Claremont Colleges

Scholarship @ Claremont

HMC Senior Theses

HMC Student Scholarship

2024

Exploring Sigmoidal Bounded Confidence Models with Mean Field Methods

Tian Dong

Follow this and additional works at: https://scholarship.claremont.edu/hmc_theses

Part of the Dynamical Systems Commons, and the Partial Differential Equations Commons

Recommended Citation

Dong, Tian, "Exploring Sigmoidal Bounded Confidence Models with Mean Field Methods" (2024). *HMC Senior Theses*. 285.
https://scholarship.claremont.edu/hmc_theses/285

This Open Access Senior Thesis is brought to you for free and open access by the HMC Student Scholarship at Scholarship @ Claremont. It has been accepted for inclusion in HMC Senior Theses by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@claremont.edu.

Exploring Sigmoidal Bounded Confidence Models with Mean Field Methods

Tian Dong

Heather Zinn Brooks, Advisor

Andrew Bernoff, Reader

HARVEY MUDD COLLEGE

Department of Mathematics

May, 2024

Copyright © 2024 Tian Dong.

The author grants Harvey Mudd College and the Claremont Colleges Library the nonexclusive right to make this work available for noncommercial, educational purposes, provided that this copyright statement appears on the reproduced materials and notice is given that the copying is by permission of the author. To disseminate otherwise or to republish requires written permission from the author.

Abstract

Mathematicians use models of opinion dynamics to describe how opinions in a group of people change over time, which can yield insight into mechanisms behind phenomena like polarization and consensus. In these models, mathematicians represent the community as a graph, where nodes represent agents and edges represent possible interactions. Opinion updates are modeled with a system of differential equations (ODEs). Our work focuses on the sigmoidal bounded confidence model (SBCM), where agents update their opinion toward a weighted average of their neighbors' opinions by weighting similar opinions more heavily. Using tools developed in physics (mean-field theory), we derive a continuity equation from the system of ODEs to further analyze the model's steady states and compare with numerical simulations.

Contents

List of Figures

Acknowledgments

Since my thesis is a capstone of my time at Mudd, I would like to use the space here to express my gratitude for all of the extraordinary people who have shaped my path here. These acknowledgements will certainly represent only a small fraction of the acknowledgements they deserve.

First, of course I'd like to thank my advisor, Prof. Heather Zinn Brooks, for introducing me to this exciting project, and for her patience, invaluable advice, and constant support both throughout this past year, and over my time at Mudd. As my thesis advisor, she has given me the guidance, resources, and freedom to become a budding researcher of mathematics. Though I've considered almost every major at Mudd, I have always been some flavor of math major, and as my major advisor, Prof. Heather encouraged me to pursue my interests in physics and math despite the many extra study abroad forms she had to sign. I am so honored to have had the opportunity to work with someone who inspires me as a mathematician, a teacher, and a person.

Secondly, I am grateful to Prof. Andrew Bernoff for agreeing to be my second reader, and for his support in guiding the direction of my thesis. His curiosity and interest helped fuel my motivation, and the questions he asked always introduced me to new and interesting directions to explore.

Next, to Prof. Jon Jacobsen, Melissa Hernandez-Alvarez, and DruAnn Thomas, a huge thanks for making the thesis class what it is. Prof. Jakes has been a constant source of advice and encouragement throughout the year. Melissa always made sure the tech side of our thesis was running smoothly, and DruAnn always made sure I had what I needed—whether that consisted of sustenance after a long thesis writing session, or a laser pointer for my conference talk.

I also owe a great deal of thanks to my thesis classmates Niles Babin, Delaney Cohn, Kai Rajesh, Toby Anderson, Emma Zhang, Clay Adams, and Jasper Bown for being a supportive cohort, and for sharing their research

interests. Also, a special thanks goes to Clay and Jasper for the company during late night work sessions and for introducing me to so many exciting parts of math. You all are such strong mathematicians, and I'm so honored to have gone through this journey with y'all.

When I came to Mudd, I had absolutely no intention on being a *math* major, as it felt very out of reach. However, thanks to the help and advice of Prof. Su, Prof. Donnelly's encouragement and conviction that I could become a successful physicist, the mentorship of Prof. Shuve as my research advisor, and the amazing teaching of Prof. Omar and Prof. Karp which helped me discover my excitement and joy in understanding mathematical ideas, and communicating them to others; I finally decided to pursue math. Since then, I've grown immensely as a mathematical communicator through Prof. Kagey's math forum class, and his passion for teaching will inspire me through grad school for years to come.

