{t \to \infty} E_G(t) = 0. \tag{3.29}$$

Proof 13 *System* (3.1)-(3.6)-(3.7), *with* $\Psi(\cdot) \equiv \sin(\cdot)$ *reads:*

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} \sin(d_G(x_i, x_j)) \nu_{ij}. \tag{3.30}$$

Considering the embedding of the system in $\mathbb{R}^3$, *we notice that for all* $i, j \in \{1, \ldots, N\}$,

$$a_{ij} \sin(d_G(x_i, x_j)) \nu_{ij} = a_{ij} \Pi_{T_{x_i} M}(x_j - x_i), \tag{3.31}$$

thus the system is (3.4), *and by proposition 11,* $\lim_{t \to \infty} E_P(t) = 0$.

Finally,

$$\lim_{t \to \infty} E_P(t) = 0 \implies \lim_{t \to \infty} \dot{x}_i = 0 \quad \textit{for all } i \implies \lim_{t \to \infty} E_G(t) = 0. \tag{3.32}$$

3.4.2 Simulations

We use a fourth order Runge-Kutta scheme to approximate the trajectories. The derivative is calculated as a vector in $\mathbb{R}^3$, and then we express this vector in spherical coordinates, $\dot{\theta} t_\theta$ and $\dot{\phi} t_\phi$ where t_θ and t_ϕ are the unit vectors in the direction of the azimuth angle(θ) and the polar angle(ϕ) respectively for the ith agent. $\{t_\theta, t_\phi\}$ form an orthonormal basis for the tangent space of x_i. Using the angular derivatives avoids having to calculate the agent's trajectory in $\mathbb{R}^3$ and then project onto the sphere for every iteration which would cause significant numerical errors.

For an agent $x_i = (\theta_i, \phi_i)$ we can write t_{θ_i} and t_{ϕ_i} as

$$t_\theta = \begin{pmatrix} -\sin\theta \\ \cos\theta \\ 0 \end{pmatrix}, \qquad t_\phi = \begin{pmatrix} \cos\theta\cos\phi \\ \sin\theta\cos\phi \\ -\sin\phi \end{pmatrix}.$$

We express the derivative of an agent as

$$\dot{x}_i = \frac{\partial x_i}{\partial \theta_i}\dot{\theta} + \frac{\partial x_i}{\partial \phi_i}\dot{\phi}. \tag{3.33}$$

By direct computation, we get:

$$\frac{\partial x_i}{\partial \theta_i} = \begin{pmatrix} -\sin\theta_i \sin\phi_i \\ \cos\theta_i \sin\phi_i \\ 0 \end{pmatrix}, \qquad \frac{\partial x_i}{\partial \phi_i} = \begin{pmatrix} \cos\theta_i \cos\phi_i \\ \sin\theta_i \cos\phi_i \\ -\sin\phi_i \end{pmatrix}.$$

We can also express the derivative of x_i as the projection of the derivative in $\mathbb{R}^3$ with t_{θ_i} and t_{ϕ_i}

$$\dot{x}_i = \langle \dot{x}_i, t_{\theta_i} \rangle t_{\theta_i} + \langle \dot{x}_i, t_{\phi_i} \rangle t_{\phi_i}. \tag{3.34}$$

It follows from (3.33) and (3.34) that

$$\dot{\theta}_i = \frac{1}{\sin\phi_i}\langle \dot{x}_i, t_{\theta_i} \rangle t_{\phi_i} \quad \text{and} \quad \dot{\phi}_i = \langle \dot{x}_i, t_{\phi_i} \rangle t_{\phi_i}. \tag{3.35}$$

Singularities: In (3.35), the factor $\frac{1}{\sin\phi}$ causes a singularity around $\phi = k\pi$ for a non-negative integer k. To avoid this practical problem, before each iteration of the RK4 scheme, we identify critical agents that have a polar angle close to 0 or π ($\phi = k\pi$), for non-negative integer k; for these critical agents, we rotate all agents $\frac{\pi}{2}$ around the x-axis to calculate the derivative. This is a concern for both Approach A and Approach B.

An additional concern is singularities for agents forming consensus. In equations (3.3) and (3.7), ν_{ij} is normalized, and when agent i and agent j are very close together, dividing by d_P and d_G causes a singularity. We avoid this problem in our simulations by defining a minimum distance $d_{\min}$ between agents for the sake of the normalization

term in the denominator. When $d(x_i, x_j) < d_{\min}$ then

$$\nu_{ij} = \frac{x_j - \langle x_j, x_i \rangle x_i}{d_{\min}}$$

3.4.3 Examples

We ran simulations using different choices of Ψ to see how this choice can impact the system. We show two simulations, the first uses Approach A (Figure 3.7); the second uses Approach B (Figure 3.9). In both pictures the same interaction matrix is used (given in the appendix), and we see that our choice of Ψ (Figure 3.8) may dramatically change the system behavior. The second example shows the effect of curvature on the system (Figure 3.10) for comparison to $\mathbb{T}^2$ and $\mathbb{R}^2$ (Figure 3.11). Another example in the appendix shows unexpected behavior using Approach A.

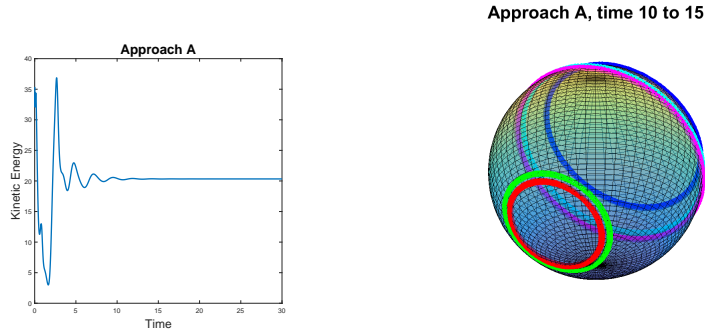Figure 3.7: A dancing equilibrium for Approach A. The energy becomes constant in time after initial fluctuations.

Example 3.4.1 *Five agents with a general interaction matrix A and Ψ as defined in (3.26) with $a \in \{\frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$. The behavior of the system can change dramatically from our choice of Ψ (Figure 3.8), and dancing equilibria may arise from A with certain*

configurations, (Figure 3.9).
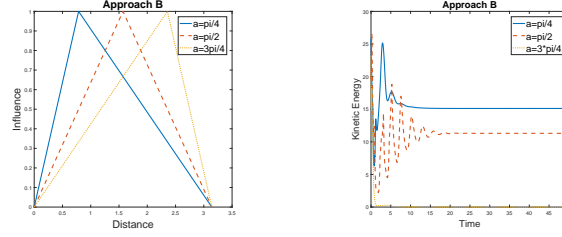
Figure 3.8: The left side shows candidates for the choice of function Ψ. The right side shows how choice of function determines the energy of the system, for the case of $a = \frac{3\pi}{4}$ the system forms an antipodal equilibrium.
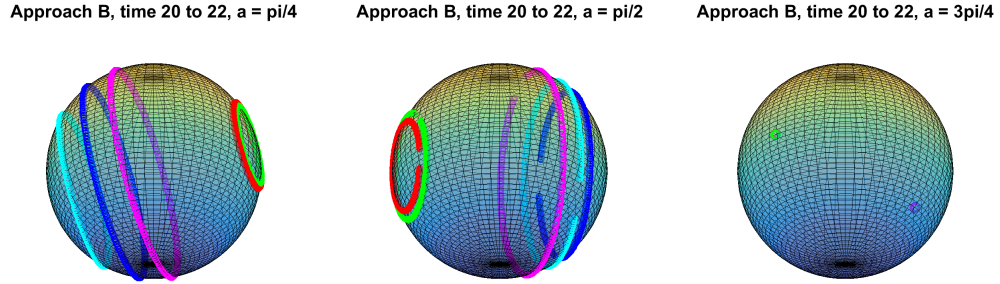
Figure 3.9: A comparison of the effect of the choice of influence function for Approach B. For $a = \frac{3\pi}{4}$ an antipodal equilibrium occurs (see Definition 7).

Example 3.4.2 *To assess the influence of the curvature of $\mathbb{S}^2$ on the dynamics, observe a simple case involving 3 agents evolving according to the interaction matrix:*

$$A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix} \tag{3.36}$$

In Section 3.6.2, we prove that those dynamics in $\mathbb{R}^2$ lead to periodic trajectories on a

single orbit shared by all three agents, the orbit's parameters being fully determined by the initial conditions (see Theorem 4). However, the same dynamics on the sphere do not give rise to periodic trajectories. In sections 3.5.3, we also discuss the dynamics with this interactions matrix on $\mathbb{T}^2$, to assess the effect of curvature of the manifold.
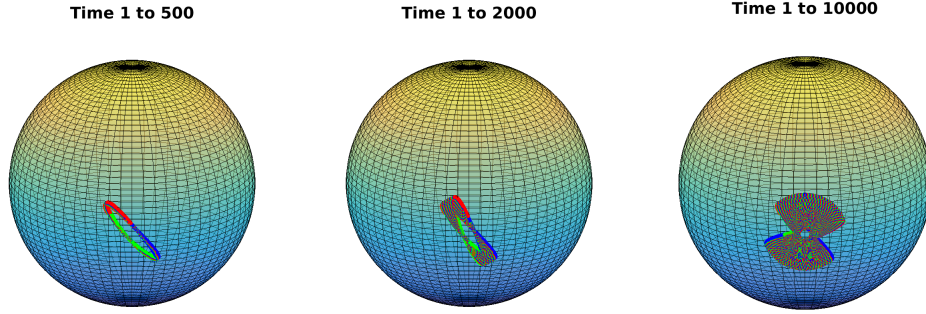
Figure 3.10: Dynamics with Approach A on $\mathbb{S}^2$, using the interactions matrix (3.36). If the agents' initial positions are close enough to each other, the agents with will form trajectories that remain in a neighborhood of their initial position.

3.5 Analysis and simulations on $\mathbb{T}^2$

We now study how the general dynamics given by equation (3.1) apply to the specific case of the torus $\mathbb{T}^2 \subset \mathbb{R}^3$. Let (e_x, e_y, e_z) denote the Euclidean basis of $\mathbb{R}^3$. Let $(R, r) \in (\mathbb{R}^+)^2$, with $R > r$. We define the manifold $\mathbb{T}^2$ as the torus obtained by rotating the circle $(x - R)^2 + z^2 = r^2$ around the z-axis. Hence $\mathbb{T}^2$ is defined by the equation $(R - \sqrt{x^2 + y^2})^2 + z^2 = r^2$. The parametric equations for such a torus are:

$$\begin{cases} x = (R + r\cos\theta)\cos\phi \\ y = (R + r\cos\theta)\sin\phi \qquad \text{for } (\phi, \theta) \in [0, 2\pi)^2. \\ z = r\sin\theta \end{cases}$$

The angles ϕ and θ are respectively referred to as the toroidal and poloidal angles. A set of points with the same toroidal angle is called a meridian.

3.5.1   Model

We first investigate the behavior of system (3.1) with Approach B (using the geodesic distance) in the case of $\mathbb{T}^2$. Unlike in the cases of $\mathbb{S}^1$ and $\mathbb{S}^2$ presented in the sections 3.3 and 3.4, there exists no simple expression for the geodesic distance between two points on the torus. In 1903, Bliss studied and classified the different kinds of geodesic lines on the standard torus [52], using elliptic functions. Gravesen et al. determined the structure of the cut loci of a torus of revolution [48].

Several challenges arise when defining Approach B on $\mathbb{T}^2$. Firstly, computing the Riemannian distance between two points is highly non-trivial. One could consider approximating it numerically, but in the numerical discretization of equations (3.1)-(3.6)-(3.7), $N(N-1)/2$ geodesics would have to be computed per time-step. That would require tremendous computing power.

Secondly, assuming that one is able to efficiently compute the geodesics on $\mathbb{T}^2$, one must take into account the cut-loci of each point to ensure that the dynamics (3.1)-(3.6)-(3.7) are well-defined. A method to guarantee well-defined dynamics would be to use a bounded confidence model [2], where the neighborhood of influence for an agent x_i at point p is of smaller radius than the closest element in the cut locus of p. See section 3.2 for conditions on Ψ to make the right hand side of equation (3.1) continuous.

For simplicity, we thus focus on Approach A, where the dynamics are a function of the projection of each vector $x_j - x_i$ onto the tangent space at x_i. We will show that some restrictions still apply to the interaction function Ψ, but they are less restrictive and more easily determined than in Approach B.

Equations (3.1)-(3.2) reads:

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij} \Psi(\|\Pi_{T_{x_i} \mathbb{T}^2}(x_j - x_i)\|)\nu_{ij}, \quad i \in \{1, \ldots, N\}. \tag{3.37}$$

The vector ν_{ij} depends on the influence that x_j has over x_i. It is zero if $\Pi_{T_{x_i} \mathbb{T}^2}(x_j - x_i) = 0$, and it is a unit vector otherwise. Let $\mathcal{N}_i$ be the set of points that have no influence on x_i (see Table 3.1). Then, given $i, j \in \{1, \ldots, N\}$, ν_{ij} has the following expression:

$$\nu_{ij} = \begin{cases} \dfrac{\Pi_{T_{x_i} \mathbb{T}^2}(x_j - x_i)}{\|\Pi_{T_{x_i} M}(x_j - x_i)\|} & \text{if } x_j \notin \mathcal{N}_i \\ 0 & \text{if } x_j \in \mathcal{N}_i. \end{cases} \tag{3.38}$$

Let $x_i \in \mathbb{T}^2$. We start by determining the set $\mathcal{N}_i$. For all i, we define the vectors $u_{\phi_i} = \cos\phi_i e_x + \sin\phi_i e_y$ and $u_{\theta_i} = \cos\theta_i u_{\phi_i} + \sin\theta_i e_z$, so that each agent's position vector reads: $x_i = Ru_{\phi_i} + ru_{\theta_i}$. With these notations, u_{θ_i} is the normal to the tangent space at the point x_i. A basis for the tangent space at a point $x_i(\phi_i, \theta_i)$ is given by the two tangent vectors $t_{\phi_i} = (-\sin\phi_i, \cos\phi_i, 0)$ and $t_{\theta_i} = (-\sin\theta_i \cos\phi_i, -\sin\theta_i \sin\phi_i, \cos\theta_i)$. Notice that $\langle x_i, t_{\phi_i} \rangle = 0$. Hence the condition $\Pi_{T_{x_i} \mathbb{T}^2}(x_j - x_i) = 0$ reads:

$$\begin{cases} \langle x_j, t_{\phi_i} \rangle = 0 \\ \langle x_j - x_i, t_{\theta_i} \rangle = 0. \end{cases}$$

After computations, we get:

$$\langle x_j, t_{\phi_i} \rangle = 0 \iff \sin(\phi_j - \phi_i) = 0 \iff \phi_j = \phi_i + k\pi, k \in \mathbb{Z}.$$

If $\phi_j = \phi_i$, the second condition becomes:

$$\langle x_j - x_i, t_{\theta_i} \rangle = 0 \iff \sin(\theta_j - \theta_i) = 0 \iff \theta_j = \theta_i + k\pi, k \in \mathbb{Z}.$$

If $\phi_j = \phi_i \pm \pi$, the second condition becomes:

$$\sin(\theta_i + \theta_j) = -\frac{2R}{r}\sin\theta_i.$$

Notice that this last equation only has a solution if $|\sin\theta_i| \leq \frac{r}{2R}$. The set of positions that have no influence on x_i thus comprises up to four points on the torus, depending on the values of r, R and $\sin\theta_i$. We then have: $\mathcal{N}_i = \{(\phi_i, \theta_i), (\phi_i, -\theta_i), (-\phi_i, -\theta_i - \mathrm{sgn}(\sin\theta_i)\arcsin(|\frac{2R}{r}\sin\theta_i|)), (-\phi_i, \pi - \theta_i + \mathrm{sgn}(\sin\theta_i)\arcsin(|\frac{2R}{r}\sin\theta_i|))\}$. To ensure the continuity of the right-hand side of equation (3.37), one must impose the conditions of table 3.1.

We now go back to equation (3.37). We study the specific case where $\Psi \equiv \mathrm{Id}$, which indeed satisfies (3.1). Then the system becomes:

$$\dot{x}_i = \Pi_{T_{x_i}\mathbb{T}^2}\left(\sum_{j=1}^{N} a_{ij}(x_j - x_i)\right). \tag{3.39}$$

Hence the velocity reads:

$$\dot{x}_i = \sum_{j=1}^{N} a_{ij}(x_j - x_i) - \langle\sum_{j=1}^{N} a_{ij}(x_j - x_i), u_{\theta_i}\rangle u_{\theta_i}$$

$$= \alpha_i - \langle\alpha_i, u_{\theta_i}\rangle u_{\theta_i} - \left(\sum_{j=1}^{N} a_{ij}\right)\langle x_i, t_{\theta_i}\rangle t_{\theta_i}$$

where $\alpha_i := \sum_{j=1}^{N} a_{ij}x_j$ is the sum of the influences of all agents on agent i. Notice that with the same notation, the system does not reduce to the simple form $\dot{x}_i = \alpha_i - \langle\alpha_i, x_i\rangle x_i$ for the same dynamics on the sphere (see [34]). This is due to the fact that on the torus, the position vector x_i does not define the normal to the tangent space at x_i, unlike in the cases of $\mathbb{S}^1$ and $\mathbb{S}^2$.

The velocity of each agent is given by:

$$\dot{x}_i = \begin{pmatrix} -\dot{\phi}_i \sin\phi_i(R + r\cos\theta_i) - r\dot{\theta}_i \sin\theta_i \cos\phi_i \\ \dot{\phi}_i \cos\phi_i(R + r\cos\theta_i) - r\dot{\theta}_i \sin\theta_i \sin\phi_i \\ r\dot{\theta}_i \cos\theta_i \end{pmatrix} = \dot{\phi}_i(R + r\cos\theta_i)t_{\phi_i} + r\dot{\theta}_i t_{\theta_i}. \quad (3.40)$$

From (3.39) and (3.40) we get the angular velocities:

$$\begin{cases} \dot{\phi}_i = \frac{1}{(R + r\cos\theta_i)}\langle \sum_{j=1}^{N} a_{ij}(x_j - x_i), t_{\phi_i}\rangle \\ \dot{\theta}_i = \frac{1}{r}\langle \sum_{j=1}^{N} a_{ij}(x_j - x_i), t_{\theta_i}\rangle. \end{cases} \quad (3.41)$$

Notice that unlike in the case of $\mathbb{S}^2$, see equation (3.35), here the derivatives $\dot{\phi}_i$ and $\dot{\theta}_i$ are not singular. This makes numerical simulations straightforward, not requiring the approximations described in Section 3.4.2.

3.5.2   Properties

We now analyze the dynamics (3.1)-(3.2)-(3.3) on $\mathbb{T}^2$. We identify families of initial conditions that trivialize the dynamics.

Proposition 9 *Consider the dynamics (3.1)-(3.2)-(3.3) on $M = \mathbb{T}^2$. Let $P_z :=$ $\{(x, y, z)$ $\in \mathbb{R}^3 \mid z = 0\}$. Let $x_i(t)$ be the position of the ithe agent at time t. If for all $i \in \{1, \ldots, N\}$, $x_i(0) \in \mathbb{T}^2 \cap P_z$, then for all $t \geq 0$, for all $i \in \{1, \ldots, N\}$, $x_i(t) \in \mathbb{T}^2 \cap P_z$.*

Proof 14 *Suppose that for all $i \in \{1, \ldots, N\}$, $x_i(0) \in \mathbb{T}^2 \cap P_z$. Then for all $i \in$*

$\{1, \ldots, N\}$, $\theta_i(0) = 0$ *or* $\theta_i(0) = \pi$. *Hence, for all* $i, j \in \{1, \ldots, N\}$,

$$t_{\theta_i}(0) = \begin{pmatrix} 0 \\ 0 \\ \pm\pi \end{pmatrix} \text{ and } x_j(0) - x_i(0) = \begin{pmatrix} (R + r\cos\theta_j)\cos\phi_j - (R + r\cos\theta_i)\cos\phi_i \\ (R + r\cos\theta_j)\sin\phi_j - (R + r\cos\theta_i)\sin\phi_i \\ 0 \end{pmatrix}$$

From equation (3.41) we get: for all $i \in \{1, \ldots, N\}$, $\dot{\theta}_i = 0$. *By uniqueness of solution, for all* $i \in \{1, \ldots, N\}$, $\theta_i(t) = \theta_i(0)$. *All the initial velocities belong to the plane* P_z. *Hence all agents remain on* P_z *at all time.*

Remark 5 *As a consequence of Proposition 9, if all agents are initially in* $\mathbb{T}^2 \cap P_z$, *all agents initially on the bigger circle* $\theta = 0$ *remain on the major circle at all time and all agents on the minor circle* $\theta = \pi$ *remain on the minor circle at all time. In particular, if all agents are initially all on the same circle (i.e. for all* $i \in \{1, \ldots, N\}$, $\theta_i = 0$ *or for all* $i \in \{1, \ldots, N\}$, $\theta_i = \pi$), *then the torus dynamics simplify to the dynamics on* $\mathbb{S}^1$ *given by (3.22) or (3.23).*

Proposition 10 *Consider the dynamics (3.1)-(3.2)-(3.3) on* $M = \mathbb{T}^2$. *Let* $\tilde{\phi} \in [0, 2\pi]$ *and let* $P_{\tilde{\phi}} := \{(x, y, z) \in \mathbb{R}^3 \mid y = \tan(\tilde{\phi})x\}$. *If for all* $i \in \{1, \ldots, N\}$, $x_i(0) \in \mathbb{T}^2 \cap P_{\tilde{\phi}}$, *then for all* $t \geq 0$, *for all* $i \in \{1, \ldots, N\}$, $x_i(t) \in \mathbb{T}^2 \cap P_{\tilde{\phi}}$.

Proof 15 *Suppose without loss of generality that* $\bar{\phi} = 0$. *Similarly to the proof for Proposition 9, we can show that for all* $i \in \{1, \ldots, N\}$, $\dot{\phi}_i(0) = 0$. *By uniqueness of solution, for all* $i \in \{1, \ldots, N\}$, $\phi_i(t) = \phi_i(0)$. *Hence all agents remain in* $P_{\tilde{\phi}}$ *at all time.*

Remark 6 *As a consequence of Proposition 10, if all agents are initially in* $\mathbb{T}^2 \cap P_{\tilde{\phi}}$, *all agents initially on the circle* $\phi = \bar{\phi}$ *remain on that circle at all time and all agents on the circle* $\phi = -\bar{\phi}$ *remain on that circle at all time. In particular, if all agents are initially all on the same circle (i.e. for all* $i \in \{1, \ldots, N\}$, $\phi_i = \bar{\phi}$ *or*

for all $i \in \{1, \ldots, N\}$, $\phi_i = -\bar{\phi}$), then the torus dynamics simplify to the dynamics on $\mathbb{S}^1$ given by (3.22) or (3.23).

3.5.3   Simulations

To assess the influence of the curvature of the manifold on the dynamics, we compare a simple case involving 3 agents evolving according to the interaction matrix given in equation (3.36). As in the case of $\mathbb{S}^2$, the dynamics on the torus do not give rise to periodic trajectories (as opposed to the dynamics in $\mathbb{R}^2$, see Theorem 4). Instead, since $\mathbb{T}^2$ can locally be identified with $\mathbb{R}^2$, if the initial mutual distances are small enough, the dynamics resemble those in $\mathbb{R}^2$. More specifically, the trajectories are quasi-periodic with a gradual shift of the center of mass (see Figure 3.11). However, if the initial distances between agents are large, the geometry and curvature of the torus changes radically the behavior of the system.
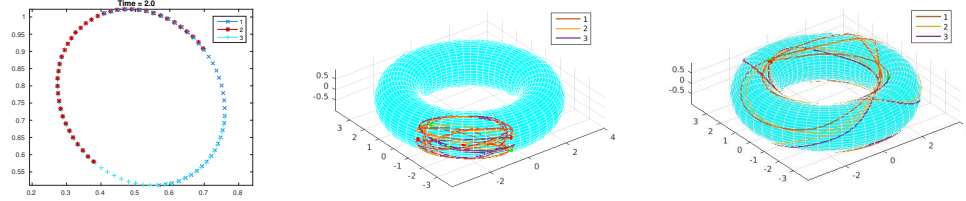
Figure 3.11: Trajectories of three agents interacting according to the matrix A given in (3.36). Left: Dynamics in $\mathbb{R}^2$, with periodic trajectories on a unique orbit. Center: Dynamics on $M = \mathbb{T}^2$ with small initial mutual distances. Right: Dynamics on $M = \mathbb{T}^2$ with large initial distances.

3.6   Social choreographies

As seen in Sections 3.3 and 3.5.3, when the interaction matrix A satisfies certain properties, for instance given by (3.28) on $\mathbb{S}^1$ or by (3.36) in $\mathbb{R}^2$, then the trajectories exhibit special properties of symmetry or periodicity. In [34], configurations on $\mathbb{S}^2$ in which all mutual distances between agents remain constant were named *dancing*

Figure 3.12: Evolutions of the coordinates of the three agents evolving on $\mathbb{T}^2$ with interaction matrix A from equation (3.36), with small initial mutual distances. Left: Evolution of ϕ. Center: Evolution of θ. Right: Evolution of the kinetic energy.

equilibrium.

In this section, we investigate systems with similar properties of periodicity or symmetry. We use the term *social choreography*, drawing a parallel with the well-known "n-body choreographies" discovered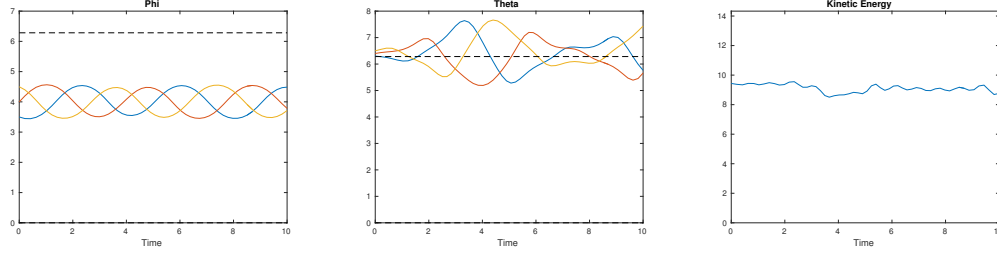 by Moore [53, 54] in the context of point masses subject to gravitational forces. In the n-body problem, the interaction potentials between masses are predetermined, as they depend exclusively on the masses and distances between agents. Hence the conditions for a n-body choreography to occur only depend on the initial state of the system. In the case of social choreography, there are more degrees of freedom, as we design the interaction matrix as well as to set the initial conditions.

We study sufficient conditions on the interaction matrices for the trajectories of the system to be periodic or symmetric by focusing on the Euclidean space $\mathbb{R}^2$, with the specific choice of interaction potential $\Psi \equiv \mathrm{Id}$. A future direction of this paper can consist in extending these results to general Riemannian manifolds. In $\mathbb{R}^2$ and with $\Psi \equiv \mathrm{Id}$, both approaches A and B are equivalent and the system simply reads as:

$$\text{for all } i \in \{1, \ldots, N\}, \quad \dot{x}_i = \sum_{j=1}^{N} a_{ij}(x_j - x_i). \tag{3.42}$$

We define the kinetic energy $E = E_G = E_P$ as in Definition 8:

$$E(t) := \frac{1}{2} \sum_{i=1}^{N} \|\dot{x}_i(t)\|^2. \tag{3.43}$$

A simple case of social choreography is that of a system with periodic trajectories, which we define as follows:

Definition 10 *Let $(x_i)_{i=1\ldots N}$ be a solution of (3.42). We refer to the system as having **periodic trajectories** if there exists $\tau > 0$ such that*

$$\textit{for all } i \in \{1, ..., N\}, \textit{ for all } t > 0, \quad x_i(t + \tau) = x_i(t).$$

We will examine possible periodic behaviors of the system in sections 3.6.2, 3.6.3 and 3.6.4.

3.6.1   Rotationally invariant system

We now give sufficient conditions on the interaction matrix and on the initial conditions for the system to be invariant by rotation.

Theorem 3 *Let $k \in \mathbb{N}$ such that k divides N. Let $P_k = \begin{pmatrix} 0 & I_{N-k} \\ I_k & 0 \end{pmatrix}$ be the matrix of change of basis from $(e_1, \ldots, e_N)$ to $(e_k, \ldots, e_N, e_1, \ldots, e_{k-1})$. Let $R(\theta)$ denote the rotation matrix in $\mathbb{R}^2$ for the angle $\theta \in [0, 2\pi)$. Suppose that initially, the system is invariant by rotation of angle $\frac{2k\pi}{N}$, that is:*

$$\textit{for all } i \in \{1, \ldots, N\}, \quad R(\frac{2k\pi}{N})x_i(0) = \begin{cases} x_{i+k}(0) & \textit{if } i + k \leq N \\ x_{i+k-N}(0) & \textit{if } i + k > N \end{cases}.$$

Suppose that the interaction matrix A is invariant by change of basis, i.e. $P_k^{-1} A P_k =$

A. Then the system remains invariant by rotation of angle $\frac{2k\pi}{N}$ at all time:

$$\text{for all } t > 0, \quad \text{for all } i \in \{1, \ldots, N\}, \quad R(\frac{2k\pi}{N})x_i(t) = \begin{cases} x_{i+k}(t) & \text{if } i+k \leq N \\ x_{i+k-N}(t) & \text{if } i+k > N \end{cases}.$$

Proof 16 *Let $A \in \mathcal{M}^N(\mathbb{R})$ be the interaction matrix, i.e. $A = (a_{ij})_{i,j=1,\ldots N}$, and define $D = \text{diag}(\sum_j a_{ij})$. Let $x = (x_1, \ldots, x_N)$ denote the set of all x_i's. It is a vector of length N with entries in $\mathbb{R}^2$. Let $X \in \mathcal{M}^{N \times 2}(\mathbb{R})$ denote the corresponding matrix of $\mathbb{R}^{N \times 2}$ such that for all $i \in \{1, \ldots, N\}$, for all $j \in \{1, 2\}$, X_{ij} is the j-th coordinate of x_i. With these notations, $\dot{X} = \tilde{A}X$, where $\tilde{A} = A - D$. We denote by $(e_1, \ldots, e_N)$ the canonical orthonormal basis of $(\mathbb{R})^N$ such that $X = \sum_{i=1}^N e_i x_i^T$.*

From the definition of the matrix X, the condition

$$\text{for all } i \in \{1, \ldots, N\}, \quad R(\frac{2k\pi}{N})x_i(0) = \begin{cases} x_{i+k}(0) & \text{if } i+k \leq N \\ x_{i+k-N}(0) & \text{if } i+k > N \end{cases}$$

can be rewritten as: $P_k X(0) = (R(\frac{2k\pi}{N})X(0)^T)^T$. Let $Y := P_k X$ and $Z := (R(\frac{2k\pi}{N})X^T)^T$. From the theorem's hypotheses, $Y(0) = Z(0)$. Let us show that Y and Z have the same evolution. One can easily prove that $P_k^{-1}\tilde{A}P_k$ if and only if $P_k^{-1}AP_k$. Then notice that

$$\dot{X} = \tilde{A}X = P_k^{-1}\tilde{A}P_k X.$$

From that we compute:

$$\dot{Y} = P_k \dot{X} = P_k(P_k^{-1}\tilde{A}P_k X) = \tilde{A}P_k X = \tilde{A}Y.$$

Similarly,

$$\dot{Z} = (R(\frac{2k\pi}{N})\dot{X}^T)^T = (R(\frac{2k\pi}{N})(\tilde{A}X)^T)^T = (R(\frac{2k\pi}{N})X^T\tilde{A}^T)^T = \tilde{A}Z.$$

Since Y and Z satisfy the same differential equation and $Y(0) = Z(0)$, then $Y(t) = Z(t)$ for all $t \geq 0$. This implies that at all time,

$$\text{for all } i \in \{1, \ldots, N\}, \quad R(\frac{2k\pi}{N})x_i(t) = \begin{cases} x_{i+k}(t) & \text{if } i + k \leq N \\ x_{i+k-N}(t) & \text{if } i + k > N \end{cases}.$$

Figure 3.13: Left: Evolution of 12 agents with the conditions of Theorem 3, with $k = 3$, resulting in diverging trajectories. Dark to light color scale indicates earlier to later time. Right: corresponding exploding kinetic energy. The interaction matrix A and the initial positions were generated according to a random algorithm, with the conditions of Theorem 3.

3.6.2 Unique orbit

Another example of social choreography is that of a system in which all agents share one unique orbit. Such choreographies have been discovered in the context of the n-body problem, for instance the "figure 8" orbit for three equal masses [53].

Definition 11 *Let $(x_i)_{i=1\ldots N}$ be a solution of (3.42). We say that the system has a **unique orbit** if the orbits of all points are identical, i.e.*

$$\text{for all } i, j \in \{1, \ldots, N\}, \quad \{z \in M | \exists t > 0, \ x_i(t) = z\} = \{z \in M | \exists t > 0, \ x_j(t) = z\}.$$

To illustrate Theorem 3, we study the evolution of N agents initially positioned at
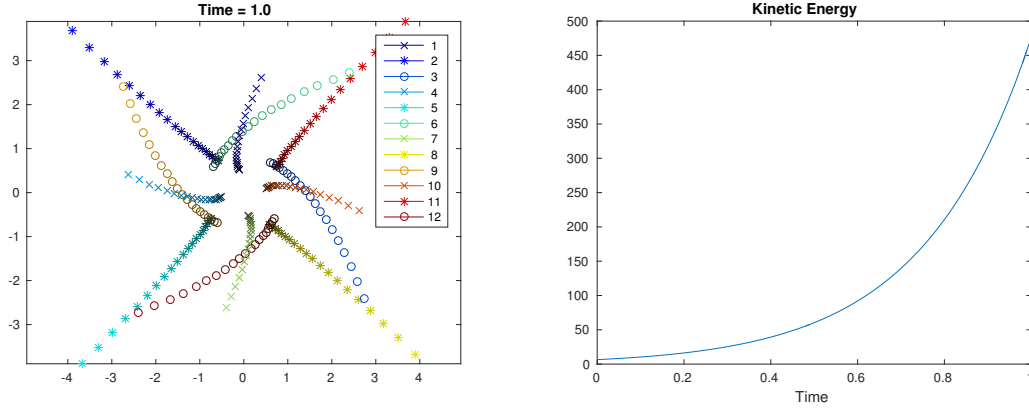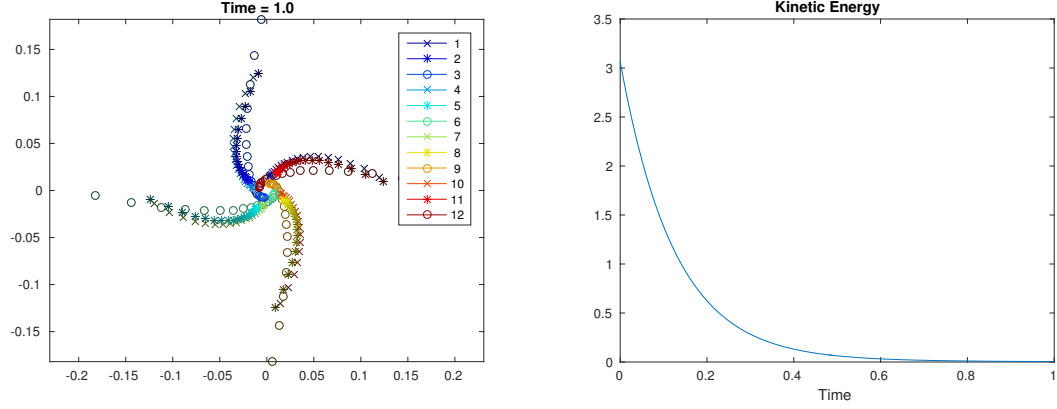
Figure 3.14: Left: Evolution of 12 agents with the conditions of Theorem 3, with $k = 3$, resulting in convergence to consensus. Dark to light color scale indicates earlier to later time. Right: corresponding kinetic energy converging to zero. The interaction matrix A and the initial positions were generated according to a random algorithm, with the conditions of Theorem 3.

regular intervals on a circle, with an interaction matrix and initial conditions given by:

$$
A = \begin{pmatrix}
0 & 1 & 0 & \dots & 0 & -1 \\
-1 & 0 & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & 1 \\
1 & 0 & \dots & 0 & -1 & 0
\end{pmatrix}
\tag{3.44}
$$

$$
\text{and for all } i \in \{1, \dots, N\}, \quad x_i(0) = \begin{pmatrix} \cos(\frac{2i\pi}{N}) \\ \sin(\frac{2i\pi}{N}) \end{pmatrix}.
$$

Notice that $\tilde{A} = A$, and the system satisfies the conditions of Theorem 3 with $k = 1$. Hence for all $i \in \{1, \dots, N-1\}$, $R(\frac{2\pi}{N})x_i(t) = x_{i+1}(t)$ and $R(\frac{2\pi}{N})x_N(t) = x_1(t)$. The $2N$-dimensional system then reduces to a 2-dimensional one for the two coordinates

x_{11} and x_{12} of x_1, and all the other variables can be recovered by rotation of x_1:

$$\dot{x}_1 = x_2 - x_N = R(\frac{2\pi}{N})x_1 - R(-\frac{2\pi}{N})x_1.$$

This can be written as:

$$\begin{pmatrix} \dot{x}_{11} \\ \dot{x}_{12} \end{pmatrix} = \begin{pmatrix} 0 & -2\sin(\frac{2\pi}{N}) \\ 2\sin(\frac{2\pi}{N}) & 0 \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix}.$$

Solving this linear system yields:

$$\begin{cases} x_{11}(t) = x_{11}(0)\cos(2\sin(\frac{2\pi}{N})t) - x_{12}(0)\sin(2\sin(\frac{2\pi}{N})t) = \cos(2\sin(\frac{2\pi}{N})t) \\ x_{12}(t) = x_{11}(0)\sin(2\sin(\frac{2\pi}{N})t) + x_{12}(0)\cos(2\sin(\frac{2\pi}{N})t) = \sin(2\sin(\frac{2\pi}{N})t) \end{cases}.$$

This proves that all agents share one common circular orbit, and their trajectories are periodic of period $2\pi(2\sin(\frac{2\pi}{N}))^{-1}$. Figure 3.15 provides a numerical illustration of this behavior, with 10 agents initially positioned at regular intervals on the unit circle.
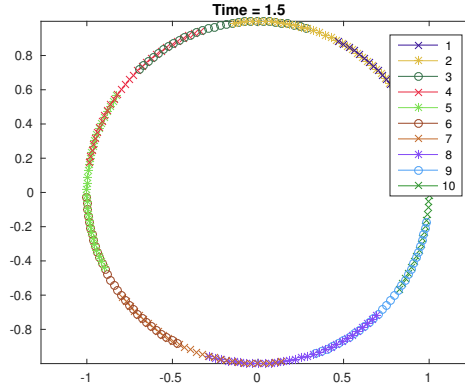

Figure 3.15: Evolution of 10 agents with initial conditions and interaction matrix given in (3.44). The agents have periodic trajectories along one shared circular orbit.

Another interesting example is that of 3 agents interacting according to the inter-

action matrix given previously, which, reduced to $N = 3$, gives:

$$A = \begin{pmatrix} 0 & 1 & -1 \\ -1 & 0 & 1 \\ 1 & -1 & 0 \end{pmatrix}. \tag{3.45}$$

Theorem 4 *Let $N = 3$. Consider the system (3.42) with interaction matrix given by (3.45). Then there exists a unique orbit shared by all agents, and all three trajectories are periodic.*

Proof 17 *The x and y-coordinates of the systems are decoupled, so that the 6-dimensi-onal system can be reduced to two 3-dimensional ones. Notice that $\tilde{A} = A$. Then for each coordinate $j \in \{1, 2\}$, the system reads:*

$$\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \end{pmatrix}(t) = \exp(tA) \begin{pmatrix} x_{1j}^0 \\ x_{2j}^0 \\ x_{3j}^0 \end{pmatrix}$$

with

$$e^{tA} =$$

$$\frac{1}{3} \begin{pmatrix} 1 + 2\cos(\sqrt{3}t) & 1 - \cos(\sqrt{3}t) + \sqrt{3}\sin(\sqrt{3}t) & 1 - \cos(\sqrt{3}t) - \sqrt{3}\sin(\sqrt{3}t) \\ 1 - \cos(\sqrt{3}t) - \sqrt{3}\sin(\sqrt{3}t) & 1 + 2\cos(\sqrt{3}t) & 1 - \cos(\sqrt{3}t) + \sqrt{3}\sin(\sqrt{3}t) \\ 1 - \cos(\sqrt{3}t) + \sqrt{3}\sin(\sqrt{3}t) & 1 - \cos(\sqrt{3}t) - \sqrt{3}\sin(\sqrt{3}t) & 1 + 2\cos(\sqrt{3}t) \end{pmatrix}.$$

Due to the special structure of e^{tA}, this can be rewritten as:

$$\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \end{pmatrix}(t) = \frac{1}{3} \begin{pmatrix} x_{1j}^0 & x_{2j}^0 & x_{3j}^0 \\ x_{2j}^0 & x_{3j}^0 & x_{1j}^0 \\ x_{3j}^0 & x_{1j}^0 & x_{2j}^0 \end{pmatrix} \begin{pmatrix} 1 + 2\cos(\sqrt{3}t) \\ 1 - \cos(\sqrt{3}t) + \sqrt{3}\sin(\sqrt{3}t) \\ 1 - \cos(\sqrt{3}t) - \sqrt{3}\sin(\sqrt{3}t) \end{pmatrix}.$$

This shows that all three trajectories are periodic, or period $\frac{2\pi}{\sqrt{3}}$. One can compute the positions of each agent after a third of a period and notice that:

$$\begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \end{pmatrix} (t + \frac{2\pi}{3\sqrt{3}})$$

$$= \frac{1}{3} \begin{pmatrix} x_{1j}^0 & x_{2j}^0 & x_{3j}^0 \\ x_{2j}^0 & x_{3j}^0 & x_{1j}^0 \\ x_{3j}^0 & x_{1j}^0 & x_{2j}^0 \end{pmatrix} \begin{pmatrix} 1 - \cos(\sqrt{3}t) - \sqrt{3}\sin(\sqrt{3}t) \\ 1 + 2\cos(\sqrt{3}t) \\ 1 - \cos(\sqrt{3}t) + \sqrt{3}\sin(\sqrt{3}t) \end{pmatrix} = \begin{pmatrix} x_{2j} \\ x_{3j} \\ x_{1j} \end{pmatrix} (t).$$

This shows that there is one unique shared orbit.

3.6.3 Coupled periodic trajectories

Other conditions on the interaction matrix A give rise to different kinds of periodic behaviors. Here we provide sufficient conditions for the system to exhibit periodic trajectories, such that each orbit is shared by two agents.

Theorem 5 (Coupled periodic trajectories) *Let N be even. Suppose that initially, the system is invariant by rotation of angle $\frac{4\pi}{N}$, that is:*

$$\text{for all } i \in \{1, \ldots, N\}, \quad R(\frac{4\pi}{N})x_i(0) = \begin{cases} x_{i+2}(0) & \text{if } i + 2 \leq N \\ x_{i+2-N}(0) & \text{if } i + 2 > N \end{cases}.$$

Let $a, b > 0$ and let

$$
A = \begin{pmatrix}
0 & a & 0 & \cdots & 0 & -b \\
-a & 0 & b & \ddots & \ddots & 0 \\
0 & -b & \ddots & a & \ddots & \vdots \\
\vdots & \ddots & \ddots & \ddots & \ddots & 0 \\
0 & \ddots & \ddots & \ddots & \ddots & a \\
b & 0 & \cdots & 0 & -a & 0
\end{pmatrix}.
\tag{3.46}
$$

Then the system is periodic of period $\tau = \frac{\pi}{\sqrt{ab}\sin(2\pi/N)}$. Furthermore, if N is divisible by 4, opposite agents share orbits two by two, i.e.:

$$
\text{for all } t > 0, \text{ for all } i \in \{1, \ldots, \frac{N}{2}\}, x_i(t + \tau) = x_{i+\frac{N}{2}}(t),
$$

and the kinetic energy is periodic with period $\tau/2$.