I stay connected to the math and physics communities at Mudd mainly through our Gender Equity in Math (GEM) club and through being an Academic Excellence tutor. I am so grateful to be a member of these two amazing teams—so much of my mathematical communication practice has been through these avenues. In particular, I would like to thank Prof. Wendy Menefee-Libey for making AE the amazing and robust program it is today, and Abby Steele for her incredible leadership as the president of WiM/GEM. I would also like to thank all of the frosh I met this year through AE for inspiring me with their curiosity and for bearing with me through my often long-winded explanations and growth as a tutor; and my WiM mentor group for being a source of joy and inspiration in my final year.

To my friends who have helped make Mudd feel like home—I cannot thank you enough for your emotional and scholarly support. Sasha, Kelly, Alysa, Clay, Jack, Amelia, and Jasper—thank you all for constantly encouraging me to explore and for reminding me to take care of myself. Please let me know if you're ever near Ann Arbor over the next few years :)

Lastly, to my family: Sakura and Scott, for their music recommendations, my grandparents for always looking out for me (and learning about string theory when I told them I was doing particle physics work), my brother Yuan for his company, and my father for helping fuel my interest in math and puzzles through countless games of Gomoku (I'm looking forward to winning against you next time we meet). Finally, I cannot thank my mother enough for her strength and endless love and support—she is the person to whom I owe all of this.

And of course, thank you, dear reader, for keeping my thesis company :)

Preface

Welcome to my thesis! This thesis serves as a capstone in many ways: in addition to uniting the work I've done with my advisor, Prof. Heather Zinn Brooks, over the past year, it also builds on some of my favorite experiences at Mudd and abroad, including learning real analysis with Prof. Karp (both one of my biggest mathematical struggles and one of my favorite math courses), doing particle theory work with Prof. Shuve, learning statistical mechanics abroad, and later grutoring for both analysis and statmech.

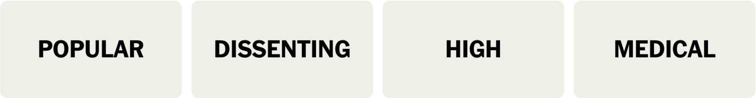
In some ways, this thesis has it all: puzzles, math, physics, birds, and many many fireflies. But it would also just be a (very inconvenient) paper paperweight without you. Thank you for reading, and I'm excited to introduce you to some of the math which has served as a close friend[1] during my senior year! Whether you're good friends with many mathematical subjects, or it's been a while since you've caught up with them, I hope you'll find parts of this thesis which speak to you. And if following calculations does not spark joy, please skip them! I hope you enjoy the pages to come :)

[1] and sometimes enemy

Chapter 1

The ABCs of Models of Opinion Dynamics

The thing that gets me up in the morning is Connections, one of my favorite New York Times games. The goal of Connections is to separate sixteen words or phrases into four categories (of four words), varying in difficulty. Some categories might be “Places in New York”, or “Anagrams of ‘live”’, or words which fill in the blank “___ child” (e.g. problem, flower, only). Let’s start with a small puzzle: what unites the following four words (see the footnote for the answer)[1]?



It’s quite exciting to discover connections, whether it’s between words, subjects, fields, or disciplines. For my thesis, we will be working with a *model of opinion dynamics*, which is used to study the way opinions spread in a group of *agents* by assigning each agent an “opinion”, keeping track of their interpersonal connections, and studying the way their opinions change over time. First developed by psychologists like Abelson (1967) to study social behavior, these models have since been used to study how populations become polarized, reach consensus, and even how sentiments spread on social media Noorazar et al. (2020). The particular bounded confidence model I’ll be working with, called the *sigmoidal bounded confidence*

[1] The category is “Second Opinion”, as these words are all featured in phrases where “opinion” comes second (e.g. dissenting opinion).

model (SBCM), unites two seemingly unrelated, well-studied models of opinion dynamics. In order to study these dynamics, my thesis will connect mathematical fields of inquiry, as well as a variety of disciplines ranging from our social science origins to the physics-inspired tools we'll use to explore the SBCM.

We'll start by getting to know the SBCM. But before we acquaint ourselves with the sigmoidal bounded confidence model, let's meet two models of opinion dynamics which inspired it.

1.1 A is for Averaging

Suppose, for concreteness, that we would like to model the evolution of a population of fireflies' favorite colors. One of these fireflies, Glowrdon Lightfoot, is tasked with organizing a light show, where the community of fireflies comes together to show off their colors. They know that each firefly will shine with their favorite color, but they would like to know which colors they have to work with. Glowrdon knows what each firefly's favorite color is now, 365 days before the show, and that each firefly interacts with their friends daily, so he sets about finding a model which he can use to help predict the colors he can use.

Glowrdon decides to take inspiration from their interactions with their own friends. Glowrdon's favorite color is blue, their two friends' favorite colors are purple and red. After interacting with their friends, Glowrdon realizes that their favorite color has changed under their influence. Their new favorite color is purple—the average of blue, purple, and red. Extrapolating, Glowrdon hypothesizes that everyone in their community updates their opinions approximately this way. This model which Glowrdon has created, where agents update their opinion based on the average of theirs and their *neighbors'* opinions, is called the Abelson model. Introduced by Abelson (1967), it is one of the first models of opinion dynamics.