Proof 18 *First remark that the system satisfies the hypotheses of Theorem 3, so*

$$
\text{for all } t > 0, \quad \text{for all } i \in \{1, \ldots, N\}, \quad R(\frac{4\pi}{N})x_i(t) = \begin{cases} x_{i+2}(t) & \text{if } i + 2 \leq N \\ x_{i+2-N}(t) & \text{if } i + 2 > N \end{cases}.
$$

Hence the system is entirely known from the positions of the first two agents, since all others can be obtained by simple rotations. We show that this $2N$-dimensional problem can be rewritten as a 4-dimensional one. Indeed, using the fact that $x_N = R(-4\pi/N)x_2$ and $x_3 = R(4\pi/N)x_1$, the system

$$
\begin{cases}
\dot{x}_1 = a(x_2 - x_1) - b(x_N - x_1) \\
\dot{x}_2 = b(x_3 - x_2) - a(x_1 - x_2)
\end{cases}
$$

becomes:

$$
\begin{cases}
\dot{x}_1 = \begin{pmatrix} \dot{x}_{11} \\ \dot{x}_{12} \end{pmatrix} = a \left[\begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} - \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \right] - b \left[\begin{pmatrix} \cos(\frac{4\pi}{N}) & \sin(\frac{4\pi}{N}) \\ -\sin(\frac{4\pi}{N}) & \cos(\frac{4\pi}{N}) \end{pmatrix} \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} - \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} \right] \\[3em]
\dot{x}_2 = \begin{pmatrix} \dot{x}_{21} \\ \dot{x}_{22} \end{pmatrix} = b \left[\begin{pmatrix} \cos(\frac{4\pi}{N}) & -\sin(\frac{4\pi}{N}) \\ \sin(\frac{4\pi}{N}) & \cos(\frac{4\pi}{N}) \end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} - \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \right] - a \left[\begin{pmatrix} x_{11} \\ x_{12} \end{pmatrix} - \begin{pmatrix} x_{21} \\ x_{22} \end{pmatrix} \right]
\end{cases}.
$$

This can be rewritten in matrix form as:

$$
\begin{pmatrix} \dot{x}_{11} \\ \dot{x}_{12} \\ \dot{x}_{21} \\ \dot{x}_{22} \end{pmatrix} = A_4 \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{pmatrix}
\tag{3.47}
$$

where

$$
A_4 := \begin{pmatrix}
-a+b & 0 & a - b\cos(\frac{4\pi}{N}) & -b\sin(\frac{4\pi}{N}) \\
0 & -a+b & b\sin(\frac{4\pi}{N}) & a - b\cos(\frac{4\pi}{N}) \\
-a + b\cos(\frac{4\pi}{N}) & -b\sin(\frac{4\pi}{N}) & a-b & 0 \\
b\sin(\frac{4\pi}{N}) & -a + b\cos(\frac{4\pi}{N}) & 0 & a-b
\end{pmatrix} \begin{pmatrix} x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{pmatrix}.
$$

One can easily show that this reduced interaction matrix A_4 has two purely imaginary conjugate eigenvalues, $i\lambda$ and $-i\lambda$, each of multiplicity 2, where $\lambda = 2\sqrt{ab}$ $\sin(\frac{2\pi}{N})$. Hence the solution of the system (3.47) can be written as a weighted sum of the functions $t \mapsto \cos(\lambda t)$ and $t \mapsto \sin(\lambda t)$. This implies that the system is periodic, of period

$$
\tau = \frac{2\pi}{\lambda} = \frac{\pi}{\sqrt{ab}\sin(\frac{2\pi}{N})}.
$$

Furthermore, if N is divisible by 4, according to Theorem 3, $x_{\frac{N}{2}+1} = -x_1$ and $x_{\frac{N}{2}+2} = -x_2$. This implies that for all $t > 0$, $x_1(t + \tau) = -x_1(t) = x_{\frac{N}{2}+1}(t)$ and $x_2(t + \tau) =$
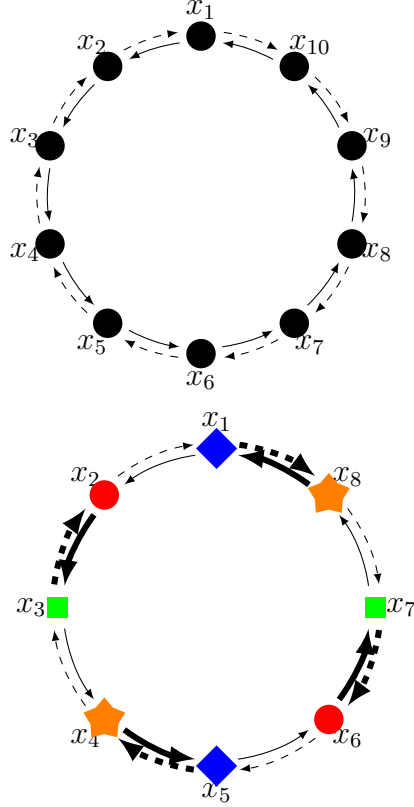
Figure 3.16: Left: Directed graph corresponding to the matrix A given in (3.44). Full arrows represent positive coefficients ($a_{ij} > 0$) while dashed ones represent negative coefficients ($a_{ij} < 0$). Right: Weighted directed graph corresponding to the matrix A given in (3.46). Thin arrows represent the weighted edges $|a_{ij}| = a$ while bold ones represent the weight $|a_{ij}| = b$. Nodes with the same color and symbol share orbits but are not directly connected in the graph.

$-x_2(t) = x_{\frac{N}{2}+2}(t)$, so the agents x_1 and $x_{\frac{N}{2}+1}$ share an orbit, as well as all pairs of agents x_i and $x_{\frac{N}{2}+i}$ for $i \in \{1, \ldots, \frac{N}{2}\}$.

As a consequence, the kinetic energy is periodic, of period $\tau_E = \tau = \pi/(\sqrt{ab} \sin(\frac{2\pi}{N}))$. If N is divisible by 4, every half period, the system is rotated by an angle π, so the kinetic energy is periodic with period $\tau_E = \tau/2$.

Remark 7 *Notice that the agents sharing orbits do not interact with one another, as shown in Figure 3.16.*

An example of such a choreography is given in Figure 3.17.
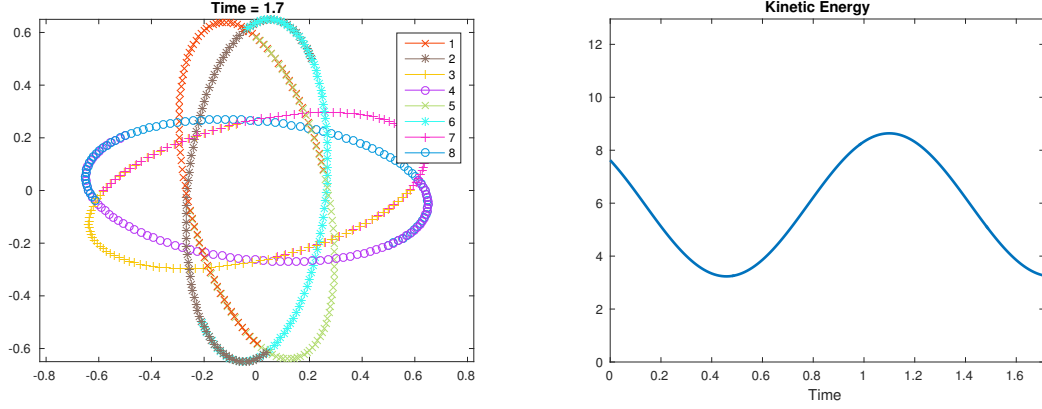
Figure 3.17: Left: Periodic trajectories of 8 agents sharing orbits two by two, in the situation of Theorem 5. Matrix A from (3.46) was constructed with $(a, b) = (1, 3)$. The initial positions $x_1(0)$ and $x_2(0)$ were randomly generated and the other 6 were obtained by rotation. The period is $\tau = 2\pi/\sqrt{6}$. Right: Corresponding kinetic energy, of period $\tau/2$.

Remark 8 *As a slight generalization, we provide numerical simulations illustrating a similar behavior, but with slightly different conditions: the periodic evolution of 9 agents on three distinct orbits shared three by three, see figures 3.18 and 3.19.*

3.6.4 Helical trajectories

In sections 3.6.2 and 3.6.3, we provided conditions for the trajectories of the system to be periodic. Here, we explore further the notion of periodicity by studying systems with drift, displaying helical trajectories but periodic kinetic energy.

Definition 12 *Let $(x_i)_{i=1...N}$ be a solution of (3.42). We call the corresponding trajectories* ***helical trajectories*** *if there exists $v \in \mathbb{R}^2$ and $\tau \in \mathbb{R}^*$ such that*

$$\text{for all } i \in \{1, ..., N\}, \text{ for all } t > 0, \quad x_i(t + \tau) = x_i(t) + \tau v.$$

Notice that this definition generalizes the notion of periodic trajectories recalled in Definition 10, which corresponds to the case $v = 0$. When $v \neq 0$, the system has a drift term, meaning that the relative positions between agents remain periodic but
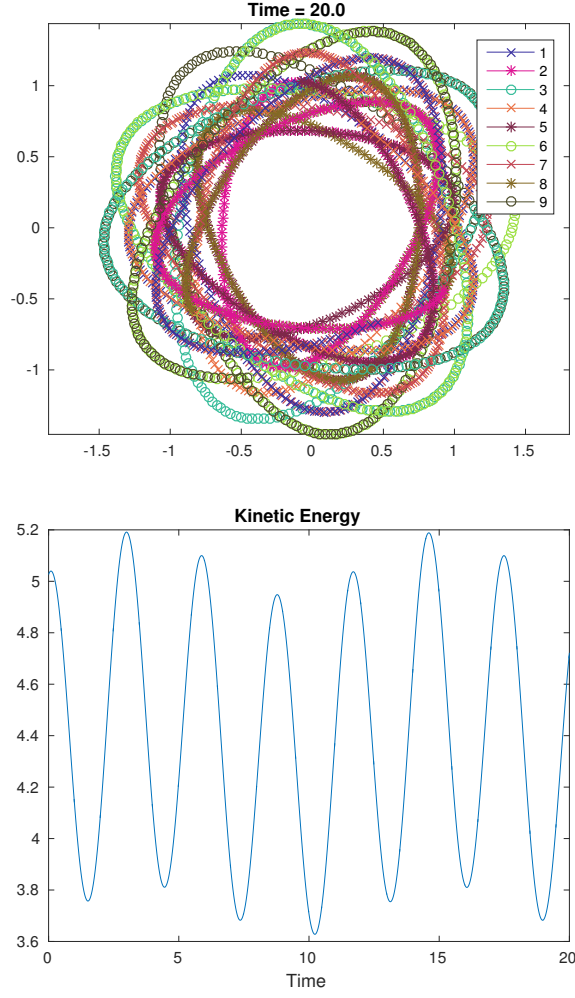
Figure 3.18: Left: evolution of 9 agents with periodic trajectories, each orbit shared by 3 agents. Right: periodic kinetic energy.

their absolute positions evolve in space.

Theorem 6 *Sufficient conditions for helical trajectories. Let $N = 4$. Let $(a, b, c, d) \in (\mathbb{R}^+)^4$ such that the interaction matrix reads*

$$A = \begin{pmatrix} 0 & a & 0 & -d \\ -a & 0 & b & 0 \\ 0 & -b & 0 & c \\ d & 0 & -c & 0 \end{pmatrix}. \tag{3.48}$$
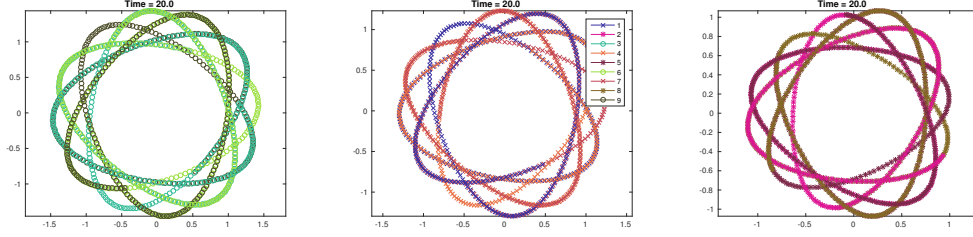
Figure 3.19: Isolated orbits of the evolution shown in Figure 3.18. Left: trajectories of agents 3, 6, 9. Middle: trajectories of agents 1, 4, 7. Right: trajectories of agents 2, 5, 8).

Then the system exhibits helical trajectories.

Proof 19 *First notice that the first and second components x_{i1} and x_{i2} of the i-th agent's position are decoupled, so that the system in matrix form reads*

$$
\dot{x}^j = \begin{pmatrix} \dot{x}_{1j} \\ \dot{x}_{2j} \\ \dot{x}_{3j} \\ \dot{x}_{4j} \end{pmatrix} = \begin{pmatrix} d-a & a & 0 & -d \\ -a & a-b & b & 0 \\ 0 & -b & b-c & c \\ d & 0 & -c & c-d \end{pmatrix} \begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \end{pmatrix} := \tilde{A} \begin{pmatrix} x_{1j} \\ x_{2j} \\ x_{3j} \\ x_{4j} \end{pmatrix}, \text{ for } j \in \{1,2\}.
$$

$$(3.49)$$

Hence the projections of x on the first and second axes solve the same differential equation. The matrix $\tilde{A}$ has three distinct eigenvalues:

$$
\lambda_1 = 0, \quad \lambda_2 = i\sqrt{(a+c)(b+d)} \quad \text{and} \quad \lambda_3 = -i\sqrt{(a+c)(b+d)}.
$$

There is one eigenvector associated with λ_1: $v_1 := (1,1,1,1)^T$. One can show that the vectors $x(t) = v_1$ and $x(t) = v_1 t + \nu$ are both solutions of System (3.49), where, denoting $\Delta := bcd - abc + abd - acd$,

$$
\nu := \frac{1}{\Delta}(ab + bc + \Delta, ab - cd + \Delta, ab + ad + \Delta, \Delta)^T.
$$

Let v_2 denote the eigenvector associated with λ_2 and let v_2^R and v_2^I denote respectively

its real and imaginary components, i.e. $v_2 := v_2^R + i v_2^I$. Then the solution of System (3.49) can be written as:

$$x^j(t) = C_1^j v_1 + C_2^j(v_1 t + \nu) + C_3^j \left[v_2^R \cos(\lambda_2 t) - v_2^I \sin(\lambda_2 t) \right]$$
$$+ C_4^j \left[v_2^R \sin(\lambda_2 t) + v_2^I \cos(\lambda_2 t) \right]$$

where $(C_1, C_2, C_3, C_4) \in \mathbb{R}^4$ are constants depending on the initial conditions. Let $\tau = \frac{2\pi}{\lambda_2}$. Then for all $t > 0$, for all $i \in \{1, \ldots, 4\}$, for all $j \in \{1, 2\}$, $x_{ij}(t + \tau) = x_{ij}(t) + C_2^j \tau$. This can be rewritten as:

$$\text{for all } i \in \{1, \ldots, 4\}, \text{ for all } t > 0, \quad x_i(t + \tau) = x_i(t) + \begin{pmatrix} C_2^1 \\ C_2^2 \end{pmatrix} \tau.$$
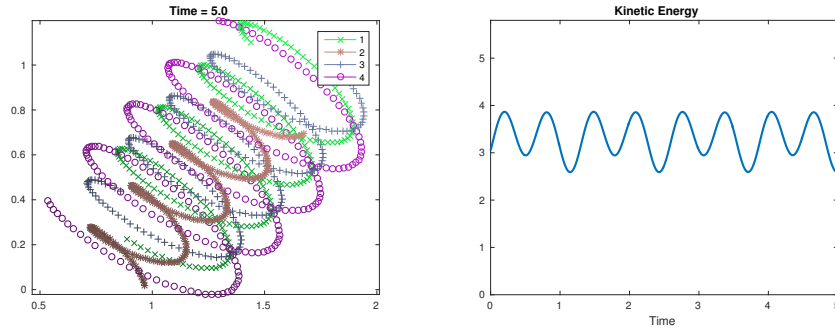
Figure 3.20: Left: Trajectories of 4 agents with helical trajectories. Parameters for matrix A (3.48) chosen to be $(a, b, c, d) = (1, 2, 3, 4)$. Dark to light color indicates earlier to later time. Right: Corresponding kinetic energy. The period is $\tau = 2\pi((a + c)(b + d))^{-1/2} = \pi/\sqrt{6}$ (see proof of Theorem 6).
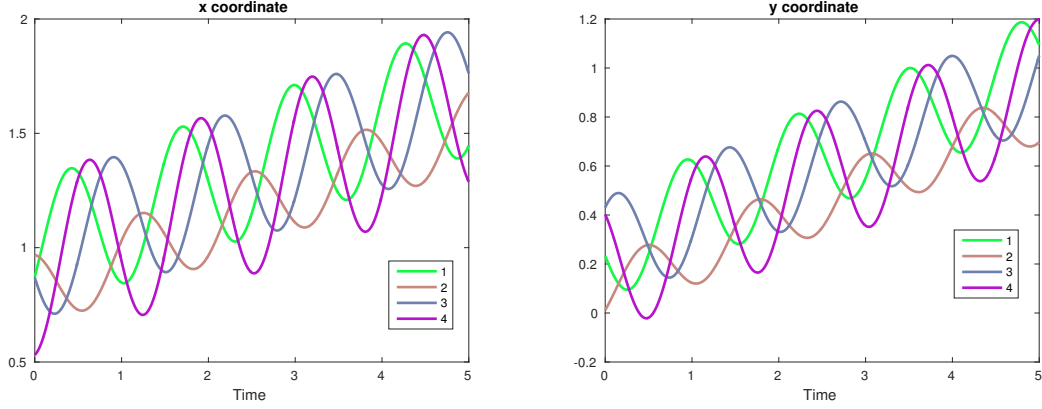
Figure 3.21: Evolution of the first and second coordinates of 4 agents with helical trajectories.

Theorem 7 *A system with helical trajectories has periodic kinetic energy.*

Proof 20 *Supose that $(x_i)_{i=1...N}$ has helical trajectories, i.e. there exists $\tau \in \mathbb{R}$, $v \in \mathbb{R}^2$ such that for all $i \in \{1, \ldots, N\}$, for all $t \geq 0$, $x_i(t + \tau) = x_i(t) + \tau v$. Then $\dot{x}_i(t + \tau) = \dot{x}_i(t)$ and so $E(t + \tau) = E(t)$.*

3.7   Introduction to opinion models with time-varying influence

In this section we investigate opinion models where each agent has a time-varying weight that determines the influence over other agents. We propose four different models for which the influence evolves. For all models we impose the conservation of the sum of the weights, so that the weights simply shift between agents. In Model 1, agents gain mass if they attract the other agents more than they are attracted in pairwise interactions. This is achieved by looking at the midpoint dynamics. In Model 2, an agent gains mass if its stance or opinion is well received by many others. An example could model a politician using a wholly popular idea to gain influence. Many students and young voters may give support to free college in the United States which increases the influence of the agent with this idea. In Model 3, an agent gains mass

by attracting other influential individuals. This model is appropriate for modeling opinions in the situation where very influential people appear to support opinions publicly. An example is when a musician, movie star, or popular artist supports a political candidate. The political candidate may gain influence because they appear to attract other influential individuals. Lastly, Model 4 represents a continuous version of Model 3, where we replace the set of most influential agents with a continuously varying coefficient.

We first address the problem of existence of solutions. Our models are smooth in the space variable but present natural discontinuities in the weight dynamics. Thus we consider solutions in the Filippov sense and prove that the space trajectory is always unique. This allows to prove that agents never coalesce in finite time and, for Models 1 and 2, if we start with some agents in the same position we can replace them by a single one whose weight is the sum of the weights.

We then prove some general properties of the system: the maximal norm of opinions is contracting as well as the convex hull of opinions. We also provide estimates on the diameter dynamics in terms of the interaction functional. We next focus on specific properties of the interaction function. If the interaction does not vanish for large distances, then consensus is always achieved exponentially fast in time. For the bounded-confidence case we prove that clustering occurs using an appropriate energy functional. We then deal with heterophilious dynamics, in the specific case of a constant interaction function. In this simplified case, we highlight the main characteristics and constitutive differences in the large-time dynamics of our four models. For Model 1, the weights never vanish and have an a-priori lower bound. Therefore, no agent looses completely its weight (not even asymptotically) and no agent gains all the weight. On the contrary, for Model 2, a single agent will asymptotically gain all the weight and such agent can be identified a priori. Finally, for Models 3 and 4, equilibria exhibit two agents sharing all the weight and we conjecture that this are

also the asymptotic states.

Then we turn our attention to numerical simulations. We show the behavior of all models with various agents starting from random positions and with random weights. We also measure the speed of convergence to consensus (when it is achieved) and show that the weight dynamics tend to facilitate consensus and increase the convergence speed for some of the proposed models. We then compare the clustering behavior of the four models in the case of an interaction function with bounded support. Lastly, we provide some conclusions and future directions.

3.8 General results

3.8.1 Notations and model definitions

Let $(x_i)_{i \in \{1, \cdots, N\}} \in (\mathbb{R}^d)^N$ represent the opinions (or positions) of a group of N agents. A general opinion dynamics model, as introduced by Hegselmann and Krause[2], can be written as:

$$
\begin{cases}
\dot{x}_i(t) = \dfrac{1}{N} \sum_{j=1}^{N} a(\|x_i(t) - x_j(t)\|) \ (x_j(t) - x_i(t)), \\
x_i(0) = x_i^0,
\end{cases}
\qquad i \in \{1, \cdots, N\}, \qquad (3.50)
$$

where $a(\cdot)$ represents the interaction function. This general system models the dynamics of a group of agents in which each agent j's influence on another individual i depends on the distance between them and on the position of j with respect to i.

We extend this model by attributing to each agent a weight of influence $m_i \in \mathbb{R}^+$, which evolves in time due to its own dynamics (note that we will also refer to it as the agent's mass). Let $M := \sum_{i=1}^{N} m_i$ be the total weight (or mass) of the system. A

general form of this augmented model will be written as:

$$
\begin{cases}
\dot{x}_i(t) = \dfrac{1}{M} \displaystyle\sum_{j=1}^{N} m_j(t) a(\|x_i(t) - x_j(t)\|) \ (x_j(t) - x_i(t)), \\[4mm]
\dot{m}_i(t) = \Psi(m(t), x(t), m_i(t), x_i(t)), \qquad\qquad i \in \{1, \cdots, N\}, \\[4mm]
x_i(0) = x_i^0, \quad m_i(0) = m_i^0,
\end{cases}
\tag{3.51}
$$

where we denoted by m the vector of N weights $m = (m_i)_{i \in \{1, \cdots, N\}}$ and by x the N opinions $x = (x_i)_{i \in \{1, \cdots, N\}}$. In this new approach, an agent j's influence on i no longer depends solely on the vector $x_j - x_i$, but also on the weight m_j. This time-varying weight dynamics will be generally prescribed by a function Ψ depending on the group's opinions $(x_i)_{i \in \{1, \cdots, N\}}$ and weights of influence $(m_i)_{i \in \{1, \cdots, N\}}$. From here onward, for more simplicity we will denote by ψ_i the function Ψ evaluated at x_i and m_i, i.e. $\psi_i(m, x) := \Psi(m, x, m_i, x_i)$. We rewrite system (3.51) in the condensed form:

$$
\begin{cases}
\dot{x}_i(t) = X_i(x(t), m(t)), \\[3mm]
\dot{m}_i(t) = \psi_i(x(t), m(t)), \qquad i \in \{1, \cdots, N\}. \\[3mm]
x_i(0) = x_i^0, \quad m_i(0) = m_i^0,
\end{cases}
\tag{3.52}
$$

where we denoted by $X_i : (\mathbb{R}^d)^N \times \mathbb{R}^N \to (\mathbb{R}^d)^N$ the function

$$
X_i : (x, m) \mapsto \frac{1}{M} \sum_{j=1}^{N} m_j a(\|x_i - x_j\|) \ (x_j - x_i).
\tag{3.53}
$$

Furthermore, we assume that for each $i \in \{1, \cdots, N\}$, there exists a function $\overline{\psi}_i$ such that

$$
\forall (x, m) \in (\mathbb{R}^d)^N \times \mathbb{R}^N, \quad \psi_i(x, m) = m_i \overline{\psi}_i(x, m).
\tag{3.54}
$$

We will also use the notations

$$X := (X_i)_{i \in \{1, \cdots, N\}}, \quad \psi := (\psi_i)_{i \in \{1, \cdots, N\}}, \quad \text{and} \quad \overline{\psi} := (\overline{\psi}_i)_{i \in \{1, \cdots, N\}},$$

and from here onward we assume the following:

Hypothesis 1 *The function $s \mapsto a(\|s\|)s$ is locally Lipschitz in $\mathbb{R}^d$, and the function ψ is locally bounded in $(\mathbb{R}^d)^N \times \mathbb{R}^N$.*

Remark 9 *The fact that the mass dynamics satisfy* (3.54) *together with Hypothesis 1 implies that non-negative weights remain non-negative at all time.*

We also require the following condition on the weight dynamics:

Hypothesis 2 *For all $(x, m) \in (\mathbb{R}^d)^N \times \mathbb{R}^N$,*

$$\sum_{i=1}^{N} \psi_i(x, m) = 0. \tag{3.55}$$

This implies that the total weight M of the group stays constant, so that the weight distribution shifts between agents while the sum of the weights remains unchanged. As a consequence, the weights satisfy $m \in [0, M]^N$, where $M = \sum_{i=1}^{N} m_i^0$.

From here onward, unless we specify otherwise, all results concern the general model with weights (3.52)$-$(3.53)$-$(3.54)**, together with Hypotheses 1 and 2.**

We now present four specific models in which we prescribe the dynamics ψ_i of the weights of influence. The models we build rely on giving a meaning to the following quantities:

- x_i, referred to as the *opinion* (or *position*) of agent i

- m_i, referred to as the *weight* (or *mass*) of agent i

- $M = \sum_{i=1}^{N} m_i^0$, referred to as the *total mass* of the system

- $a(\|x_i - x_j\|)\|x_i - x_j\|$, referred to as the *positional influence* of i over j (and symmetrically, of j over i)

- $p_i := \sum_{j=1}^{N} a(\|x_i - x_j\|)\|x_i - x_j\|$, referred to as the *total positional influence* of i over the group

- $e_{ij} := m_j a(\|x_i - x_j\|)\|x_i - x_j\|$, referred to as the the *influence* of j over i (notice that e_{ij} is obtained by multiplying the positional influence of j over i by the weight of j)

- $I_i := \operatorname{argmax}_{j \in \{1, \cdots, N\}} e_{ij} = \{j \in \{1, \cdots, N\} \mid \forall k \in \{1, \cdots, N\}, e_{ij} \geq e_{ik}\}$, representing the set of agents that have the largest influence on i

Model 1: Increasing weight by pairwise competition. In this first model, the dynamics of an agent's weight m_i depend on the dynamics of the midpoints between i and each other agent j:

$$
\begin{cases}
\dot{x}_i = \dfrac{1}{M} \sum_{j=1}^{N} m_j \, a(\|x_i - x_j\|)(x_j - x_i) \\
\dot{m}_i = \dfrac{m_i}{M} \sum_{j=1}^{N} m_j \langle \dfrac{\dot{x}_i + \dot{x}_j}{2}, u_{ji} \rangle
\end{cases}
\qquad i \in \{1, \cdots, N\}, \qquad \text{(M1)}
$$

where u_{ji} is the vector defined by:

$$
u_{ji} = \begin{cases}
\dfrac{x_i - x_j}{\|x_i - x_j\|} & \text{if } x_i \neq x_j \\[2mm]
0 & \text{if } x_i = x_j.
\end{cases}
$$

If the midpoint $\frac{x_i + x_j}{2}$ moves in the direction of x_i, i.e. $\langle \frac{\dot{x}_i + \dot{x}_j}{2}, \frac{x_i - x_j}{\|x_i - x_j\|} \rangle \geq 0$, then i gains influence and m_i increases, proportionally to the projection of the velocity of

the midpoint on the direction $\frac{x_i - x_j}{\|x_i - x_j\|}$ and proportionally to j's own mass m_j. Notice that for compactness of notation, we expressed $\dot{m}_i$ in terms of x, m and $\dot{x}$. However, by plugging the expressions of $\dot{x}_i$ and $\dot{x}_j$ in the second equation of (M1), we recover an equation of the form $\dot{m}_i = \psi_i(x, m)$, so we are in the framework of the general system $(3.52) - (3.53) - (3.54)$.

Model 2: Increasing weight by positional influence. In this model, an agent's weight m_i increases if the agent's positional influence p_i is higher than the weighted average of the group's positional influences. We write:

$$
\begin{cases}
\dot{x}_i = \dfrac{1}{M} \sum_{j=1}^{N} m_j \, a(\|x_i - x_j\|)(x_j - x_i) \\
\dot{m}_i = \dfrac{m_i}{M} \left(\dfrac{p_i}{\frac{1}{M} \sum_{j=1}^{N} m_j p_j} - 1 \right)
\end{cases}
\qquad i \in \{1, \cdots, N\}, \qquad \text{(M2)}
$$

where $p_i := \sum_{j=1}^{N} a(\|x_i - x_j\|)\|x_i - x_j\|$ represents an agent's total positional influence on the group.

Model 3: Increasing weight by exerting the biggest influence on the most influential. In this model, m_i will increase if the agents that i influences the most are themselves influential. We give:

$$
\begin{cases}
\dot{x}_i = \dfrac{1}{M} \sum_{j=1}^{N} m_j a(\|x_i - x_j\|)(x_j - x_i) \\
\dot{m}_i = m_i \left(\sum_{j \mid i \in I_j} \dfrac{m_j}{|I_j|_m} - 1 \right)
\end{cases}
\qquad i \in \{1, \cdots, N\}, \qquad \text{(M3)}
$$

where $|I_j|_m := \sum_{k \in I_j} m_k$ represents the weight of the agents that influence j the most. As defined earlier, $I_j = \operatorname{argmax}_{k \in \{1, \cdots, N\}} e_{jk}$ is the set of agents that influence j the most, with $e_{jk} = m_k a(\|x_j - x_k\|)\|x_j - x_k\|$ representing the influence of k over j. A sufficient condition for m_i to increase is if there exists j with a greater weight than

i (i.e. $m_j > m_i$) such that i is the most influential agent on j, i.e. $I_j = \{i\}$. More generally, if $\#I_j = 1$ for all $j \in \{1, \cdots, N\}$ (where $\#$ denotes the cardinality of a set), $\dot{m}_i > 0$ if and only if the sum of the weights of all the agents for whom agent i is the most influential is greater than its own weight m_i. Indeed, in this case, $i \in I_j$ implies that $|I_j|_m = m_i$, and we get:

$$\dot{m}_i > 0 \quad \Leftrightarrow \quad \sum_{j \,|\, i \in I_j} \frac{m_j}{|I_j|_m} > 1 \quad \Leftrightarrow \quad \sum_{j \,|\, i \in I_j} \frac{m_j}{m_i} > 1 \quad \Leftrightarrow \quad \sum_{j \,|\, i \in I_j} m_j > m_i.$$

Model 4: Increasing weight by influencing the most influential. This last model provides a "continuous" version of the topological model (M3):

$$\begin{cases} \dot{x}_i = \dfrac{1}{M} \displaystyle\sum_{j=1}^{N} m_j a(\|x_i - x_j\|)(x_j - x_i) \\ \dot{m}_i = m_i \left(\dfrac{1}{M} \displaystyle\sum_{j=1}^{N} m_j \dfrac{e_{ji}}{\frac{1}{M} \sum_{k=1}^{N} m_k e_{jk}} - 1 \right) \end{cases} \quad i \in \{1, \cdots, N\}, \qquad \text{(M4)}$$

where $e_{ji} = m_i a(\|x_i - x_j\|)\|x_i - x_j\|$ represents the influence of agent i on agent j. In this model, an agent j contributes to increasing the weight of an agent i proportionally to its own weight m_j and to the strength of the influence of i on j relatively to the weighted average of all influences on j. As opposed to Model (M3), in the evolution of m_i we consider not only the agents for which i has the strongest influence, but all agents j, with a weight corresponding to how much i counts in the total influence perceived by j.

3.8.2 Existence and uniqueness of solutions

We begin by establishing the existence of solutions to the general system $(3.52)-(3.53)-(3.54)$. Let us first recall the definition of Filippov set-valued maps[55]:

Definition 13 *Let $\mathcal{P}(E)$ denote the collection of susbsets of the Euclidean space E.*

For a measurable vector field $X : E \to E$, the Filippov set-valued map $F[X] : E \to \mathcal{P}(E)$ is defined by:

$$\forall x \in E, \quad F[X](x) := \bigcap_{\delta > 0} \bigcap_{S \in \mathcal{P}(E),\ \mu(S) = 0} \overline{\mathrm{co}}\{X(B(x, \delta) \setminus S)\},$$

where $\overline{\mathrm{co}}\{\cdot\}$ denotes the convex closure, μ denotes the Lebesgue measure, $B(x, \delta)$ denotes the ball of center x and radius δ, and for all $S \in \mathcal{P}(E)$, $X(S) := \{X(x) \mid x \in S\}$.

Proposition 1 *Consider the differential system* $(3.52)-(3.53)-(3.54)$ *and suppose that* $s \mapsto a(\|s\|)s \in \mathrm{Lip}_{\mathrm{loc}}(\mathbb{R}^d)$ *and* $\psi \in L^{\infty}_{\mathrm{loc}}((\mathbb{R}^d)^N \times \mathbb{R}^N)$. *Then there exists a solution* (x, m) *to system* $(3.52)-(3.53)-(3.54)$ *in the sense of Filippov, i.e. an absolutely continuous map* $(x, m) : [0, T] \to (\mathbb{R}^d)^N \times \mathbb{R}^N$ *such that for almost all* $t \in [0, T]$,

$$(\dot{x}, \dot{m})(t) \in F[(X, \psi)](x(t), (m(t)),$$

where $F[(X, \psi)]$ *denotes the Filippov set-valued map of* (X, ψ). *Furthermore, for all solutions* (x, m) *and* (y, p) *satisfying* $(x(0), m(0)) = (y(0), q(0))$, *if* $m \equiv q$, *then* $x \equiv y$.

Proof 21 *The existence of solutions is a direct application of the notion of Filippov solutions[55], since the local boundedness of* (X, ψ) *implies the upper semicontinuity of* $F[(X, \psi)]$. *Uniqueness of solutions is not guaranteed since* (X, ψ) *is not locally Lipschitz. However, for each solution* $t \mapsto m(t)$, *the first equation of system* $(3.52)-(3.53)-(3.54)$ *can be rewritten as:*

$$\dot{x}_i = X_i(x, m(t)) := \tilde{X}_i(x, t) \tag{3.56}$$

where $\tilde{X} := (\tilde{X}_i)_{i \in \{1, \cdots, N\}}$ *is locally Lipschitz with respect to the* $\ell_1^N - \ell_2^d$ *norm* $x \mapsto \sum_{i=1}^{N} \|x_i\|$, *where* $\| \cdot \|$ *is the Euclidean norm of* $\mathbb{R}^d$. *Indeed, let* $\delta > 0$, *and let*

$(x, y) \in (\mathbb{R}^{dN})^2$ *such that* $\sum_{i=1}^{N} \|x_i - y_i\| \leq \delta$. *Then for all* $(i, j) \in \{1, \cdots, N\}^2$,

$\|(x_j - x_i) - (y_j - y_i)\| \leq \|(x_j - y_j)\| + \|(x_i - y_i)\| \leq \delta$, *so there exists* $K \geq 0$ *such that for each* $i \in \{1, \cdots, N\}$,

$$
\begin{aligned}
\|\tilde{X}_i(t, x) - \tilde{X}_i(t, y)\| &= \left\| \frac{1}{M} \sum_{j=1}^{N} m_j \left[a(\|x_i - x_j\|)(x_j - x_i) - a(\|y_i - y_j\|)(y_j - y_i) \right] \right\| \\
&\leq \frac{1}{M} \sum_{j=1}^{N} m_j K \|(x_j - x_i) - (y_j - y_i)\| \\
&\leq \frac{K}{M} \sum_{j=1}^{N} m_j (\|x_j - y_j\| + \|x_i - x_j\|) \leq 2K \sum_{j=1}^{N} \|x_j - y_j\|.
\end{aligned}
$$

Hence (3.56) *admits a unique solution from each initial condition* $x_0 \in (\mathbb{R}^d)^N$.

An important property follows from the uniqueness of x given the weights $t \mapsto m(t)$: if two agents are clustered at some time T, *i.e.* $x_i(T) = x_j(T)$, then they are necessarily clustered at all time before and after T. As a consequence, in one-dimension, the dynamics preserve the order.

Corollary 2 *Let* $(x, m) : \mathbb{R}^+ \mapsto (\mathbb{R}^d)^N \times [0, M]^N$ *denote a solution to system* (3.51). *If for some* $T \geq 0$, $x_i(T) = x_j(T)$, *then* $x_i(t) = x_j(t)$ *for all* $t \in \mathbb{R}^+$. *Hence, if* $d = 1$, *if* $x_i(0) \geq x_j(0)$, *then* $x_i(t) \geq x_j(t)$ *for all* $t \in \mathbb{R}^+$.

Proof 22 *Suppose that for some* $T \geq 0$, $x_i(T) = x_j(T)$. *Then from* (3.51), *one easily computes:* $\dot{x}_i(T) = \dot{x}_j(T)$. *From Proposition 1, there is a unique solution* x *for each* m *satisfying* (3.51), *hence* $x_i(t) = x_j(t)$ *for all* $t \geq T$ *and* $t \leq T$.

From here onward, we will suppose that at initial time, $x_i^0 \neq x_j^0$ for all $(i, j) \in \{1, \cdots, N\}^2$. In two of the specific models defined in Section 3.8.1, this can be done without loss of generality, as we show with the following:

Lemma 3 *Let* $t \mapsto (x(t), m(t)) \in \mathbb{R}^{dN} \times \mathbb{R}^N$ *and* $t \mapsto (y(t), q(t)) \in \mathbb{R}^{d(N-1)} \times \mathbb{R}^{N-1}$

be two solutions to system (M1) *(respectively* (M2)*). If the initial conditions satisfy*

$$\begin{cases} x_1^0 = x_N^0 = y_1^0 \\ m_1^0 + m_N^0 = q_1^0 \end{cases}$$

then $y_1 \equiv x_1 \equiv x_N$, $q_1 \equiv m_1 + m_N$, *and for all* $i \in \{2, \cdots, N-1\}$, $y_i \equiv x_i$ *and* $q_i \equiv m_i$.

Proof 23 *First, from Corollary 2, it follows immediately that* $x_1 \equiv x_N$, *independently of the weight dynamics. Secondly, let* $\tilde{x} := (x_1, \cdots, x_{N-1})$ *and* $\tilde{m} := (m_1 + m_N, \cdots, m_{N-1})$. *We show that for each of the two models* (M1) *and* (M2), $(\tilde{x}, \tilde{m})$ *and* (y, p) *solve the same differential systems. For Model* (M1),

$$\begin{cases} \dot{m}_1 + \dot{m}_N = \dfrac{m_1 + m_N}{M} \sum_{j=2}^{N-1} m_j \langle \dfrac{\dot{x}_1 + \dot{x}_j}{2}, u_{j1} \rangle \\ \dot{m}_i = \dfrac{m_i}{M} \left((m_1 + m_N) \langle \dfrac{\dot{x}_i + \dot{x}_1}{2}, u_{1i} \rangle + \sum_{j=2}^{N-1} m_j \langle \dfrac{\dot{x}_i + \dot{x}_j}{2}, u_{ji} \rangle \right), \quad 2 \le i \le N-1. \end{cases}$$

For Model (M2), *since* $x_1 = x_N$, *we have* $p_1 = p_N$, *and so*

$$\begin{cases} \dot{m}_1 + \dot{m}_N = \dfrac{m_1 + m_N}{M} \left(\dfrac{p_1}{\frac{1}{M}((m_1 + m_N)p_1 + \sum_{j=1}^{N-1} m_j p_j)} - 1 \right) \\ \dot{m}_i = \dfrac{m_i}{M} \left(\dfrac{p_i}{\frac{1}{M}((m_1 + m_N)p_1 + \sum_{j=1}^{N-1} m_j p_j)} - 1 \right), \quad 2 \le i \le N-1. \end{cases}$$

Then for either Model (M1) *or Model* (M2), $(\tilde{x}, \tilde{m})$ *and* (y, q) *follow the same dynamics, and since the initial conditions are the same,* $(\tilde{x}, \tilde{m}) \equiv (y, q)$.

Hence, in Models (M1) and (M2), if for some $(i, j) \in \{1, \cdots, N\}^2$, $x_i^0 = x_j^0$, we remove x_j from the system and replace m_i^0 by $m_i^0 + m_j^0$. The same cannot be done in the general case, when $\overline{\psi}_i$ depends explicitely on m_i.

Proposition 2 *The models* (M1), (M2), (M3) *and* (M4) *satisfy Hypothesis 2.*

Proof 24 *We compute $\sum_{i=1}^{N} \dot{m}_i$ for each model.*

For Model 1, the property $\sum_{i=1}^{N} \dot{m}_i = 0$ follows easily from the antisymmetry of the summed coefficients.

For Model 2,

$$\sum_{i=1}^{N} \dot{m}_i = \frac{\frac{1}{M} \sum_{i=1}^{N} m_i p_i}{\frac{1}{M} \sum j = 1^N m_j p_j} - \frac{\sum_{i=1}^{N} m_i}{M} = 1 - 1 = 0.$$

For Model 3,

$$\sum_{i=1}^{N} \dot{m}_i = \sum_{i=1}^{N} m_i \Big(\sum_{j \,|\, i \in I_j} \frac{m_j}{|I_j|_m} - 1 \Big) = \sum_{j=1}^{N} \frac{m_j}{|I_j|_m} \sum_{i \in I_j} m_i - M = \sum_{j=1}^{N} \frac{m_j}{|I_j|_m} |I_j|_m - M = 0.$$

Lastly, for Model 4,

$$\sum_{i=1}^{N} \dot{m}_i = \sum_{i=1}^{N} m_i \left(\frac{1}{M} \sum_{j=1}^{N} m_j \frac{e_{ji}}{\frac{1}{M} \sum_{k=1}^{N} m_k e_{jk}} - 1 \right) = \sum_{j=1}^{N} m_j \frac{\sum_{i=1}^{N} m_i e_{ji}}{\sum_{k=1}^{N} m_k e_{jk}} - M = 0.$$

We show that although the four specific models defined in Section 3.8.1 have generally discontinuous mass dynamics, some restrictions render ψ continuous, and we can therefore consider their unique solutions in the classical sense.