To carry out their analysis, Glowrdon decides to make use of some mathematical tools, the first of which is an object called a *graph*:

Definition 1.1.1. A *graph* $\mathcal{G} = (V, E)$ consists of a set of vertices V along with a set of edges E, where each edge connects a pair of vertices. We can write each edge $e \in E$ ("element e in the set E") as a pair of vertices $e = \{v_1, v_2\}$ where $v_1, v_2 \in V$. In this thesis, I will use "vertex" and "node" interchangeably.

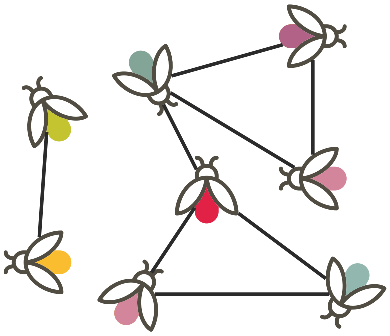


Figure 1.1 A model of a community using a graph, where fireflies are vertices, and edges are denoted with a line between fireflies. Each color can be translated to a number through its wavelength.

Since inter-firefly connections are an important part of the way opinions change in a community and they don't expect for friendships to change drastically in just one year, Glowrdon decides to model their community using a graph $\mathcal{G} = (V, E)$ (also shown in fig. 1.1), where each vertex (firefly) $i \in V$ is labeled with some opinion (favorite color) $x_i \in \mathbb{R}$. This approximation, that the time scale we're considering is short enough that inter-agent connections stay roughly static, is often used for the real-world models we're interested in[2]. Additionally, Glowrdon makes use of the following graph terminology:

Definition 1.1.2. If two vertices $i, j \in V$ are connected by an edge, we call them *adjacent* (written $i \sim j$, and pronounced "i is adjacent to j").

Note that we can rewrite set of edges E as a set of adjacent vertex-pairs so that:

$$E = \{(i, j) : i \sim j \text{ and } i, j \in V\}.$$

Definition 1.1.3. For a given vertex i, the total number of vertices adjacent to it is called the *degree* of i, written $\deg(i)$. It can be expressed with $\deg(i) = \sum_{j \sim i} 1$. For example, the yellow vertex in fig. 1.1 has a degree of 1, and the red vertex in fig. 1.1 has a degree of 3.

Definition 1.1.4. A *subgraph* $G' = (V', E')$ of a graph $G = (V, E)$ is a graph whose vertices are a subset of the vertices of G (i.e. $V' \subset V$) and whose

[2] Recently, researchers (e.g. Nugent et al. (2023)) have also been exploring the way an evolving underlying network affects the model's outcomes.

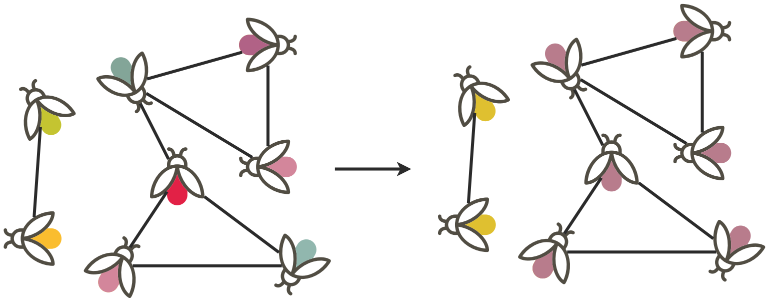


Figure 1.2 A visualization of the evolution of the Abelson model on the graph shown in fig. 1.1 with opinions residing in the color spectrum. We start with the initial condition on the left. After letting the system come to equilibrium, we reach the state on the right, where within each connected component, agents share the same opinion, i.e. *consensus*.

edges are precisely those which are in G and connect two vertices in V', i.e.

$$E' = \{(u, v) \in E \mid u, v \in V'\}.$$

Definition 1.1.5. A *path* between two vertices v_1 and v_n is a sequence of distinct vertices $(v_1, v_2, \ldots, v_n)$ where each neighboring pair of vertices are joined by edges, i.e. $(v_i, v_{i+1}) \in E$ for $i = 1, \ldots, n-1$.

Definition 1.1.6. In a *connected* graph, there is a path between any two vertices. As we can see, there is not a path between the yellow and red vertices in fig. 1.1 so this graph is not connected.

Definition 1.1.7. A *connected component* of a graph G is a connected subgraph which is not part of any larger connected subgraph. Any graph can be split into connected components. For example, fig. 1.1 has two connected components: one containing the yellow vertex, and another containing the red one.