Proposition 3 *Let (x, m) be a Filippov solution to system (M1), with $\psi_i(x, m) = \frac{m_i}{M} \sum_{j=1}^{N} m_j \langle \frac{\dot{x}_i + \dot{x}_j}{2}, u_{ji} \rangle$. If for all $(i, j) \in \{1, \cdots, N\}^2$, $x_i^0 \neq x_j^0$, then along the dynamics, ψ_i is continuous with respect to x and m and (x, m) is the unique solution of the Cauchy problem in the classical sense.*

Proof 25 *The only discontinuity of ψ occurs when $x_i = x_j$ for some $(i, j) \in \{1, \cdots, N\}^2$. However, if initially $x_i^0 \neq x_j^0$, for all $(i, j) \in \{1, \cdots, N\}^2$, we have $x_i(t) \neq x_j(t)$ for all $t \geq 0$. Hence along the dynamics, ψ is locally Lipschitz with respect to x and m and each solution (x, m) is continuously differentiable and satisfies (M1) for all $t \in \mathbb{R}$.*

Proposition 4 *Let (x, m) be a Filippov solution to system* (M2), *with $\psi_i(x, m) = \frac{m_i}{M}\left(p_i(\frac{1}{M}\sum_{j=1}^{M} m_j p_j)^{-1} - 1\right)$, and suppose that $a(\cdot)$ is bounded away from zero. Then ψ_i is only discontinuous at consensus, i.e. when $x_i = x_j$ for all $(i, j) \in \{1, \cdots, N\}^2$. If for all $(i, j) \in \{1, \cdots, N\}^2$, $x_i^0 \neq x_j^0$, then along the dynamics, ψ_i is continuous with respect to x and m, and (x, m) is the unique solution of the Cauchy problem in the classical sense.*

Proof 26 *The only possible point of discontinuity is at (x, m) such that $\sum_{j=1}^{M} m_j p_j = 0$. From Proposition 2, there exists at least one $i \in \{1, \cdots, N\}$ such that $m_i \neq 0$. Then since for all $j \in \{1, \cdots, N\}$, $m_j \geq 0$ and $p_j \geq 0$, we have $p_i = 0$. Recall that the positional influence p_i is defined by $p_i = \sum_{j=1}^{N} a(\|x_i - x_j\|)\|x_i - x_j\|$. Hence $p_i = 0$ if and only if for all $j \in \{1, \cdots, N\}$, $\|x_i - x_j\| = 0$, that is at a consensus point. We now show that indeed ψ_i is discontinuous around consensus. Consider two sequences (x, m) and $(\tilde{x}, \tilde{m})$ converging to consensus while satisfying the following:*

$$\begin{cases} x_1(t) \to_{t\to\infty} 0 \\ x_i(t) = 0 \quad \text{for all } i \geq 2 \end{cases} \quad \text{and} \quad \begin{cases} \tilde{x}_2(t) = -\tilde{x}_1(t) \to_{t\to\infty} 0 \\ \tilde{x}_i(t) = 0 \quad \text{for all } i \geq 3. \end{cases}$$

We compute for the first and the second sequence respectively:

$$\begin{cases} p_1 = (N-1)a(\|x_1\|)\|x_1\| \\ p_i = a(\|x_1\|)\|x_1\|, \; i \geq 2 \end{cases} \quad \begin{cases} \tilde{p}_1 = \tilde{p}_2 = (N-2)a(\|\tilde{x}_1\|)\|\tilde{x}_1\| + 2a(2\|\tilde{x}_1\|)\|\tilde{x}_1\| \\ \tilde{p}_i = 2a(\|\tilde{x}_1\|)\|\tilde{x}_1\|, \; i \geq 3 \end{cases}$$

Then for the first sequence, for all $i \geq 2$,

$$\frac{p_i}{\sum_{j=1}^{N} m_j p_j} = \frac{a(\|x_1\|)\|x_1\|}{(M - m_1)a(\|x_1\|)\|x_1\| + m_1(N-1)a(\|x_1\|)\|x_1\|} = \frac{1}{M + (N-2)m_1}.$$

For the second sequence, for all $i \geq 3$,

$$\frac{\tilde{p}_i}{\sum\limits_{j=1}^{N} \tilde{m}_j \tilde{p}_j} = \frac{1}{(M - \tilde{m}_{12}) + \tilde{m}_{12}\left(\frac{N-2}{2} + \frac{a(2\|\tilde{x}_1\|)}{a(\|\tilde{x}_1\|)}\right)} = \frac{1}{M + \tilde{m}_{12}\left(\frac{N-4}{2} + \frac{a(2\|\tilde{x}_1\|)}{a(\|\tilde{x}_1\|)}\right)},$$

where we denoted by $\tilde{m}_{12}$ the sum $\tilde{m}_1 + \tilde{m}_2$. These quotients are in general different, hence $\psi_i(x, m)$ and $\psi_i(\tilde{x}, \tilde{m})$ converge to different limits as x and $\tilde{x}$ converge to consensus. Therefore ψ is discontinuous for all (x, m) with x at consensus. However, according to Proposition 2, if initially $x_i^0 \neq x_j^0$, for all $(i, j) \in \{1, \cdots, N\}^2$, we have $x_i(t) \neq x_j(t)$ for all $t \geq 0$. Hence along the dynamics, ψ is locally Lipschitz with respect to x and m and each solution (x, m) is continuously differentiable and satisfies (M1) in the classical sense. It is the unique solution to the Cauchy problem associated with (M2).

Proposition 5 *Let (x, m) be a Filippov solution to system (M4), with $\psi_i(x, m) = m_i \left(\sum_{j=1}^{M} \frac{m_j}{M} e_{ji} \left(\frac{1}{M} \sum_{k=1}^{M} m_k e_{jk}\right)^{-1} - 1 \right)$, and suppose that $a(\cdot)$ is bounded away from zero. If for all $(i, j) \in \{1, \cdots, N\}^2$, $x_i^0 \neq x_j^0$, then along the dynamics, ψ_i is continuous with respect to x and m, and (x, m) is the unique solution of the Cauchy problem in the classical sense.*

Proof 27 *The only possible points of discontinuity are the points (x, m) satisfying*

$$\sum_{k=1}^{M} m_k^2 a(\|x_j - x_k\|)\|x_j - x_k\| = 0$$

for some $j \in \{1, \cdots, N\}$. Unlike for Model (M2), as seen in Proposition 4, this can occur for many configurations of the system, since the weights can be equal to zero. However, from Proposition 2, if for all $(i, j) \in \{1, \cdots, N\}^2$, $x_i^0 \neq x_j^0$, then $x_i(t) \neq x_j(t)$ for all $t \geq 0$. Since for all time there exists $k \in \{1, \cdots, N\}$ such that

$m_k(t) \neq 0$, *for all* $t \geq 0$ *we have*

$$\sum_{k=1}^{M} m_k(t)^2 a(\|x_j(t) - x_k(t)\|)\|x_j(t) - x_k(t)\| \neq 0.$$

Remark 10 *We have proven that under the assumption that* a *is bounded below and that the system is initially not clustered, the functions driving the weight dynamics of models* (M1), (M2) *and* (M4) *are continuous along the trajectories of* x *and* m. *We then consider solutions* (x, m) *in the classical sense. This is not the case for model* (M3) *which is by nature discontinuous.*

3.8.3   General properties of the system

Let

$$x_{\text{max}}^0 := \max_{i \in \{1, \cdots, N\}} \|x_i^0\|. \tag{3.57}$$

We show that the system is not expanding in time in the following sense:

Proposition 6 *Let* (x, m) *be a solution to the general system* $(3.52) - (3.53) - (3.54)$. *For all* $t \in \mathbb{R}^+$,

$$\max_{i \in \{1, \cdots, N\}} \|x_i(t)\| \leq x_{\text{max}}^0. \tag{3.58}$$

Proof 28 *We show that* $\max_{i \in \{1, \cdots, N\}}(\|x_i(t)\|)$ *decreases with respect to time. For all* $i \in \{1, \cdots, N\}$, *if* $\|x_i(t)\| \neq 0$,

$$\begin{aligned}
\frac{d}{dt}\|x_i\| &= \frac{1}{2\|x_i\|} \frac{d}{dt}(\|x_i\|^2) \\
&= \frac{1}{\|x_i\|} \frac{1}{M} \sum_{j=1}^{N} m_j a(\|x_i - x_j\|)\langle x_i, x_j - x_i \rangle \\
&= \frac{1}{\|x_i\|} \frac{1}{M} \left(\sum_{j=1}^{N} m_j a(\|x_i - x_j\|)\langle x_i, x_j \rangle - \sum_{j=1}^{N} m_j a(\|x_i - x_j\|)\|x_i\|^2 \right).
\end{aligned} \tag{3.59}$$

Let $i \in \{1, \cdots, N\}$ *such that* $\|x_i(t)\| = \max_{j \in \{1, \cdots, N\}} \|x_j(t)\|$. *If* $\|x_i(t)\| = 0$, *then*

(3.51) *is trivial and* $\|x_i(t')\| = 0$ *for all* $t' \geq t$. *If* $\|x_i(t)\| \neq 0$, *then from* (3.59), $\frac{d}{dt}\|x_i(t)\| \leq 0$, *which proves* (3.58).

An important consequence of the Lipschitz property of $s \mapsto a(\|s\|)s$ is that a solution to (3.51) cannot form clusters in finite time, as we show in the following proposition.

Proposition 7 *Let* (x,m) *be a solution to system* (3.52)$-$(3.53)$-$(3.54), *with initial condition* $x(0) = x^0$. *Then there exists* $K \geq 0$ *such that for all* $(i,j) \in \{1,\cdots,N\}^2$,

$$\|x_i(t) - x_j(t)\| \geq \|x_i(0) - x_j(0)\|e^{-Kt}.$$

Proof 29 *From Proposition 6, for all* $t \in \mathbb{R}^+$, *for all* $(i,j,k) \in \{1,\cdots,N\}^3$, $\|(x_k - x_i) - (x_k - x_j)\|(t) = \|x_i - x_j\|(t) \leq 2x_{\max}^0$. *Then from Hypothesis 1, there exists* $K \geq 0$ *(depending on* $x_{\max}^0$*) such that*

$$\frac{d}{dt}\|x_i - x_j\|^2 = \frac{2}{M}\sum_{k=1}^{N} m_k \langle x_i - x_j, a(\|x_k - x_i\|)(x_k - x_i) - a(\|x_k - x_j\|)(x_k - x_j)\rangle$$

$$\geq -\frac{2}{M}\sum_{k=1}^{N} m_k \|x_i - x_j\|\|a(\|x_k - x_i\|)(x_k - x_i) - a(\|x_k - x_j\|)(x_k - x_j)\|$$

$$\geq -\frac{2}{M}\sum_{k=1}^{N} m_k \|x_i - x_j\|K\|(x_k - x_i) - (x_k - x_j)\| = -2K\|x_i - x_j\|^2$$

and we conclude with Gronwall's lemma.

We now examine the existence of solutions in the case of the four specific models (M1), (M2), (M3) and (M4). Let $\overline{x} := \frac{1}{N}\sum_{i=1}^{N} x_i$ denote the average group opinion. We also define $\overline{X} := \frac{1}{M}\sum_{i=1}^{N} m_i x_i$ the weighted average opinion. Recall that in the classical opinion dynamics model (3.50), the average opinion $\overline{x}$ is constant with respect to time.

Proposition 8 *For the augmented model with varying weights* $(3.52)-(3.53)-(3.54)$, *if the weights do not evolve in time (i.e. $\psi_i \equiv 0$ for all $i \in \{1, \cdots, N\}$), then the weighted average $\overline{X} = \frac{1}{M} \sum_{i=1}^{N} m_i x_i$ is constant with respect to time.*

Proof 30 *If $\dot{m}_i = 0$ for all $i \in \{1, \cdots, N\}$,*

$$\frac{d}{dt}\overline{X} = \frac{1}{M} \sum_{i=1}^{N} m_i \dot{x}_i = \frac{1}{M} \sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j a(\|x_i - x_j\|)(x_j - x_i) = 0$$

by antisymmetry of the summed coefficients.

We now investigate the propensity of the system to form consensus. Let us suppose that for all $(i,j) \in \{1, \cdots, N\}^2$, $\|x_i^0 - x_j^0\| > 0$. From Proposition 7, for all $(i,j) \in \{1, \cdots, N\}^2$, $\|x_i(t) - x_j(t)\| \geq \|x_i(0) - x_j(0)\|e^{-Kt}$. From here onward we denote by

$$d_{\min}(t) := \min_{(i,j)\in\{1,\cdots,N\}^2} \|x_i(0) - x_j(0)\|e^{-Kt} \tag{3.60}$$

the lower bound on the smallest distance between two agents at time t, *i.e.* for all $(i,j) \in \{1, \cdots, N\}^2$, $\|x_i(t) - x_j(t)\| \geq d_{\min}(t)$. Furthermore, according to Proposition 6, $\max_{i\in\{1,\cdots,N\}} \|x_i(t)\| \leq x_{\max}^0$. Then for all $t \in \mathbb{R}^+$, $0 < d_{\min}(t) \leq \|x_i(t) - x_j(t)\| \leq 2x_{\max}^0$. Let us define the function $t \mapsto \alpha(t)$ as

$$\alpha(t) = \sup_{z\in[d_{\min}(t),2x_{\max}^0]} a(z). \tag{3.61}$$

For all $t \geq 0$, we then define a normalized interaction function $z \mapsto a_t(z)$ such that:

$$a_t(z) := \frac{a(z)}{\alpha(t)}. \tag{3.62}$$

This normalized interaction function now satisfies the property $0 \leq a_t(\|x_i(t) - x_j(t)\|) \leq 1$ for all $t \geq 0$ and all $(i,j) \in \{1, \cdots, N\}^2$. This will allow us to rewrite the

system with a row-stochastic matrix $(A_{ij})_{(i,j)\in\{1,\cdots,N\}^2}$ defined by:

$$
\begin{cases}
A_{ij}(t) = \frac{m_j(t)}{M} a_t(\|x_i(t) - x_j(t)\|) & \text{if } i \neq j \\
A_{ii}(t) = 1 - \sum_{j\neq i} \frac{m_j(t)}{M} a_t(\|x_i(t) - x_j(t)\|).
\end{cases}
\tag{3.63}
$$

Then for all $t \geq 0$, for all $(i,j) \in \{1, \cdots, N\}^2$, $0 \leq A_{ij}(t) \leq 1$. Moreover, we have constructed A in such a way that for all $i \in \{1, \cdots, N\}$, $\sum_{j=1}^{N} A_{ij}(t) = 1$, making it row-stochastic as claimed above.

Then denoting by $x(t)$ the opinion vector $(x_i)_{i\in\{1,\cdots,N\}}$, the opinion dynamics of system (3.51) can be rewritten in matrix form as:

$$
\dot{x}(t) = \alpha(t) \left(A(t)x(t) - x(t) \right).
\tag{3.64}
$$

Let $\Omega(t)$ denote the convex hull of $x(t)$, defined as follows.

Definition 14 *Let $(x_i)_{i\in\{1,\cdots,N\}}$. Its convex hull Ω is defined by:*

$$
\Omega = \left\{ \sum_{i=1}^{N} \xi_i x_i \mid \forall i \in \{1, \cdots, N\}, 0 \leq \xi_i \leq 1 \text{ and } \sum_{i=1}^{N} \xi_i = 1 \right\}.
$$

We show that $\Omega(t)$ contracts in time with the following proposition.

Proposition 9 *The convex hull $\Omega(t)$ of $x(t)$ is contracting in time in the sense that*

$$
\Omega(t_2) \subseteq \Omega(t_1) \quad \text{for all } t_2 \geq t_1 \geq 0.
\tag{3.65}
$$

This implies that there exists a convex compact Ω_∞ such that

$$
\lim_{t\to+\infty} \Omega(t) = \Omega_\infty.
$$

Proof 31 *From Equation (3.64), we can rewrite the system as:*

$$\dot{x}_i(t) = \alpha(t)(\overline{x}_i(t) - x_i(t)), \quad i \in \{1, \cdots, N\}, \tag{3.66}$$

where $\overline{x}_i := \sum_{j=1}^{N} A_{ij} x_j$ is a convex combination of the agents x_j. Then, by definition, $\overline{x}_i$ is inside the convex hull of x, and $\dot{x}_i$ is directed towards the convex hull, which ensures the property (3.65).

Theorem 8 *Let $D : t \mapsto D(t) := \max_{(i,j) \in \{1, \cdots, N\}^2} \|x_i(t) - x_j(t)\|$ represent the diameter of the system. Then it satisfies the concentration estimate*

$$\frac{d}{dt}D(t) \leq -\alpha(t)\eta(t)D(t), \tag{3.67}$$

with

$$\alpha(t) = \sup\{a(z) \mid z \in [d_{min}(t), 2x_{\max}^0]\} \quad and$$
$$\eta(t) := \min_{(i,j) \in \{1, \cdots, N\}^2} \sum_{k=1}^{N} \min\{A_{ik}, A_{jk}\} \tag{3.68}$$

for the row-stochastic matrix $A(t)$ defined in equation (3.63). and for the quantities $x_{\max}^0$ and d_{min} respectively defined in equations (3.57) and (3.60).

In particular, if $\lim_{t \to +\infty} \int_0^t \alpha(s)\eta(s)ds = +\infty$, then $\lim_{t \to +\infty} D(t) = 0$. Furthermore, if $\int_0^{+\infty} \alpha(t) \exp\left(-\int_0^t \alpha(s)\eta(s)ds\right) dt < \infty$, then all the agents converge to an emerging consensus x^∞ such that

$$\|x_i(t) - x^\infty\| \leq \exp\left(-\int_0^t \alpha(s)\eta(s)ds\right) D(0). \tag{3.69}$$

Proof 32 *After writing the system in matrix form (3.64), where A is row-stochastic, the proof is an easy adaptation of the proof of Theorem 2.2 of Motsch and Tadmor[56].*

Let us first show that

$$\max_{(i,j)\in\{1,\cdots,N\}^2} \|(A(t)x(t))_i - (A(t)x(t))_j\| \leq (1-\eta(t))D(t). \qquad (3.70)$$

Let $(i,j) \in \{1, \cdots, N\}^2$ and let $\eta_k^{ij} := \min\{A_{ij}, A_{jk}\}$. We omit the time t for convenience of notation and compute:

$$\|(Ax)_i - (Ax)_j\| =$$

$$\left\|\sum_{k=1}^{N} A_{ij}x_k - \sum_{k=1}^{N} A_{jk}x_k\right\| = \left\|\sum_{k=1}^{N}(A_{ij} - \eta_k^{ij})x_k - \sum_{k=1}^{N}(A_{jk} - \eta_k^{ij})x_k\right\|$$

$$\leq \sum_{k=1}^{N}(A_{ij} - \eta_k^{ij}) \max_{k\in\{1,\cdots,N\}} \|x_k\| - \sum_{k=1}^{N}(A_{jk} - \eta_k^{ij}) \min_{k\in\{1,\cdots,N\}} \|x_k\|$$

$$\leq (1 - \sum_{k=1}^{N}\eta_k^{ij}) \left(\max_{k\in\{1,\cdots,N\}} \|x_k\| - \min_{k\in\{1,\cdots,N\}} \|x_k\|\right)$$

$$\leq (1 - \eta) \max_{(i,j)\in\{1,\cdots,N\}^2)} \|x_i - x_j\| = (1 - \eta)D.$$

Let us now compute from (3.64):

$$\frac{d}{dt}\|x_i - x_j\|^2 = 2\left\langle x_i - x_j, \alpha\left(\sum_{k=1}^{N}A_{ik}x_k - x_i\right) - \alpha\left(\sum_{k=1}^{N}A_{jk}x_k - x_j\right)\right\rangle$$

$$\leq 2\alpha\left(-\|x_i - x_j\|^2 + \|(Ax)_i - (Ax)_j\|\|x_i - x_j\|\right)$$

$$\leq 2\alpha\left(-\|x_i - x_j\|^2 + (1 - \eta)D\|x_i - x_j\|\right)$$

Let $(i,j) \in \{1, \cdots, N\}^2$ such that $\|x_i - x_j\| = D$. Then

$$\frac{d}{dt}D(t)^2 \leq 2\alpha(t)\left(-D(t)^2 + (1 - \eta(t))D(t)^2\right) = -2\alpha(t)\eta(t)D(t)^2,$$

which implies (3.67). It follows that $D(t) = D(0)\exp(-\int_0^t \alpha(s)\eta(s)ds)$, so

$$\lim_{t\to+\infty}\int_0^t \alpha(s)\eta(s)ds = +\infty \quad \Rightarrow \quad \lim_{t\to+\infty} D(t) = 0.$$

Then the diameter of the system converges to zero. We further show that each agent converges to consensus. Let $i \in \{1, \cdots, N\}$, and let $0 \le t_1 < t_2$.

$$\|x_i(t_2) - x_i(t_1)\| = \left\| \int_{t_1}^{t_2} \dot{x}_i dt \right\| = \left\| \int_{t_1}^{t_2} \alpha(t) \left((Ax)_i(t) - x_i(t) \right) dt \right\|$$

$$\le \int_{t_1}^{t_2} \alpha(t) \max_{(i,j) \in \{1, \cdots, N\}^2} \|x_i(t) - x_j(t)\| dt = \int_{t_1}^{t_2} \alpha(t) D(t)$$

$$\le D(0) \int_{t_1}^{t_2} \alpha(t) \exp\left(- \int_0^t \alpha(s)\eta(s)ds \right) dt$$

Hence,

$$\lim_{\tau \to +\infty} \int_0^\tau \alpha(t) \exp\left(- \int_0^t \alpha(s)\eta(s)ds \right) dt < +\infty \Rightarrow$$

$$\lim_{|t_1 - t_2| \to 0} \|x_i(t_2) - x_i(t_1)\| = 0, \tag{3.71}$$

which implies the convergence of x_i to a consensus opinion x^∞ satisfying equation (3.69).

3.8.4 Consensus in the case of a positive interaction function

We now show that if the interaction function a is bounded below on the initial diameter of the system, then the system converges to consensus.

Proposition 10 *Let (x, m) be a solution to $(3.52)-(3.53)-(3.54)$ and let $t \mapsto D(t)$ be the diameter of the system as defined in Theorem 8. If $\inf\{a(z) \,|z \le D(0)\} := a_{min} > 0$ then the system (3.51) converges to consensus, with the rate*

$$D(t) \le D(0)e^{-a_{min}t}$$

Proof 33 *Due to the assumption on a, at $t = 0$, since for all $(i, j) \in \{1, \cdots, N\}^2$, $\|x_i^0 - x_j^0\| \le D(0)$, we have $a(\|x_i^0 - x_j^0\|) \ge a_{min}$. According to Theorem 8, for all*

$t \in \mathbb{R}^+$, $D(t) \leq D(0)$, so for all $t \in \mathbb{R}^+$, for all $(i,j) \in \{1, \cdots, N\}^2$, $a(\|x_i(t) - x_j(t)\|) \geq a_{min}$. Consider $(i,j) \in \{1, \cdots, N\}^2$ such that $D(t) = \|x_i(t) - x_j(t)\| = \max_{(k,l) \in \{1, \cdots, N\}^2} \|x_k(t) - x_l(t)\|$. We compute $\frac{d}{dt} D(t)^2 = \frac{d}{dt} \|x_i(t) - x_j(t)\|^2$:

$$\frac{d}{dt} \left(\|x_i - x_j\|^2 \right) = \frac{2}{M} \sum_{k=1}^{N} m_k a(\|x_i - x_k\|) \langle x_k - x_i, x_i - x_j \rangle$$
$$- \frac{2}{M} \sum_{k=1}^{N} m_k a(\|x_j - x_k\|) \langle x_k - x_j, x_i - x_j \rangle.$$

Because $\|x_i(t) - x_j(t)\| = D(t)$, for all $(k,l) \in \{1, \cdots, N\}^2$, $\langle x_k - x_i, x_i - x_j \rangle \leq 0$ and symmetrically, $\langle x_k - x_j, x_i - x_j \rangle \geq 0$. Consequently,

$$\frac{d}{dt} \left(\|x_i - x_j\|^2 \right) \leq$$
$$\frac{2}{M} \sum_{k=1}^{N} m_k a_{min} \langle x_k - x_i, x_i - x_j \rangle - \frac{2}{M} \sum_{k=1}^{N} m_k a_{min} \langle x_k - x_j, x_i - x_j \rangle$$
$$\leq -\frac{2}{M} a_{min} \sum_{k=1}^{N} m_k \|x_i - x_j\|^2 \leq -2 a_{min} \|x_i - x_j\|^2.$$

Then $D(t) \leq D(0) e^{-a_{min} t}$, which implies that the system converges to consensus.

Remark 11 *The condition on the boundedness away from zero of the interaction function does not prevent the effective interactions $m_k a(\|x_i - x_k\|)$ from being zero, as the individual weights m_k can tend to 0. However, due to the positivity of the total mass of the system $M > 0$, the directed graph of interactions remains weakly connected and consensus is still achieved.*

3.8.5 Clustering in the case of an interaction function with bounded support

In the previous section, we saw that a positive interaction function leads to consensus of the system. We now explore the behavior of the system in the case of an interaction

function with bounded support, *i.e.* we suppose the following:

Hypothesis 3 *Suppose that there exists $R \geq 0$ such that a satisfies*

$$\begin{cases} a(r) > 0 \ \text{for all } r \in (0, R) \\ a(r) = 0 \ \text{for all } r \geq R. \end{cases} \tag{3.72}$$

In the case of the classical Hegselmann-Krause model (3.50), this was shown to lead to clustering equilibria[57], *i.e.* states of the system in which for all $(i, j) \in \{1, \cdots, N\}^2$, either $\|x_i - x_j\| = 0$ or $\|x_i - x_j\| \geq R$, where R denotes the size of the support of the interaction function. In this section, we prove that in the case of non-evolving positive weights, we obtain the same result. The case of non-evolving positive weights is covered by the framework of type-symmetric and cut-balanced systems, and convergence was shown by Hendrickx and Tsitsiklis[58]. Cut-balanced systems model situations in which if a group of agents influences another group, then the first group is also influenced by the second one by at least a proportional amount. In this paper, we consider the opposite situation, by allowing a subgroup of agents to become the only ones with influence. Our system is cut-balanced only when the weights are bounded away from zero. We saw in Section 3.8.4 that as long as the total mass M is positive, if the interaction function is positive, the system converges to consensus, even if some weights vanish. We now examine the case of an interaction function with bounded support. We show that we can adapt the work of Jabin and Motsch[57] to prove the clustering of the system, with the strong condition that the weights are positive and do not evolve. We also show that when the masses evolve and are allowed to vanish, we cannot expect the results to hold.

Let the energy of the system be defined by:

$$E(x) = \frac{1}{M^2} \sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j \Phi(\|x_i - x_j\|^2), \quad \Phi : r \mapsto \int_0^r a(\sqrt{s})ds. \tag{3.73}$$

Proposition 11 *Consider the simplified system* $(3.52)-(3.53)-(3.54)$ *with constant masses, i.e.* $\Psi_i \equiv 0$ *for all* $i \in \{1, \cdots, N\}$. *Then the energy decreases in time with:*

$$\frac{dE}{dt} = -\frac{4}{M} \sum_{i=1}^{N} m_i \|\dot{x}_i\|^2 \tag{3.74}$$

and

$$\frac{dE}{dt} \leq -\frac{1}{M^3} \frac{\left(\sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j a(\|x_i - x_j\|) \|x_i - x_j\|^2 \right)^2}{\sum_{i=1}^{N} m_i \|x_i\|^2}. \tag{3.75}$$

Furthermore, the following inequality follows:

$$\int_0^{+\infty} m_i^2 m_j^2 a^2(\|x_i(t) - x_j(t)\|) \|x_i(t) - x_j(t)\|^4 dt \leq C_M \tag{3.76}$$

where C_M *is a constant depending only on the total mass* M.

Proof 34 *For the equality* (3.74), *we compute the time derivative of the energy functional when the masses are constant in time:*

$$\frac{dE}{dt} = \frac{1}{M^2} \sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j a(\|x_i - x_j\|) 2 \langle x_i - x_j, \dot{x}_i - \dot{x}_j \rangle$$

$$= \frac{4}{M^2} \sum_{i=1}^{N} \sum_{j=1}^{N} m_i m_j a(\|x_i - x_j\|) \langle x_i - x_j, \dot{x}_i \rangle = -\frac{4}{M} \sum_{i=1}^{N} m_i \|\dot{x}_i\|^2.$$

Furthermore, from the Cauchy-Schwarz inequality,

$$\left(\sum_{i=1}^{N} m_i \langle \dot{x}_i, x_i \rangle \right)^2 \leq \left(\sum_{i=1}^{N} m_i \|\dot{x}_i\|^2 \right) \left(\sum_{i=1}^{N} m_i \|x_i\|^2 \right) = -\frac{M}{4} \frac{dE}{dt} \left(\sum_{i=1}^{N} m_i \|x_i\|^2 \right),$$

from which we get

$$\left(\sum_{i=1}^{N} \frac{1}{M} \sum_{j=1}^{N} m_i m_j a(\|x_i - x_j\|) \langle x_j - x_i, x_i \rangle \right)^2 \leq -\frac{M}{4} \frac{dE}{dt} \left(\sum_{i=1}^{N} m_i \|x_i\|^2 \right),$$

so

$$\left(\frac{1}{2M}\sum_{i=1}^{N}\sum_{j=1}^{N}m_im_ja(\|x_i-x_j\|)\|x_i-x_j\|^2\right)^2 \leq \frac{M}{4}\frac{dE}{dt}\left(\sum_{i=1}^{N}m_i\|x_i\|^2\right),$$

and this gives the desired inequality (3.75). Lastly, from Proposition 6, the opinion vector x is bounded at all time, so there exists C_M such that

$$\int_0^{+\infty}\sum_{i=1}^{N}\sum_{j=1}^{N}m_i^2m_j^2a^2(\|x_i(t)-x_j(t)\|)\|x_i(t)-x_j(t)\|^4dt \leq C_M,$$

which implies the inequality (3.76).

Proposition 11 implies that the stationary states S of the system are the states satisfying $\frac{dE}{dt}=0$, *i.e.* if a satisfies Hypothesis 3,

$$S = \{x \in (\mathbb{R}^d)^N \mid \forall (k,l) \in \{1,\cdots,N\}^2,\ m_km_l = 0 \text{ or } \|x_k-x_l\| = 0 \text{ or } \|x_k-x_l\| \geq R\}.$$

We now prove that if the masses are all strictly positive, the system converges to an equilibrium satisfying:

$$\forall (k,l) \in \{1,\cdots,N\}^2,\ \|x_k - x_l\| = 0 \text{ or } \|x_k - x_l\| \geq R.$$

Proposition 12 *Suppose that $a(\cdot)$ satisfies Hypothesis 3. Suppose that $\psi_i \equiv 0$ and $m_i > 0$ for all $i \in \{1,\cdots,N\}$. Then for all $\epsilon > 0$, there exists T_ϵ such that for all $(i,j) \in \{1,\cdots,N\}^2$, for all $t \geq T_\epsilon$, $\|x_i(t) - x_j(t)\| \in [0,\epsilon] \cup [R-\epsilon,+\infty)$.*

Proof 35 *By contradiction, suppose that there exists $(i,j) \in \{1,\cdots,N\}$ and $\{t_n\}_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$, $\|x_i(t_n) - x_j(t_n)\| \in (\epsilon, R - \epsilon)$. From the hypotheses on $a(\cdot)$, there exists $\delta > 0$ such that for all $r \in (\epsilon, R - \epsilon)$, $a(r) \geq \delta$. Furthermore, since from Theorem 8, for all $t \geq 0$, $\|x_i(t) - x_j(t)\| \leq D(0)$, from Hypothesis 1 there exists*

$C > 0$ *such that for all* $t \geq 0$, $a(\|x_i - x_j\|)\|x_i - x_j\| \leq C$. *Then*

$$\|\dot{x}_i\| \leq \frac{1}{M}\sum_{j=1}^{N} m_j a(\|x_i - x_j\|)\|x_i - x_j\| \leq \frac{1}{M}\sum_{j=1}^{N} m_j C \leq C,$$

so x is uniformly continuous in time. This implies that there exists $\tau > 0$ such that $a(\|x_i(t) - x_j(t)\|) \geq \frac{\delta}{2}$ *and* $\|x_i(t) - x_j(t)\| \geq \frac{\epsilon}{2}$ *for all* $r \in [t_n, t_n + \tau]$, *for all* $n \in \mathbb{N}$. *Then*

$$\lim_{T \to +\infty} \int_0^T m_i^2 m_j^2 a(\|x_i(t) - x_j(t)\|)^2 \|x_i(t) - x_j(t)\|^4 dt$$

$$\geq \lim_{k \to +\infty} \sum_{n=1}^{k} \int_{t_n}^{t_n+\tau} m_i^2 m_j^2 a(\|x_i(t) - x_j(t)\|)^2 \|x_i(t) - x_j(t)\|^4 dt$$

$$\geq \lim_{k \to +\infty} \sum_{n=1}^{k} \tau \frac{\delta^2}{4} \frac{\epsilon^4}{16} m_{\min}^4 = +\infty$$

where $m_{\min} := \min_{i \in \{1,\cdots,N\}}\{m_i\}$. *This contradicts the inequality* (3.76).

The next corollary follows automatically from Proposition 12:

Corollary 3 *If a satisfies* (3.72), *and* $\psi_i \equiv 0$, $m_i > 0$ *for all* $i \in \{1, \cdots, N\}$, *the system forms clusters separated by a distance greater than R, i.e. for any* $(i,j) \in \{1, \cdots, N\}^2$,

$$\lim_{t \to +\infty} \|x_i(t) - x_j(t)\| = 0 \quad or \quad \lim_{t \to +\infty} \|x_i(t) - x_j(t)\| \geq R. \qquad (3.77)$$

Corollary 1 implies that the solution to (3.51) with constant positive weights forms clusters, but it is not enough in order to prove its convergence. In dimension 1, the convergence of the system is a direct consequence of the preserved order stated in Proposition 2, of the clustering property stated in Corollary 3 and of the convergence of the convex hull Ω stated in Proposition 9. The proof is identical to that of Theorem 4 in Jabin and Motsch[57], hence we do not provide the details. Instead, we focus on

the multidimensional case, which is more complicated due to the fact that lost connectivity between two agents can be recovered at a later time. To prove convergence of the solution to system (3.51) with constant positive weights, we follow the method introduced therein[57]. We start by refining the result of Proposition 11:

Proposition 13 *Let (x, m) be a solution to the simplified system $(3.52)-(3.53)-(3.54)$ with constant masses, i.e. $\Psi_i \equiv 0$ for all $i \in \{1, \cdots, N\}$. For all $(i, j) \in \{1, \cdots, N\}^2$, for all $T \geq 0$,*

$$\int_0^T m_i m_j a(\|x_i(t) - x_j(t)\|) \|x_i(t) - x_j(t)\|^2 dt \leq \frac{M^2}{4} (x_{\max}^0)^2. \qquad (3.78)$$

Proof 36 *We start by writing (we omit (t) notation of function x for brevity):*

$$\int_0^T \sum_{i=1}^N m_i \langle x_i, \dot{x}_i \rangle dt = \int_0^T \sum_{i=1}^N \sum_{j=1}^N \frac{m_i m_j}{M} a(\|x_i - x_j\|) \langle x_i, x_j - x_j \rangle dt.$$

Then we have:

$$\int_0^T \sum_{i=1}^N \sum_{j=1}^N \frac{m_i m_j}{M} a(\|x_i - x_j\|) \|x_i, x_j\|^2 dt = -\frac{1}{4} \int_0^T \sum_{i=1}^N m_i \frac{d}{dt} \|x_i(t)\|^2 dt.$$

The right-hand side can be written as:

$$-\frac{1}{4} \int_0^T \sum_{i=1}^N m_i \frac{d}{dt} \|x_i\|^2 dt = -\frac{1}{4} \left[\sum_{i=1}^N m_i \|x_i\|^2 \right]_0^T = \frac{M}{4} ((x_{\max}^0)^2 - (x_{\max}^T)^2),$$

where we used the boundedness of $\|x_i\|$ proven in Proposition 6. Inequality (3.78) follows from the non-negativity of each term.

We are now equipped to prove the convergence theorem. From Corollary 3, we know that the agents form clusters separated by distances at least equal to R, the support of the interaction function.

Theorem 9 *Let (x, m) be a solution to $(3.52)-(3.53)-(3.54)$. Suppose that $a(\cdot)$ satisfies Hypothesis 3, and $\psi_i \equiv 0$, $m_i > 0$ for all $i \in \{1, \cdots, N\}$. Let $N_C \leq N$ denote the number of clusters formed by the system. We denote by $\mathbb{C}_k$ the set of agents forming the $k-$th cluster, by $M_k := \sum_{i \in \mathbb{C}_k} m_i$ the mass of $\mathbb{C}_k$, and by $y_k := \frac{1}{M_k} \sum_{i \in \mathbb{C}_k} m_i x_i$ its barycenter. Then there exists a constant C such that for each $k \in \{1, ..., N_C\}$, for all $T \geq 0$, the barycenter of $\mathbb{C}_k$ satisfies:*

$$\int_0^T \|\dot{y}_k(t)\| dt \leq C. \tag{3.79}$$

Hence the system (3.51) with constant weights converges as $t \to +\infty$.

Proof 37 *Let $k \in \{1, ..., N_C\}$. We compute the velocity of the center of mass y_k of the cluster $\mathbb{C}_k$:*

$$\dot{y}_k = \frac{1}{M_k} \sum_{i \in \mathbb{C}_k} m_i \dot{x}_i$$

$$= \frac{1}{M_k} \sum_{i \in \mathbb{C}_k} \left[\sum_{j \in \mathbb{C}_k} \frac{m_i m_j}{M} a(\|x_i - x_j\|)(x_j - x_i) + \sum_{j \notin \mathbb{C}_k} \frac{m_i m_j}{M} a(\|x_i - x_j\|)(x_j - x_i) \right].$$

The first term vanishes by antisymmetry of the summed coefficients. For the second one, we sum over $i \in \mathbb{C}_k$ and $j \notin \mathbb{C}_k$. Then as a consequence of Corollary 3, there exists $T' \geq 0$ such that for all $t \geq T'$, $\|x_i - x_j\| \geq \frac{R}{2}$. Hence for all $t \geq T'$, $\frac{1}{R}\|x_i - x_j\| \leq \frac{2}{R^2}\|x_i - x_j\|^2$. If $T' \leq T$, the integral between T' and T can be bounded as follows:

$$\int_{T'}^T \sum_{i \in \mathbb{C}_k} \sum_{j \notin \mathbb{C}_k} \frac{m_i m_j}{M M_k} a(\|x_i(t) - x_j(t)\|)\|x_j(t) - x_i(t)\| dt$$

$$\leq \int_{T'}^T \sum_{i \in \mathbb{C}_k} \sum_{j \notin \mathbb{C}_k} \frac{m_i m_j}{M M_k} a(\|x_i(t) - x_j(t)\|)\frac{2}{R}\|x_j(t) - x_i(t)\|^2 dt \leq \frac{M^2}{2R}(x_{\max}^0)^2$$

where we used the inequality (3.78). Furthermore, as a consequence of Hypothesis 1

and of Theorem 8, there exists $C' > 0$ such that for all $t \geq 0$, $a(\|x_i(t) - x_j(t)\|)\|x_j(t) - x_i(t)\| \leq C'$.

$$\int_0^{T'} \sum_{i \in \mathbb{C}_k} \sum_{j \notin \mathbb{C}_k} \frac{m_i m_j}{MM_k} a(\|x_i(t) - x_j(t)\|)\|x_j(t) - x_i(t)\| dt \leq T'C'.$$

Then (3.79) follows with $C := \frac{M^2}{2R}(x_{\max}^0)^2 + T'C'$ if $T \geq T'$, and $C := T'C'$ if $T' \geq T$. We deduce that the barycenters y_k of the clusters converge as $t \to +\infty$, and from Corollary 3, this implies that x_i converges for each $i \in \{1, \cdots, N\}$.

Remark 12 *The proof of Theorem 8 follows closely the proof of Proposition 3.2 in Jabin and Motsch[57]. The main difference is that we prove the convergence of the barycenters of each cluster, instead of their centers. This allows us to recover a symmetry property that greatly simplifies the computation of $\dot{y}_k$.*

Hence the system with non-evolving positive masses behaves in the same way as the non-weighted dynamics[57]. However, as soon as one mass vanishes, we may observe clusters separated by any distance. The following simple 1-dimensional example shows such a case of equilibrium at a distance smaller than R.

Example 1 *Let (x, m) be a solution to $(3.52)-(3.53)-(3.54)$ with $d = 1$ and $N = 3$. Suppose that $a(\cdot)$ satisfies Hypothesis 3, and $|x_1 - x_2| = R$, with $m_3 = 0$. Then $\dot{x}_1 = \dot{x}_2 = 0$, and the system is at equilibrium if $\dot{m}_1 = \dot{m}_2 = \dot{m}_3 = \dot{x}_3 = 0$, i.e. if $\psi_1(x, m) = \psi_2(x, m) = \psi_3(x, m) = 0$ and*

$$m_1 a(|x_1 - x_3|)(x_1 - x_3) + m_2 a(|x_2 - x_3|)(x_2 - x_3) = 0.$$

Simplifying even further, if we consider Model (M1) and if we suppose $a(r) = 1$ for all $r \in [0, R)$, the system is at equilibrium if

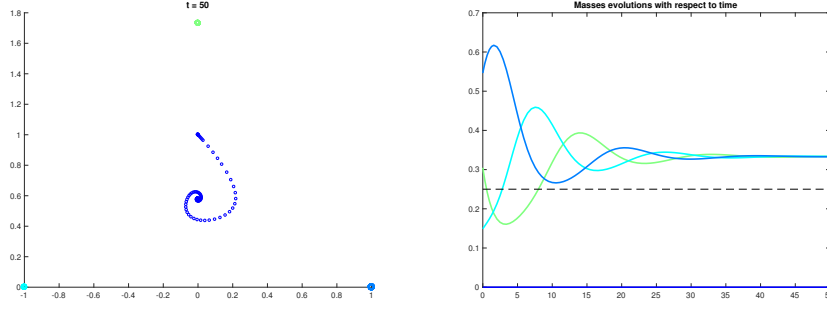$$\frac{|x_2 - x_3|}{|x_1 - x_3|} = \frac{m_1}{m_2}.$$

Figure 3.22: Evolution of the system corresponding to Example 2. Left: Evolution of the positions $\{x_i\}_{i\in\{1,...,4\}}$. Right: Evolution of the masses $\{m_i\}_{i\in\{1,...,4\}}$. The dashed line shows the constant average mass of the system.

Simulations in the case of time-evolving masses show other cases of equilibria with clusters separated by distances smaller than R.

Example 2 *Let $d = 2$ and $N = 4$. Suppose that $a(r) = 1$ for $r \in [0, \frac{3}{2})$ and 0 otherwise. We place three agents at the vertices of an equilateral triangle of side equal to 2, with $x_1 = (0, \sqrt{3})$, $x_2 = (-1, 0)$ and $x_3 = (1, 0)$. We suppose that $m_4 = 0$ and we simulate the evolution of the system with the mass dynamics given by Model (M2). The positions of the the first three agents stay invariant in time with $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$, and the mass of the fourth agent remains equal to 0. Then the only variable quantities are x_4, and the three masses m_1, m_2 and m_3. Simulations show that if $\|x_4 - x_i\| < R$ is satisfied for all $i \in \{1, ..., 3\}$, the system tends to an equilibrium given by $m_1 = m_2 = m_3 = \frac{1}{3}$ and $x_4 = (\frac{1}{\sqrt{3}}, 0)$ being the center of the circumscribing circle of (x_1, x_2, x_3), see Figure 3.22. Furthermore, although the system converges to equilibrium, we observe oscillations both in the mass evolution and in the spiral trajectory of x_4.*

3.9 Long-term behavior and equilibria for heterophilious dynamics

We now focus on the specific case of a constant interaction function, for example $a \equiv 1$. This models the case of heterophilious dynamics, as the influence of one agent over another grows with the distance separating them[56]. We use this case of simple interaction function to illustrate some fundamental differences between the four models (M1), (M2), (M3) and (M4), more specifically in the long-term behavior of the weights and in the existence of equilibria.

Proposition 14 *Suppose that the interaction function is constant, $a \equiv 1$. Then the solution to system $(3.52)-(3.53)-(3.54)$ converges exponentially to consensus, and we have:*

$$\frac{d}{dt}D(t) \leq D(0)e^{-t}. \tag{3.80}$$

Furthermore, the system converges to the consensus opinion x^∞ and

$$\forall i \in \{1, \cdots, N\}, \quad \|x_i(t) - x^\infty\| \leq e^{-t}D(0). \tag{3.81}$$

Lastly,

$$\forall (i,j) \in \{1, \cdots, N\}^2, \quad x_i(t) - x_j(t) = e^{-t}(x_i^0 - x_j^0). \tag{3.82}$$

Proof 38 *With $a \equiv 1$, we have $\alpha \equiv 1$ and the opinion dynamics can be written in matrix form (3.64) with the row-stochastic matrix A defined as:*

$$\begin{cases} A_{ij}(t) = \frac{m_j(t)}{M} & \text{if } i \neq j \\ A_{ii}(t) = 1 - \sum_{j \neq i} \frac{m_j(t)}{M}. \end{cases}$$

We apply Theorem 8 and calculate:

$$\sum_{k=1}^{N} \min\{A_{ik}, A_{jk}\}$$

$$= \sum_{k \neq i,j}^{N} \min\{\frac{m_k(t)}{M}, \frac{m_k(t)}{M}\}+$$

$$\min\{1 - \sum_{k \neq i} \frac{m_k(t)}{M}, \frac{m_i(t)}{M}\} + \min\{1 - \sum_{k \neq j} \frac{m_k(t)}{M}, \frac{m_j(t)}{M}\}$$

$$= \sum_{k \neq i,j}^{N} \frac{m_k(t)}{M} + \frac{m_i(t)}{M} + \frac{m_j(t)}{M} = 1.$$

Then $\eta(t) = \min_{(i,j) \in \{1,\cdots,N\}^2} 1 = 1$, and we get (3.80) from (3.67). Furthermore,

$$\int_0^{+\infty} \exp\left(-\int_0^t \eta(s)ds\right) dt = \int_0^{+\infty} e^{-t}dt < \infty,$$

so according to Theorem 8, all the agents converge to the consensus x^{∞} with the rate given by (3.81). Lastly, we compute explicitly:

$$\dot{x}_i(t) = \frac{1}{M} \sum_{j=1}^{N} m_j(t) \ (x_j(t) - x_i(t)) = (\overline{X}(t) - x_i(t)), \quad i \in \{1, \cdots, N\},$$

which gives: $\dot{x}_i(t) - \dot{x}_j(t) = -(x_i(t) - x_j(t))$ and (3.82) follows.