Now that we have a graph, we can think about how it changes over time. Glowrdon starts by writing down a rule which governs how the opinions of each vertex changes with each time step. In their case, the new opinion at time t_{n+1} of the ith firefly $x_i(t_{n+1})$ is determined by the average of opinions of it and its neighbors at time t_n, so that:

$$x_i(t_{n+1}) = \frac{1}{1 + \deg(i)}\left(x_i(t_n) + \sum_{j \sim i} x_j(t_n)\right). \tag{1.1}$$

Now, Glowrdon just needs to solve this equation 364 times for each firefly. Though this is doable, it will take Glowrdon too long—and leave them with too little time for the other preparation work they must do. However, given how often Glowrdon is updating their model, they figure that just by changing the scale on which they view time, each time step is only $1/365 \approx 0.0027$ years—an approximately continuous scale. If Glowrdon can make opinions x_i depend continuously on time, then they only need to solve *one* equation for each firefly! In general, this approximation works well as long as we assume that many (fairly regular) interactions happen within our time frame. To do this, Glowrdon reformulates the update equation (eq. (1.1)) as a *differential equation*. To start, Glowrdon takes the difference between time steps:

$$
\begin{aligned}
x_i(t_{n+1}) - x_i(t_n) &= \frac{1}{1+\deg(i)}\left(x_i(t_n) + \sum_{j\sim i} x_j(t_n)\right) - x_i(t_n) && (1.2)\\
&= \frac{1}{1+\deg(i)} \sum_{j\sim i}(x_j(t_n) - x_i(t_n)). && (1.3)
\end{aligned}
$$

Now, in order to take the limit $t_{n+1} \to t_n$, we'll introduce a coupling strength ε, which controls for how quickly our system converges.

$$
x_i(t_{n+1}) - x_i(t_n) = \frac{\varepsilon}{1+\deg(i)} \sum_{j\sim i}(x_j(t_n) - x_i(t_n))
$$

When ε is larger, our system changes more dramatically at each time step. Since we want to make sure our system changes smoothly, we require that in the limit as $t_{n+1} \to t_n$, $\frac{\varepsilon}{t_{n+1}-t_n} \to 1$ so that by the definition of a derivative:

$$
\begin{aligned}
\lim_{t_{n+1}\to t_n} \frac{x_i(t_{n+1}) - x_i(t_n)}{t_{n+1} - t_n} &= \lim_{t_{n+1} - t_n} \frac{\varepsilon}{t_{n+1} - t_n} \frac{1}{1+\deg(i)} \sum_{j\sim i}(x_j(t_n) - x_i(t_n))\\
\dot{x}_i(t) = \frac{\mathrm{d}x_i}{\mathrm{d}t} &= \frac{1}{1+\deg(i)} \sum_{j\sim i}(x_j(t) - x_i(t)). && (1.4)
\end{aligned}
$$

In one last simplification in their analysis, Glowrdon reasons that since 365 days is a long time, on the 364th day, they do not expect that their communities' opinions will change much. Though it's not clear yet why this approximation should apply, we'll see later that we can run simulations of the model to verify this assumption. Mathematically, Glowrdon expects that the system will settle down near a *steady state*, which occurs when $\dot{x}_i(t) = 0$ for each $i \in V$.

It turns out that all steady states for this model have $x_j(t) = x_i(t)$ for all $j \sim i$. In this case, whenever there is a *path* between two vertices, i.e. $i \sim a \sim b \sim \cdots \sim j$, since $x_i(t) = x_a(t) = \cdots = x_j(t)$, all of the vertices along the path must share the same opinion. In particular, in any connected component of a graph (where I can find a path between any two vertices in the component) we expect the steady state opinion of each vertex to be constant. We call this state within each connected component a *consensus state*. A visual depiction of an example model is shown in **??**.

To show why this is the only steady state, consider a connected component G' of G. Let's call y the maximum opinion expressed by agents in G', and suppose Max is one such vertex with opinion $x_{\text{Max}} = y$. Since G' is connected, any other vertex v in G' is connected to Max by a path $(\text{Max}, u_1, u_2, \ldots, u_n, v)$. Note that since we are at a steady state, $\dot{x}_{\text{Max}} = 0$ by eq. (1.4), so we must have $x_j(t) - x_{\text{Max}}(t) = 0$ for each $j \sim \text{Max}$. Thus, Max shares the same opinion as all of its neighbors. In particular Max shares the same opinion as u_1, the first vertex along the path connecting Max and v. Since now u_1 is also a vertex possessing the maximum opinion of agents in G', we can repeat this argument for u_1 and find that u_2 must also share the maximum opinion. Continuing along the path, we can conclude that v must share the same opinion as Max. Since our choice of v was arbitrary, the same argument applies to all vertices in G', so G' has, as its only steady state, a consensus state.