Remark 13 *For the classical weightless opinion dynamics (3.50) with constant interaction function $a \equiv 1$, the system converges exponentially to consensus at the value $\overline{x}$ with the same rate:*

$$\forall i \in \{1, \cdots, N\}, \quad x_i(t) = \overline{x} + e^{-t}(x_i^0 - \overline{x}).$$

This is to be expected since the classical dynamics (3.50) are a special case of system (3.52)$-$(3.53)$-$(3.54) with equal and constant weights.

We now further study the evolution of the weight distribution, and exhibit funda-

mental differences in the behaviors of the three models (M1), (M2) and (M3).

Proposition 15 *Consider Model* (M1), *with constant interaction function* $a \equiv 1$. *Suppose that for all* $i \in \{1, \cdots, N\}$, $m_i(0) = m_i^0 > 0$. *Then there exist* $(\mu_i)_{i \in \{1, \cdots, N\}}$ *positive constants such that for all* $t \geq 0$ *for all* $i \in \{1, \cdots, N\}$ *the weights satisfy:* $m_i(t) \geq \mu_i > 0$.

Proof 39 *From* (M1), *the weights satisfy for all* $i \in \{1, \cdots, N\}$:

$$\dot{m}_i = \frac{m_i}{M} \sum_{j=1}^{N} m_j \left\langle \frac{\dot{x}_i + \dot{x}_j}{2}, \frac{x_i - x_j}{\|x_i - x_j\|} \right\rangle.$$

From Proposition 14,

$$\left\| \frac{\dot{x}_i + \dot{x}_j}{2} \right\| = \left\| \overline{X} - \frac{x_i + x_j}{2} \right\| \leq D(t) \leq D(0)e^{-t}.$$

From this we get:

$$\dot{m}_i \geq -\frac{m_i}{M} \sum_{j=1}^{N} m_j \left\| \frac{\dot{x}_i + \dot{x}_j}{2} \right\| \geq -\frac{m_i}{M} \sum_{j=1}^{N} m_j D(0)e^{-t} = -m_i e^{-t} D(0).$$

Integrating with respect to time yields:

$$\ln \left(\frac{m_i(t)}{m_i^0} \right) \geq - \int_0^t e^{-s} D(0) ds$$

from which we obtain:

$$m_i(t) \geq m_i^0 e^{D(0)(e^{-t}-1)} \geq m_i^0 e^{-D(0)} > 0.$$

This result shows that in the case of Model (M1), there is no saturation behavior: if the weights are initially all positive, they remain bounded away from zero at all time, which implies that they remain well-distributed among the agents. This is

illustrated in the numerical simulations presented in Figures 3.23 and 3.27.

We now show that we have the opposite behavior in the case of Model (M2).

Proposition 16 *Consider Model* (M2) *with the heterophilious interaction* $a \equiv 1$*. Let* $p_i^0 = p_i(0) = \sum_{j=1}^N \|x_i^0 - x_j^0\|$*. Suppose that for all* $(i, j) \in \{1, \cdots, N\}^2$*,* $p_i^0 \neq p_j^0$*, and let* $k := \text{argmax}_{i \in \{1, \cdots, N\}} p_i^0$*. Then*

$$\lim_{t \to \infty} m_k(t) = M \quad and \quad \forall i \neq k, \lim_{t \to \infty} m_i(t) = 0. \tag{3.83}$$

Proof 40 *From Proposition 14, for all* $i \in \{1, \cdots, N\}$*:*

$$p_i(t) = \sum_{j=1}^N \|x_i(t) - x_j(t)\| = e^{-t} \sum_{j=1}^N \|x_i^0 - x_j^0\| = e^{-t} p_i^0.$$

Then from (M2),

$$\dot{m}_i(t) = \frac{m_i(t)}{M} \left(\frac{p_i(t)}{\frac{1}{M} \sum_{j=1}^N m_j(t) p_j(t)} - 1 \right) = \frac{m_i(t)}{M} \left(\frac{p_i^0}{\frac{1}{M} \sum_{j=1}^N m_j(t) p_j^0} - 1 \right).$$

Let $k := \text{argmax}_{i \in \{1, \cdots, N\}} p_i^0$*. Then* $p_k^0 \geq \frac{1}{M} \sum_{j=1}^N m_j(t) p_j^0$*, so* $\dot{m}_k(t) \geq 0$ *for all* $t \geq 0$*. Since* $m_k(t) \leq M$*, there exists* $\overline{m}_k$ *with* $0 \leq \overline{m}_k \leq M$ *such that* $\lim_{t \to \infty} m_k(t) = \overline{m}_k$ *and* $m_k(t) \leq \overline{m}_k$ *for all* $t \geq 0$*. Suppose that* $\overline{m}_k < M$*. More specifically, suppose that there exists* $\delta > 0$ *such that* $\overline{m}_k = M - \delta$*. There also exists* $\epsilon > 0$ *such that for all* $i \neq k$*,* $p_i^0 \leq p_k^0 - \epsilon$*. We then write:*

$$\begin{aligned}
\frac{\frac{1}{M} \sum_{j=1}^N m_j(t) p_j^0}{p_k^0} &= \frac{1}{p_k^0 M} \left(m_k p_k^0 + \sum_{j \neq k} m_j p_j^0 \right) \leq \frac{m_k}{M} + \frac{(M - m_k)(p_k^0 - \epsilon)}{M p_k^0} \\
&\leq \frac{m_k}{M} \left(1 - \frac{p_k^0 - \epsilon}{p_k^0} \right) + \frac{p_k^0 - \epsilon}{p_k^0} \leq \frac{M - \delta}{M} \left(1 - \frac{p_k^0 - \epsilon}{p_k^0} \right) + \frac{p_k^0 - \epsilon}{p_k^0} \\
&\leq 1 - \frac{\delta}{M} \left(1 - \frac{p_k^0 - \epsilon}{p_k^0} \right) = 1 - \frac{\delta \epsilon}{M p_k^0}.
\end{aligned}$$

Let $C := (\delta\epsilon)/(Mp_k^0) > 0$. We choose ϵ small enough that $C > 1$. Then

$$\frac{p_k^0}{\frac{1}{M}\sum_{j=1}^{N} m_j(t)p_j^0} \geq \frac{1}{1-C} = 1 + \frac{C}{1-C} > 1,$$

and $\dot{m}_k$ satisfies:

$$\dot{m}_k \geq \frac{m_k}{M}(1 + \frac{C}{1-C} - 1) = \frac{m_k}{M}\frac{C}{1-C}.$$

By Gronwall's lemma, for all $t \geq 0$, $m_k(t) \geq m_k^0 e^{\frac{C}{M(1-C)}t}$, which contradicts $m_k(t) \leq M$. Hence $\overline{m}_k = M$ and because $\sum_{i=1}^{N} m_i = M$, for all $i \neq k$, $\lim_{t\to\infty} m_i(t) = 0$.

Hence in the case of Model (M2), the agent with the largest initial positional influence p_k^0 becomes the leader: its weight tends to the total mass of the system M, while the weights of all the other agents tend to 0. This is illustrated in Figures 3.24 and 3.28.

Models (M3) and (M4) exhibit a third behavior, different from those of Models (M1) and (M2): starting from random initial conditions, two agents gain influence while all the other weights go to zero. We prove that indeed the state $m = (\frac{M}{2}, \frac{M}{2}, 0, ..., 0)$ is an equilibrium for the weights.

Proposition 17 *In models (M3) and (M4), any state satisfying*

$$\begin{cases} m_I = m_J = \dfrac{M}{2} \\ m_i = 0 \quad \text{for all } i \neq I, J \end{cases} \tag{3.84}$$

for any indices $I \neq J$, is an equilibrium for the weight dynamics, in the sense that $\dot{m}_i = 0$ for all $i \in \{1, \cdots, N\}$.

Proof 41 *Without loss of generality, we reorder the agents so that $I = 1$ and $J = 2$. For Model (M3), notice that for all $i \in \{3, ..., N\}$, $\dot{m}_i = 0$. Furthermore, $I_1 = \{2\}$*

and $I_2 = \{1\}$, so from (M3),

$$
\begin{cases}
\dot{m}_1 = m_1(\frac{m_2}{m_1} - 1) = m_2 - m_1 = 0 \\
\dot{m}_2 = m_2(\frac{m_1}{m_2} - 1) = m_1 - m_2 = 0.
\end{cases}
$$

For Model (M4), *again $\dot{m}_i = 0$ for all $i \in \{3, ..., N\}$. The agents with non-zero masses satisfy:*

$$
\dot{m}_1 = m_1 \left(\frac{m_2}{M} \frac{m_1 a(\|x_1 - x_2\|)\|x_1 - x_2\|}{\frac{m_1^2}{M} a(\|x_1 - x_2\|)\|x_1 - x_2\|} - 1 \right) = m_2 - m_1 = 0,
$$

and symmetrically, $\dot{m}_2 = 0$.

This is illustrated in Figures 3.25 and 3.29.

3.10   Numerical simulations

3.10.1   Behavior of the four specific models with constant interaction function

We start by illustrating the behavior of each of the four models for $d = 1$. Figures 3.23, 3.24, 3.25 and 3.26 show the evolution of a group of 10 agents in $\mathbb{R}$, with the same initial conditions. The interaction function was chosen to be $a \equiv 1$. Notice that in all cases, the order of the opinions is preserved, as shown in Proposition 2. It is not the case for the order of the weights. As shown in Section 3.8.4, in all cases, the system converges to consensus asymptotically. In each of the figures 3.23 to 3.30, the plot on the left represents the evolution of the opinions, each represented by a different color. The width of each curve is proportional to the agent's weight. For the one-dimensional plots (*i.e.* Figures 3.23 to 3.26), the dotted line represents the

average position and the dashed line represents the weighted barycenter. The plot on the right shows the evolution of the corresponding weights, with the dashed line representing the average weight (constant by Proposition 2). Figure 3.23 illustrates the behavior of Model (M1) proven in Proposition 15: the weights vary little and stay bounded away from 0. Figure 3.24 illustrates the behavior of Model (M2) proven in Proposition 16: one agent eventually gains the total mass of the system, and becomes leader. Notice that consensus is achieved at a different value than in Figure 3.23, due to the fact that one agent drags the group towards itself. Figures 3.25 and 3.26 illustrate the behaviors of Models (M3) and (M4), proven in Proposition 17: two agents eventually share equally the total mass of the system, and become co-leaders. Interestingly, with the same initial conditions, the two leaders are different in Model (M3) and Model (M4).

Figures 3.27, 3.28, 3.29 and 3.30 show the evolution of a group of 10 agents in $\mathbb{R}^2$, with the same initial conditions. The interaction function was also chosen to be $a \equiv 1$. Notice that with Model (M2), consensus is reached before the leader gains the total mass of the system (Figure 3.28). Hence it does not have time to influence the position at which consensus is achieved. As in 1D, with Models (M3) and (M4), two agents become co-leaders, but they are different for each of the models (see Figures 3.29 and 3.30).
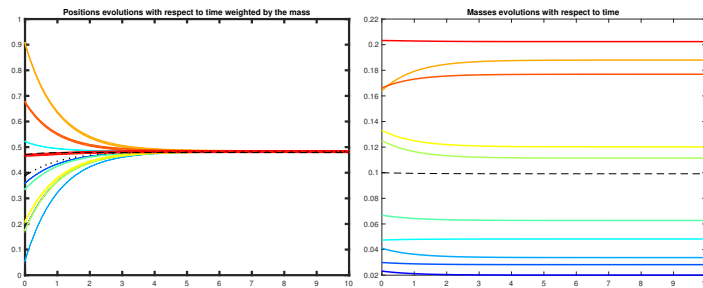

Figure 3.23: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M1).
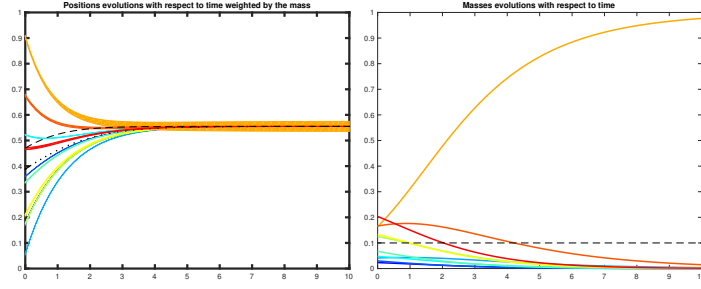
Figure 3.24: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M2).

Figure 3.25: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M3).

Figure 3.26: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M4).
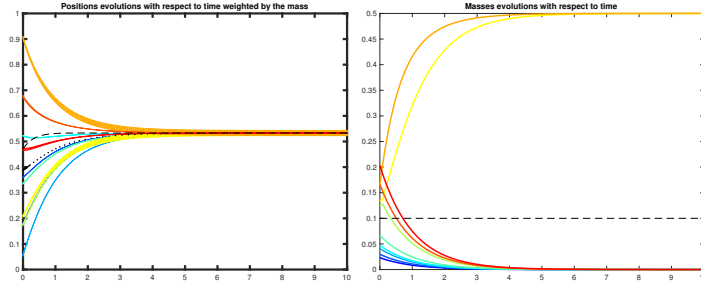

Figure 3.27: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}^2$ with (M1).
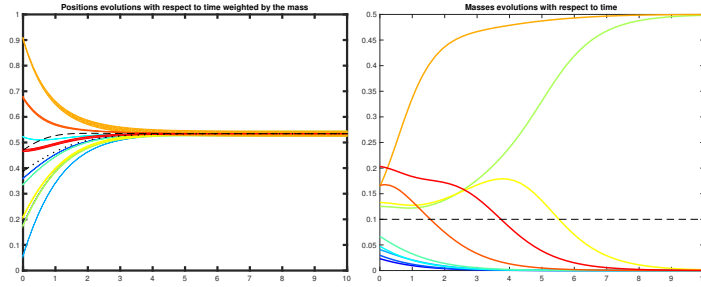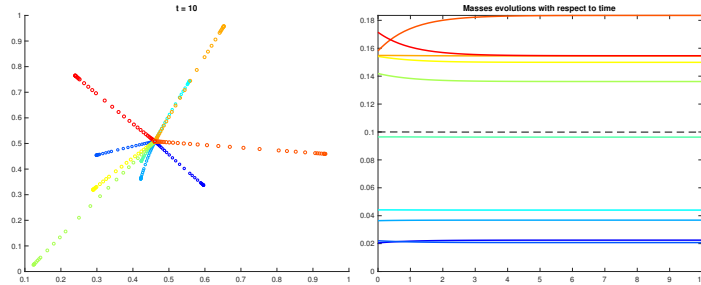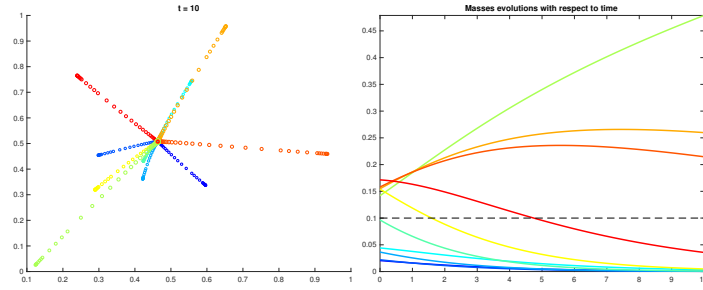
Figure 3.28: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}^2$ with (M2).
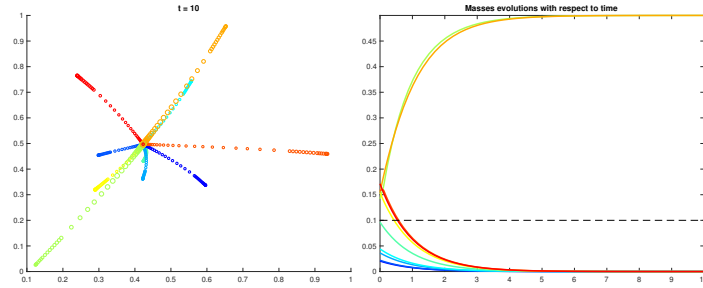
Figure 3.29: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}^2$ with (M3).
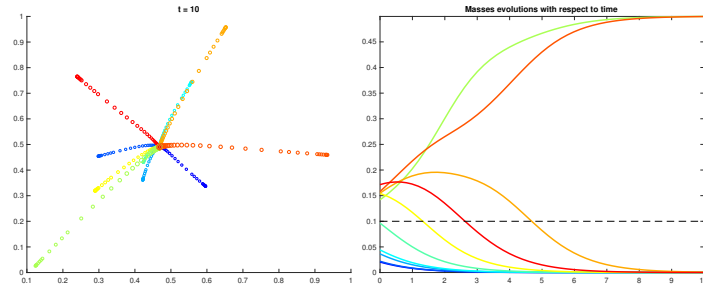
Figure 3.30: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}^2$ with (M4).

3.10.2 Behavior of selected models with bounded confidence

We now provide simulations for an interaction function with bounded support. In Section 3.8.5, we showed that for a model with constant weights, the system forms clusters separated by at least the distance of the support of the interaction function. We also showed through simple examples that the same behavior cannot be expected if the weights are allowed to evolve in time, and more specifically, to vanish. We now show general simulations in 1D exhibiting the behavior of the system when a satisfies

Hypothesis 3 with $R = 0.2$.

Figure 3.31 shows the evolution of the opinions and weights for Model (M2). We observe cluster formation, with clusters separated by a distance greater than R, the support of a. Interestingly, whereas the opinions quickly reach an equilibrium, the weights keep oscillating between clusters (in the opinions plots, the width of each curve is proportional to the agent's weight). This is due to the fact that the weight dynamics allow exchange of mass between clusters (the denominator in (M2) sums over all agents, regardless of the distances between opinions). The weighted barycenter is shown to oscillate (dashed line), whereas the average opinion converges to an equilibrium (dotted line).

Figure 3.32 shows the evolution of the opinions and weights for Model (M3). Recall that with a constant interaction function, the system was shown to converge to an equilibrium in which two agents share equally the total mass of the system. This behavior is observed again here, but now at the level of each individual cluster. Figure 3.32 shows a situation where the system converges to three clusters. The one that converge at position $x = 0.84$ is composed of three agents, but its mass is only shared between the green and the yellow agents, as shown by the weights' evolution. Similarly, the second cluster (at $x = 0.4$) is composed of five agents, but only the green and the orange agents gain the total mass of the cluster. Lastly, the third cluster is only composed of two agents (red and blue) that share the mass equally.

3.10.3   Speed of convergence

In this section, we compare the speed of convergence to consensus in each of the four models. We define the "dispersion" over time by:

$$\Delta(t) := \frac{1}{N^2} \sum_{i,j=1}^{N} \|x_i(t) - x_j(t)\|^2.$$

Figure 3.31: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M2), a satisfying Hyp. 3 with $R = 0.2$.
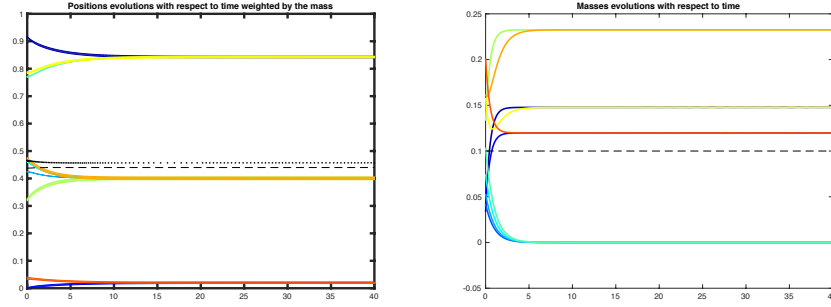
Figure 3.32: Evolution of opinions (left) and weights (right) of 10 agents in $\mathbb{R}$ with (M3), a satisfying Hyp. 3 with $R = 0.2$.
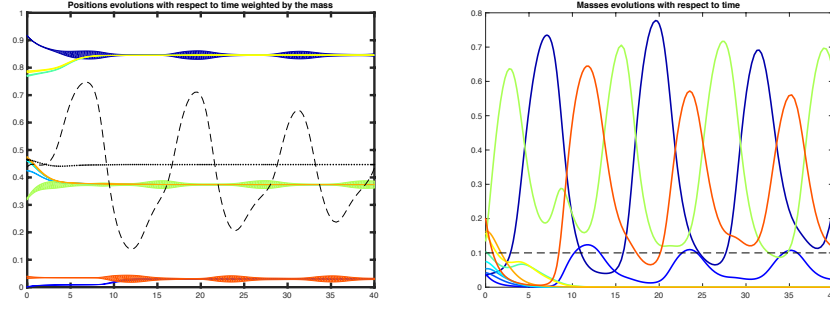
Simulations were run with $N = 30$ agents and $d = 3$ to test the effect of mass dynamics on speed of convergence to consensus. We used the interaction function $a(\cdot) \equiv 1$. Recall that the system converges asymptotically to consensus, as seen in Propositions 10 and 14. We fix a final time $T = 10$ and we consider that the system has "reached consensus" if $\Delta(T) < \varepsilon = 10^{-6}$. We denote by ω the number of times that consensus was reached out of 100 simulations in 3 dimensions.

Two series of simulations were run. In the first one, whose results are shown in Table 3.2, all 30 initial weights were sampled randomly from the uniform distribution in $[0, 1]$, and then normalized so that $M = \sum_{i=1}^{30} m_i = 1$. The 30 initial positions were sampled randomly from the uniform distribution in $[0, 1]^3$.

In the second series, whose results are shown in Table 3.3, all 30 initial weights were

Random initial mass	ω	τ	$\overline{\Delta}(10)$
M1	100	6.6580	3.2003×10^{-8}
M2	100	6.8082	3.5569×10^{-8}
M3	100	6.5051	2.7298×10^{-8}
M4	100	6.4685	2.6024×10^{-8}
Model-NMD	100	6.6801	3.2298×10^{-8}

Table 3.2: Speed of convergence with randomly generated weights

Equal initial mass	ω	τ	$\overline{\Delta}(10)$
M1	100	6.6996	3.2798×10^{-8}
M2	100	6.8559	3.7337×10^{-8}
M3	100	6.6664	3.2040×10^{-8}
M4	100	6.5616	2.8759×10^{-8}
Model-NMD	100	6.6886	3.2346×10^{-8}

Table 3.3: Speed of convergence with equal weights

set to be equal to $m_i = 1/30$, thus still satisfying $M = 1$. The 30 initial positions were sampled randomly from the uniform distribution in $[0, 1]^3$.