1.2 B is for Bounded

The conclusions of the previous section may feel unsatisfying. Why does this model predict that all connected communities find consensus, when this phenomenon rarely occurs in the real world? If real-world interactions are approximated well by the dynamics of the Abelson model, we could do away with most disagreements just by allowing the two parties time to talk.

In contrast to always finding agreement, sociologists found that people seek and favor information which confirms their currently held beliefs Fischer et al. (2010), which they called *selective exposure*. Hegselmann and Krause (2002) introduced a model incorporating this effect (the HK model), where opinions are updated with the average of only *similar* adjacent opinions.

More concretely, we introduce a *confidence level* δ—instead of averaging all opinions, we'll only average neighboring opinions which are within δ of the original opinion. Thus, our update function (eq. (1.4)) becomes:

$$x_i(t + \Delta t) = \sum_{j \sim i} x_j(t)\omega(x_i, x_j) \tag{1.5}$$

$$\text{where} \quad \omega(x_i, x_j) = \frac{\mathbb{1}[|x_j - x_i| < \delta]}{\sum_{j\sim i} \mathbb{1}[|x_j - x_i| < \delta]} \tag{1.6}$$

Here, $\mathbb{1}$ is the *indicator function*,

$$\mathbb{1}[x < \delta] = \begin{cases} 1 & x < \delta \\ 0 & x \geq \delta. \end{cases}$$

which is equal to one whenever the bracketed condition is satisfied and zero everywhere else. Additionally, the way we've written our update rule in eq. (1.5) has suggested that this update rule is a *weighted average*[3]. Indeed, we've defined our weights $\omega(x_i, x_j)$ in such a way that they sum to one:

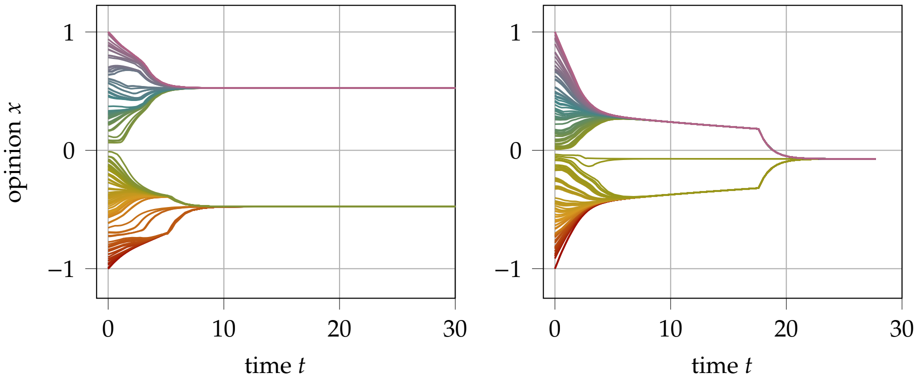
$$\sum_{j\sim i} \omega(x_i, x_j) = \sum_{j\sim i} \frac{\mathbb{1}[|x_j - x_i| < \delta]}{\sum_{j\sim i} \mathbb{1}[|x_j - x_i| < \delta]} = \frac{\sum_{j\sim i} \mathbb{1}[|x_j - x_i| < \delta]}{\sum_{j\sim i} \mathbb{1}[|x_j - x_i| < \delta]} = 1.$$

As we did in the previous section, we'll rewrite this equation in the form of a differential equation:

$$\dot{x}_i(t) = \sum_{j\sim i} (x_j(t) - x_i(t))\omega(x_i, x_j).$$

From this equation, we can also examine the values of δ which lead to steady states in the model. The model's behavior changes depending on many factors including the confidence bound δ, the (substrate) graph structure, and initial opinions of each vertex. Even by restricting ourselves to the case where each vertex is adjacent to every other vertex (so the graph structure is a *complete graph*) and focusing on the effect of δ by randomly distributing the vertices' initial opinions, the model exhibits a wide variety of behaviors, with some examples shown in fig. 1.3.

[3] In fact, the Abelson model we introduced above can also be formulated in this manner, except with a constant weight function.



a. $\delta = 0.3$.

b. $\delta = 0.5$.

Figure 1.3 Plots of the evolution of the HK model at varying values of δ with random initial conditions on the complete graph. By varying δ, we can control the number of factions our population ends up in. Figures based on Hegselmann and Krause (2002), and generated using code from Brooks et al. (2023).

One additional component Hegselmann and Krause introduced is a type of agent called a "zealot": these vertices have an update rule given by

$$\dot{x}_i = 0$$

so that they're happy to share their opinion, but not willing to change it. We can now split our vertex set V into two partitions: a set of persuadable vertices we'll call $\mathcal{P}$, and a set of zealots we'll call $\mathcal{Z} = V \backslash \mathcal{P}$. Now, in total, our update rules becomes:

$$\dot{x}_i(t) = \begin{cases} \sum_{j\sim i}(x_j(t) - x_i(t))\omega(x_i(t), x_j(t)) & i \in \mathcal{P} \\ 0 & i \in \mathcal{Z} \end{cases}. \tag{1.7}$$

A visualization of this process is shown in fig. 1.4. For simplicity, let's consider a graph with only two zealots $\mathcal{Z} = \{1, 2\}$ with $x_1 = 1$, $x_2 = -1$. Even in this simpler case, we can get a lot of interesting behavior, as shown in fig. 1.5.

1.2.1 A Brief Change in Perspective and Look Ahead

Looking between the two models we've seen so far, we have on the one hand the Abelson model, which averages all surrounding vertices (and

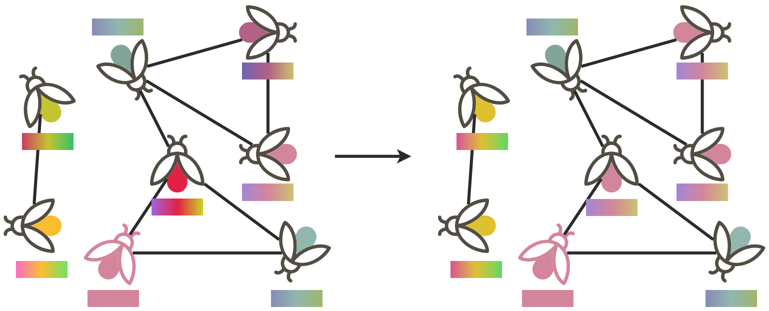
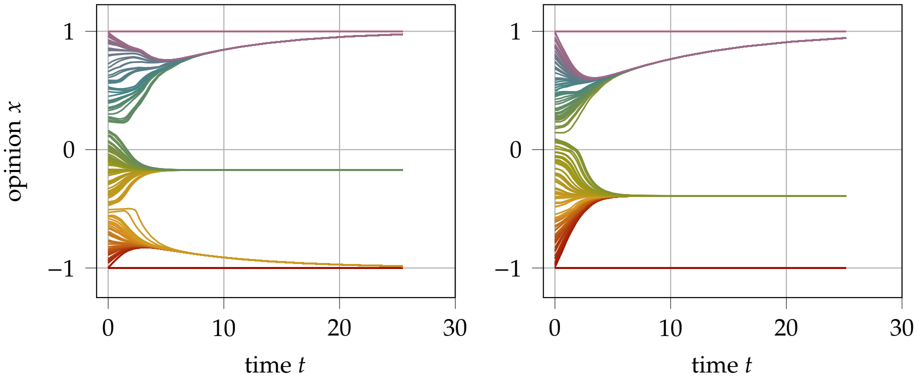


Figure 1.4 A visualization of the HK model where opinons again lie on the color spectrum. Here, each firefly has a confidence bound of allowable favorite colors. Additionally, there is one zealot firefly shown in pink who will not change its own color, but affects the colors of others. After a long time, the pair of fireflies on the left behaves the same as in the Abelson model, but the group of fireflies on the right split into two factions. Since the fireflies who like blue are not influenced by those who like purple, they retain their opinions, effectively separating the remaining purple and red fireflies into two pairs. Then, within each pair, the fireflies reach a consensus (the red firefly is indoctrinated by the zealot).



a. $\delta = 0.3$. **b.** $\delta = 0.5$.

Figure 1.5 Evolution of the HK model with two zealots. The initial conditions were generated uniformly at random −1 to 1. This was generated using the code repository from Brooks et al. (2023).

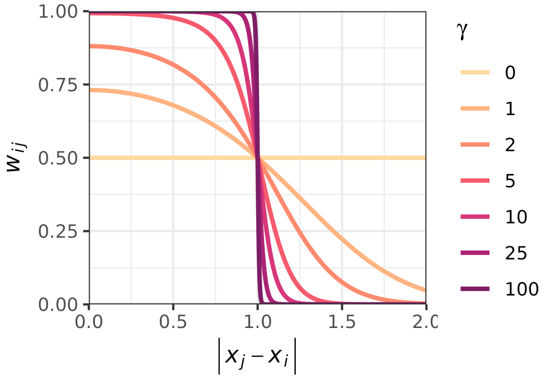


Figure 1.6 The dependence of the sigmoid on values of γ. Note that the yellow curve is the weight function for the Abelson model (a constant), while the purple curve approaches the weight function for the HK model (a step function). Figure from Brooks et al. (2023).

consequently has a constant weight function), and the HK model, which puts a bound on which values contribute to the average. Plotting the weight function for each of these, we can see in fig. 1.6 that these two weight functions can be thought of as extremes in a broader class of weight functions, parameterized by a real number γ. Conveniently for us, these weight functions are also smooth (infinitely differentiable and continuous) which is helpful when working with calculus, unlike the HK weight function.