We denote by τ the average time to consensus given that the system achieved it. Additionally, $\overline{\Delta}(10)$ represents the mean dispersion at the end of the simulation ($T = 10$) for all 100 simulations. As a benchmark for time to convergence, "Model NMD" shows statistics for the model with no mass dynamics. For Model (M1) the time to consensus is very similar to Model-NMD. This is expected because Model (M1) shows only weak mass dynamics in the sense that masses do not change dramatically from their initial values, see also simulations shown in Figures 3.23 and 3.27. Model (M2) shows greater average time to consensus. This suggests that having a leader who is the only one to have influence on the group is not optimal. Interestingly, Models (M3) and (M4) converge the fastest, which suggests that a system guided by two leaders converges more efficiently to consensus than a system guided by one unique leader, and than a system with no leader.

3.10.4 Clustering of bounded confidence models with time-varying mass

We now present simulations comparing the effect of each model's mass dynamics on the clustering behavior of the system. Ten simulations of each model (M1), (M2), (M3), (M4) and (M-NMD) were run with $N = 30$ agents in three spatial dimensions ($d = 3$). As in Section 3.10.3, M-NMD refers to the model with no mass dynamics, *i.e.* $\Psi \equiv 0$. In order to observe clustering, we chose for these simulations an interaction function $a(\cdot)$ with bounded support, satisfying Hypothesis 3 with an interaction radius set to $R = 0.4$. More specifically,

$$\begin{cases} a(r) = 1 \text{ for all } r \in (0, R) \\ a(r) = 0 \text{ for all } r \geq R. \end{cases}$$

As in Section 3.10.3, we first generated 10 sets of initial conditions (IC0 through IC9) with randomly generated opinions in $[0, 1]^3$ and randomly generated masses in $[0, 1]$, satisfying $M = \sum_{i=1}^{30} m_i = 1$. The resulting number of clusters for each model is shown in Table 3.4, where the rightmost column shows for each model the mean number of clusters $\bar{c}$ of all 10 simulations. As expected, Model (M1) has a similar behavior to M-NMD because the mass dynamics are weak compared to those of (M2), (M3), and (M4). It is notable that M-NMD forms the least number of clusters on average. Mass seems to increase the speed at which nearby agents move toward another, and this would cause a cluster to form such that the less massive agents move toward more massive ones. These less massive agents would not have the attraction to draw the clusters together. An important note about (M3) and (M4) is that a high number of clusters arises. This can be explained by the fact that pairs of co-leading agents drain the masses of all the other ones, as shown in Figure 3.32. This affects clustering by leaving many singleton clusters of mass-less agents. Even if

these agents are relatively close to each other (their distance being possibly less than the interaction radius R), they will have no attractive effect on any agent.

	IC0	IC1	IC2	IC3	IC4	IC5	IC6	IC7	IC8	IC9	$\bar{c}$
M1	5	4	5	6	8	5	5	4	4	2	4.8
M2	5	4	7	6	6	5	5	5	4	5	5.2
M3	11	9	8	12	7	9	10	7	6	8	8.7
M4	8	6	7	9	7	9	6	5	5	7	6.9
M-NMD	5	3	5	6	7	5	4	4	4	3	4.6

Table 3.4: Clustering comparison of the five models with randomly generated weights

For comparison, we ran a corresponding set of ten simulations (IP0 through IP9) with initial opinions chosen identically to those of simulations IC0 through IC9, but with initially equally distributed masses $m_i = \frac{1}{N}$ for all $i \in \{1, \cdots, N\}$. In this case, model M-NMD is exactly the classical HK model (3.50). The clustering results are displayed in Table 3.5. Several observations can be made by comparing the two tables. Firstly, as mentioned previously, the equal distribution of weights tends to reduce the number of clusters: for Models (M1), (M2), (M3) and M-NMD, the average number of clusters is lower with initial conditions IP0-IP9 than with initial conditions IC0-IC9. Moreover, observe that with initially equal masses, Model (M1) and Model M-NMD's performances are equivalent: even though they achieve different numbers of clusters for each simulation, their average is the same.

Comparing these results with those of Section 3.10.3, it is interesting to note that with a positive interaction function, Models (M3) and (M4) seem to achieve faster convergence to consensus, whereas in the case of bounded confidence, they result in greater clustering of the system. These first results might suggest ways to design efficient control strategies to either drive the system to consensus, or on the contrary to avoid clustering.

	IP0	IP1	IP2	IP3	IP4	IP5	IP6	IP7	IP8	IP9	$\bar{c}$
M1	5	5	5	6	3	5	3	3	4	3	4.2
M2	5	5	6	5	4	5	3	5	5	4	4.7
M3	10	6	6	11	7	8	8	6	5	9	7.6
M4	10	6	7	12	8	7	5	7	4	6	7.2
M-NMD	5	5	5	6	3	4	3	3	5	3	4.2

Table 3.5: Clustering comparison of the five models with initially equal weights

3.11 Appendix

An example using Approach A which shows unexpected behavior in the first example (A.1), as well as the interactions matrix and initial positions used for simulations shown in Figure 3.7 and 3.9.

Example 3.11.1 *15 agents with a general interaction matrix A. We use the same notion of a general interactions matrix as used in [34]. The interactions matrix A is composed of integers a_{ij} that are uniformly chosen between -5 and 5 inclusive. $\Psi \equiv \mathrm{Id}$. Generally, a system with this kind of interaction matrix will exhibit simple oscillating kinetic energy, as in [34]. We show a rare simulation using this general interactions matrix A in the appendix (Figures 3.33 and 3.34).*

$$A = \begin{pmatrix} -5 & -1 & 1 & 0 & 4 \\ -1 & -5 & 1 & 1 & 0 \\ -2 & 0 & -5 & -1 & -2 \\ -4 & -1 & -5 & 3 & -4 \\ -4 & 1 & -2 & 2 & 1 \end{pmatrix} \qquad X = \begin{pmatrix} 1.5755 & 1.7399 \\ 1.6523 & 0.5619 \\ 5.3026 & 2.7008 \\ 2.4971 & 0.7288 \\ 0.6571 & 0.5281 \\ 1.2180 & 0.0840 \\ 2.2812 & 1.0129 \\ 5.4949 & 1.7441 \\ 3.7685 & 2.5903 \\ 1.6218 & 2.5266 \\ 2.2521 & 0.0767 \\ 5.5766 & 1.1671 \\ 5.6582 & 1.5453 \\ 2.8146 & 1.4641 \\ 1.6892 & 0.1310 \end{pmatrix}$$

Example 3.11.2 *Five agents with a general interaction matrix A and Ψ as defined in (3.26). Simulations are shown in Figure 3.7 and 3.9 with $a \in \{\frac{\pi}{4}, \frac{\pi}{2}, \frac{3\pi}{4}\}$.*

$$A = \begin{pmatrix} -5 & -1 & 1 & 0 & 4 \\ -1 & -5 & 1 & 1 & 0 \\ -2 & 0 & -5 & -1 & -2 \\ -4 & -1 & -5 & 3 & -4 \\ -4 & 1 & -2 & 2 & 1 \end{pmatrix} \qquad X = \begin{pmatrix} 6.1743 & 2.8473 \\ 4.5883 & 2.7635 \\ 2.1606 & 2.5691 \\ 3.6698 & 0.8191 \\ 0.6771 & 138672 \end{pmatrix}$$
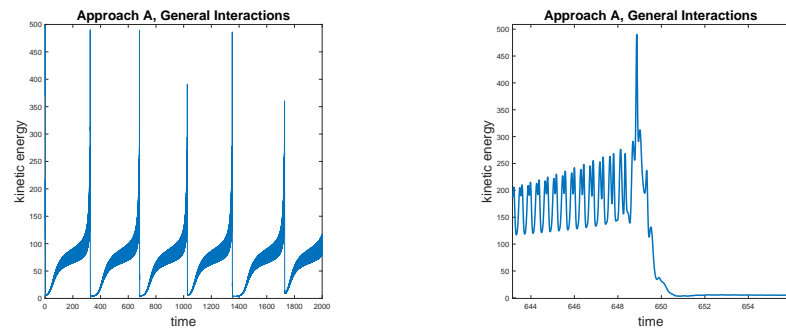
Figure 3.33: Energy of the system using Approach A, 15 agents, and a general interaction matrix (left). A snapshot of the energy oscillations to match with trajectories in Figure 3.34 (right).
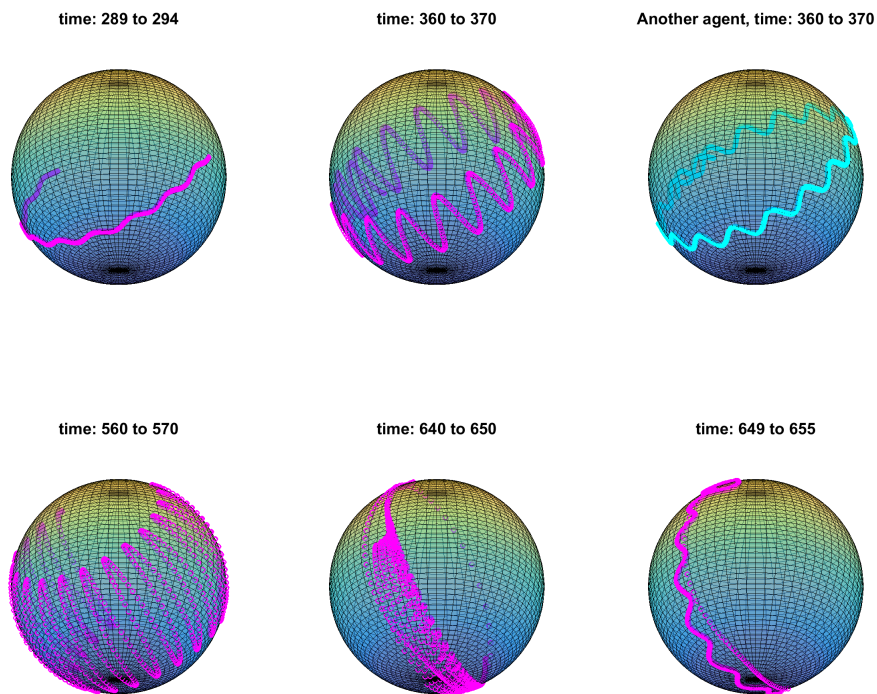

Figure 3.34: An agent's trajectory simulated with Approach A, 15 agents, and a general interaction matrix. The trajectory in shown the top right is of a second agent. The agents oscillate with amplitudes that increase with time, eventually the trajectory approximates a great circle, after which the oscillations resume with smaller amplitudes.

Bibliography

[1] Charles E Ettensohn, Gary M Wessel, and Gregory Wray. *Development of sea urchins, ascidians, and other invertebrate deuterostomes: experimental approaches*, volume 74. Elsevier, 2004.

[2] Rainer Hegselmann and Ulrich Krause. Opinion dynamics and bounded confidence models, analysis, and simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002.

[3] Peter K Sorger, Sandra RB Allerheiligen, Darrell R Abernethy, Russ B Altman, Kim LR Brouwer, Andrea Califano, David Z D'Argenio, Ravi Iyengar, William J Jusko, Richard Lalonde, et al. Quantitative and systems pharmacology in the post-genomic era: new approaches to discovering drugs and understanding therapeutic mechanisms. In *An NIH white paper by the QSP workshop group*, pages 1–48. NIH Bethesda, 2011.

[4] Violeta I Pérez-Nueno. Using quantitative systems pharmacology for novel drug discovery. *Expert opinion on drug discovery*, 10(12):1315–1331, 2015.

[5] SRB Allerheiligen. Next-generation model-based drug discovery and development: quantitative and systems pharmacology. *Clinical Pharmacology & Therapeutics*, 88(1), 2010.

[6] CM Friedrich. A model qualification method for mechanistic physiological qsp models to support model-informed drug development. *CPT: pharmacometrics & systems pharmacology*, 5(2):43–53, 2016.

[7] Inomzhon Mirzaev and Jeremy Gunawardena. Laplacian dynamics on general graphs. *Bulletin of mathematical biology*, 75(11):2118–2149, 2013.

[8] Bernhard Palsson. *Systems biology*. Cambridge university press, 2015.

[9] David J Klinke and Stacey D Finley. Timescale analysis of rule-based biochemical reaction networks. *Biotechnology progress*, 28(1):33–44, 2012.

[10] Jeremy Gunawardena. A linear framework for time-scale separation in nonlinear biochemical systems. *PloS one*, 7(5):e36321, 2012.

[11] Alberto Bressan and Benedetto Piccoli. *Introduction to the mathematical theory of control*, volume 2. American institute of mathematical sciences Springfield, 2007.

[12] K Gadkar, N Budha, A Baruch, JD Davis, P Fielder, and S Ramanujan. A mechanistic systems pharmacology model for prediction of ldl cholesterol lowering by pcsk9 antagonism in human dyslipidemic populations. *CPT: pharmacometrics & systems pharmacology*, 3(11):1–9, 2014.

[13] I Hosseini and F Mac Gabhann. Mechanistic models predict efficacy of ccr5-deficient stem cell transplants in hiv patient populations. *CPT: pharmacometrics & systems pharmacology*, 5(2):82–90, 2016.

[14] RJ Allen, Theodore R Rieger, and Cynthia J Musante. Efficient generation and selection of virtual populations in quantitative systems pharmacology models. *CPT: pharmacometrics & systems pharmacology*, 5(3):140–146, 2016.

[15] Rajan Srinivasan. *Importance sampling: Applications in communications and detection.* Springer Science & Business Media, 2013.

[16] Jefferey E. Ming, Ruth Abrams, Derek W. Barlett, Tu Nguyen, Katherine Kudrycki, Ananth Kadambi, Christina M. Friedrich, Nassim Djebli, Britta Goebel, Joseph Elassal, Poulabi Banerjee, Michael J. Reed, Jeffrey S. Barrett, and Karim Azer. A quantitative systems pharmacology platform to investigate the impact of alirocumab and cholesterol-lowering therapies on lipid profiles and plaque characteristics. *Gene Regulation and Systems Biology*, 2017.

[17] Razvan T Dadu and Christie M Ballantyne. Lipid lowering with pcsk9 inhibitors. *Nature Reviews Cardiology*, 11(10):563–575, 2014.

[18] David J Klinke II. Integrating epidemiological data into a mechanistic model of type 2 diabetes: validating the prevalence of virtual patients. *Annals of biomedical engineering*, 36(2):321–334, 2008.

[19] Brian J Schmidt, Fergal P Casey, Thomas Paterson, and Jason R Chan. Alternate virtual populations elucidate the type i interferon signature predictive of the response to rituximab in rheumatoid arthritis. *BMC bioinformatics*, 14(1):1, 2013.

[20] LB Sheiner and TM Ludden. Population pharmacokinetics/dynamics*. *Annual Review of Pharmacology and Toxicology*, 32(1):185–209, 1992.

[21] Eric H Davidson and D Arendt. The regulatory genome gene regulatory networks in development and evolution. *SCIENCE-NEW YORK THEN WASHINGTON-*, 314(5802):1085, 2006.

[22] Chiou-Hwa Yuh, Hamid Bolouri, and Eric H Davidson. Genomic cis-regulatory logic: experimental and computational analysis of a sea urchin gene. *Science*, 279(5358):1896–1902, 1998.

[23] Jongmin Nam, Ping Dong, Ryan Tarpine, Sorin Istrail, and Eric H Davidson. Functional cis-regulatory genomics for systems biology. *Proceedings of the National Academy of Sciences*, 107(8):3930–3935, 2010.

[24] Jongmin Nam and Eric H Davidson. Barcoded dna-tag reporters for multiplex cis-regulatory analysis. *PLoS One*, 7(4):e35934, 2012.

[25] Alexandre Melnikov, Anand Murugan, Xiaolan Zhang, Tiberiu Tesileanu, Li Wang, Peter Rogov, Soheil Feizi, Andreas Gnirke, Curtis G Callan Jr, Justin B Kinney, et al. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nature biotechnology*, 30(3):271, 2012.

[26] Rupali P Patwardhan, Joseph B Hiatt, Daniela M Witten, Mee J Kim, Robin P Smith, Dalit May, Choli Lee, Jennifer M Andrie, Su-In Lee, Gregory M Cooper, et al. Massively parallel functional dissection of mammalian enhancers in vivo. *Nature biotechnology*, 30(3):265, 2012.

[27] Robin P Smith, Leila Taher, Rupali P Patwardhan, Mee J Kim, Fumitaka Inoue, Jay Shendure, Ivan Ovcharenko, and Nadav Ahituv. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nature genetics*, 45(9):1021, 2013.

[28] Michael A White, Connie A Myers, Joseph C Corbo, and Barak A Cohen. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of chip-seq peaks. *Proceedings of the National Academy of Sciences*, 110(29):11952–11957, 2013.

[29] Cosmas D Arnold, Daniel Gerlach, Christoph Stelzer, Łukasz M Boryń, Martina Rath, and Alexander Stark. Genome-wide quantitative enhancer activity maps identified by starr-seq. *Science*, 339(6123):1074–1077, 2013.

[30] Constantin N Flytzanis, Andrew P McMahon, Barbara R Hough-Evans, Karen S Katula, Roy J Britten, and Eric H Davidson. Persistence and integration of cloned dna in postembryonic sea urchins. *Developmental biology*, 108(2):431–442, 1985.

[31] Constantin N Flytzanis, Roy J Britten, and Eric H Davidson. Ontogenic activation of a fusion gene introduced into sea urchin eggs. *Proceedings of the National Academy of Sciences*, 84(1):151–155, 1987.

[32] Andrew P McMahon, Constantin N Flytzanis, Barbara R Hough-Evans, Karen S Katula, Roy J Britten, and Eric H Davidson. Introduction of cloned dna into sea urchin egg cytoplasm: replication and persistence during embryogenesis. *Developmental biology*, 108(2):420–430, 1985.

[33] Jongmin Nam, Yi-Hsien Su, Pei Yun Lee, Anthony J Robertson, James A Coffman, and Eric H Davidson. Cis-regulatory control of the nodal gene, initiator of

the sea urchin oral ectoderm gene network. *Developmental biology*, 306(2):860–869, 2007.

[34] Marco Caponigro, Anna Chiara Lai, and Benedetto Piccoli. A nonlinear model of opinion formation on the sphere. 2014.

[35] Sebastien Motsch and Eitan Tadmor. Heterophilious dynamics enhances consensus. *SIAM review*, 56(4):577–621, 2014.

[36] S. Y. Ha, T. Ha, and J. H. Kim. Emergent behavior of a Cucker–Smale type particle model with nonlinear velocity couplings. *IEEE Transactions on Automatic Control*, 55(7):1679–1683, July 2010.

[37] Emiliano Cristiani, Paolo Frasca, and Benedetto Piccoli. Effects of anisotropic interactions on the structure of animal groups. *Journal of mathematical biology*, 62(4):569–588, 2011.

[38] Aylin Aydoğdu, Marco Caponigro, Sean McQuade, Benedetto Piccoli, Nastassia Pouradier Duteil, Francesco Rossi, and Emmanuel Trélat. Interaction network, state space and control in social dynamics. In N. Bellomo, P. Degond, and E. Tadmor, editors, *Active Particles Volume 1, Theory, Methods, and Applications*. Birkhauser-Springer, 2017.

[39] Pawel Sobkowicz. Modelling opinion formation with physics tools: Call for closer link with reality. *Journal of Artificial Societies and Social Simulation*, 12(1):11, 2009.

[40] Yoshiki Kuramoto. Cooperative dynamics of oscillator community a study based on lattice of rings. *Progress of Theoretical Physics Supplement*, 79:223–240, 1984.

[41] Steven H. Strogatz. From Kuramoto to Crawford: exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena*, 143(1–4):1 – 20, 2000.

[42] Florian Dörfler, Michael Chertkov, and Francesco Bullo. Synchronization in complex oscillator networks and smart grids. *Proceedings of the National Academy of Sciences*, 110(6):2005–2010, 2013.

[43] L. Scardovi, A. Sarlette, and R. Sepulchre. Synchronization and balancing on the N-torus. *Systems & Control Letters*, 56(5):335 – 341, 2007.

[44] Rodolphe Sepulchre, Derek Paley, Naomi Ehrich Leonard, et al. Stabilization of planar collective motion: All-to-all communication. *Automatic Control, IEEE Transactions on*, 52(5):811–824, 2007.

[45] Rodolphe Sepulchre, Derek Paley, Naomi Ehrich Leonard, et al. Stabilization of planar collective motion with limited communication. *Automatic Control, IEEE Transactions on*, 53(3):706–719, 2008.

[46] A. Sarlette, S. Bonnabel, and R. Sepulchre. Coordinated motion design on lie groups. *Automatic Control, IEEE Transactions on*, 55(5):1047–1058, May 2010.

[47] Alain Sarlette and Rodolphe Sepulchre. Consensus optimization on manifolds. *SIAM Journal on Control and Optimization*, 48(1):56–76, 2009.

[48] Jens Gravesen, Steen Markvorsen, Robert Sinclair, and Minoru Tanaka. *The cut locus of a torus of revolution*. Technical University of Denmark. Department of Mathematics, 2003.

[49] Benedetto Piccoli and Francesco Rossi. Measure-theoretic models for crowd dynamics. In *Crowd Dynamics, Volume 1*, pages 137–165. Springer, 2018.

[50] J. Cheeger and David G. Ebin. *Comparison theorems in Riemannian geometry*, volume 365. AMS Chelsea Publishing, 1975.

[51] Seung-Yeal Ha, Dongnam Ko, Jinyeong Park, and Xiongtao Zhang. Collective synchronization of classical and quantum oscillators. *EMS Surveys in Mathematical Sciences*, 3(2):209–267, 2016.

[52] G. Bliss. The geodesic lines on the anchor ring. *Annals of Mathematics*, 1(4):1–21, 1902.

[53] C. Moore. Braids in classical dynamics. *Physical Review Letters*, 70:3675–3679, June 1993.

[54] C. Moore and M. Nauenberg. New periodic orbits for the n-body problem. *ASME. J. Comput. Nonlinear Dynam.*, 4(1):307–311, 2006.

[55] Aleksei Fedorovich Filippov. Differential equations with discontinuous right-hand side. *Matematicheskii sbornik*, 93(1):99–128, 1960.

[56] S. Motsch and E. Tadmor. Heterophilious dynamics enhances consensus. *SIAM Review*, 56(4):577–621, 2014.

[57] Pierre Emmanuel Jabin and Sebastien Motsch. Clustering and asymptotic behavior in opinion formation. *Journal of Differential Equations*, 257(11):4165–4187, 12 2014.

[58] Julien M Hendrickx and John N Tsitsiklis. Convergence of type-symmetric and cut-balanced consensus seeking systems. *IEEE Transactions on Automatic Control*, 58(1):214–218, 2013.