Such a weight function can be interesting to study for a variety of reasons—we might ask questions like:

- For which γ values does this model approximate the Abelson model or the HK model?
- How does the new model differ from the Abelson and HK models when the weight function is in between the two extremes?
- If we vary the structure of the substrate graph of our model, how do the steady states of our system change? Do they match the behavior of the steady states for the Abelson or HK models?

Though we won't be able to immediately dive into explorations of these questions, we'll set things up by finding the mathematical description of this model.

1.3 C is for Continuous

As we described, we'd like to find a continuous weight function which bridges the gap between the Abelson and HK models. In particular, the functions shown in fig. 1.6 are called *sigmoids*[4]. They're given by the following formula:

$$\omega_\gamma(x_i, x_j) = \frac{w_\gamma(x_i, x_j)}{\sum_{k \sim i} w_\gamma(x_i, x_k)} \quad \text{with} \quad w_\gamma(x_i, x_j) = \frac{1}{e^{\gamma((x_j - x_i)^2 - \delta)} + 1}. \tag{1.8}$$

Here, γ and δ are parameters we can vary: δ determines the point at which the function crosses $w = \frac{1}{2}$, and γ determines the "sharpness" of the curve.

We can see that when $\gamma = 0$, the exponential term is a constant, so that $w(x_i, x_j) = \frac{1}{2}$ (a constant!) which gives the Abelson model. On the other hand, when $\gamma \to \infty$, the exponential term is very sensitive to whether $(x_j - x_i)^2 < \delta$:

$$\lim_{\gamma \to \infty} e^{\gamma((x_j - x_i)^2 - \delta)} = \begin{cases} \infty & (x_j - x_i)^2 > \delta \\ 1 & (x_j - x_i)^2 = \delta \\ 0 & (x_j - x_i)^2 < \delta \end{cases} \tag{1.9}$$

$$\text{so} \lim_{\gamma \to \infty} w_\gamma(x_i, x_j) = \lim_{\gamma \to \infty} \frac{1}{e^{\gamma((x_j - x_i)^2 - \delta)} + 1} = \begin{cases} 0 & (x_j - x_i)^2 > \delta \\ \frac{1}{2} & (x_j - x_i)^2 = \delta \\ 1 & (x_j - x_i)^2 < \delta \end{cases} \tag{1.10}$$

which is exactly the weight function from the HK model with the exception that when $x_i = x_j$, here we have $w(x_i, x_j) = \frac{1}{2}$, whereas in the HK model $w(x_i, x_j) = 1$. Since the sequence $w_\gamma(x_i, x_j)$ converges for fixed points x_i and x_j, we can say that the w_γ *converge pointwise* to the piecewise function after the equal sign.

This model is called the *sigmoidal bounded confidence model* (SBCM), and was introduced by Brooks et al. (2023). There, they explored the behavior

[4] Sigmoids also give us a connection to physics! From quantum physics, we learn that electrons (and in general, fermions) satisfy the Pauli exclusion principle, i.e. each energy level can be occupied by at most one electron. Consequently, they have a distribution described by the following sigmoid, also called the Fermi-Dirac distribution:

$$\bar{n} = \frac{1}{e^{\beta(\epsilon - \mu)} + 1}$$

where $\bar{n}$ tells us the average number of electrons with a given energy ϵ. We can see that these parameters correspond to ours in the following way: $\beta \to \gamma$, $\epsilon \to (x_j - x_i)^2$, and $\mu \to \delta$.

of the model on a variety of substrate graphs, and showed that the linearly stable steady states of the model indeed converges to those of the HK model as γ tends to infinity and to the Abelson model as $\gamma \to 0$. However, similar to the HK model, the SBCM is difficult to characterize for intermediate values of γ due to its sensitivity to the underlying graph structure. In this work, we hope to begin extending their exploration by borrowing tools from a parallel branch of inquiry developed for approximating large aggregations of particles in physics, and learn more about the SBCM.

1.4 Discussion and Outline

In this chapter, we started our journey by acquainting ourselves with graphs and opinion dynamics through the Abelson model. After deriving the continuous-time version of the model and discussing some of its shortcomings, we introduced the Hegselmann–Krause model, which accounts for agents' selective exposure. Finally, by noticing some similarities between these two models, we motivated the SBCM, which is the model we'll be getting to know in this thesis.

For a look ahead, in chapter 2 I'll introduce an important approximation, the mean-field limit, which was inspired by physicists' study of the way particles move in gases, and apply it to the SBCM. We'll also take a look at a few simulations of the discrete SBCM to gain intuition for how we expect the system to behave in the mean-field limit. Then, in chapter 3, we'll analyze a simplified version of the mean-field SBCM using some powerful mathematical tools from a particular category of PDE problems called gradient flows. It turns out that this simplified mean-field SBCM is a nail this hammer can strike—using this tool, we'll show that there exists a unique solution to our system, and this solution behaves in the ways that we expect from simulations. Finally, in chapter 4, we'll summarize again everything we've talked about, and list a handful of lines of inquiry we could eventually explore with the mean-field SBCM. Finally, if you're interested in the lore behind the fireflies, feel free to check out appendix B (totally optional). Now sit back, relax, and enjoy :)

Chapter 2

The DEs of the Mean-Field Approximation

Like Prof. Sahakian with squirrels, I get quite excited when I spot birds—in particular, after I got back from studying abroad in Europe where I got to know a few of the local birds, I noticed that near Harvey Mudd, there is a population of European Starlings (an invasive species). In addition to being excellent mimics, they also have the best collective noun of all birds (and it's hard to beat a murder of crows): a flock of starlings is called a "murmuration" (due to the sound of their flapping wings).

In stark contrast to the one or two starlings I often see around Mudd, there are cities where you can see millions at a time (see fig. 2.1). Though each individual starling is moving in response to its neighbors, as a flock, a pattern emerges, and they behave like a fluid moving in response to unseen forces. Similarly, in physics, even though (as far as we know), the Standard Model tells us that everything[1] in the universe is comprised of particles, we can still make good physical predictions with macroscopic descriptions (e.g. density, volume, pressure).

Originating in physics, the mean-field approximation is a powerful tool for changing the scale at which we look at a system. This approximation, which roughly translates discrete sets to continuous ones, was developed in statistical mechanics (a field in physics which studies large aggregations of particles) and allows us to make problems involving large numbers of particles tractable by averaging over certain parameters. We'll use this approach to try to gain intuition about how the SBCM behaves—instead of

[1] Excepting gravity, for now...

Figure 2.1 Photo collage: European Starling by simonglinn via Birdshare; murmuration photo by ad551 via Creative Commons. Taken from Alfano (2013).

considering the fireflies as individuals, we'll instead model them as particles in a "fluid".

2.1 D is for Deriving the Mean-Field Limit

Recall that agents in the SBCM update their opinion using a weighted average of their neighbors opinions given by eq. (1.7), with weight function given by eq. (1.8). I've also rewritten these equations here for our convenience:

$$\dot{x}_i(t) = \begin{cases} \sum_{i\sim j}(x_j(t) - x_i(t))\omega(x_i(t), x_j(t)) & i \in \mathcal{P} \\ 0 & i \in \mathcal{Z}, \end{cases} \tag{2.1a}$$

$$\omega_\gamma(x_i, x_j) = \frac{\omega_\gamma(x_i, x_j)}{\sum_{k\sim i} \omega_\gamma(x_i, x_k)} \quad \text{with} \quad w_\gamma(x_i, x_j) = \frac{1}{e^{\gamma((x_j - x_i)^2 - \delta)} + 1}. \tag{2.1b}$$

From now on, we'll write $w(x_i, x_j)$ to refer to $w_\gamma(x_i, x_j)$, and to simplify things, we'll consider the case where every agent interacts with every other agent—that is, the graph we're working with is a complete graph. This replaces all sums over adjacent nodes with sums over all nodes.

Now, as we've left it, the SBCM is still quite difficult to solve, and the number of coupled differential equations we need to solve only gets larger as

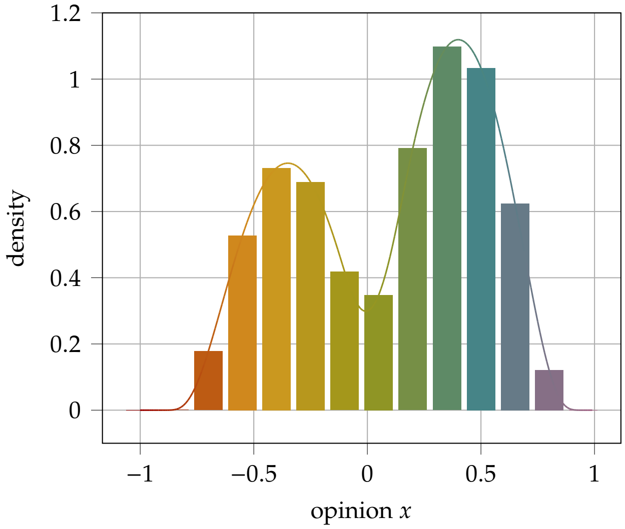


Figure 2.2 The bars' area represent $p(y,t)$, whereas the continuous line represents the underlying distribution $\rho(y,t)$. By partitioning the opinion space (x-axis) into a finite number of buckets, we obtain the relation $p(y,t) \approx \rho(y,t)\Delta y$.

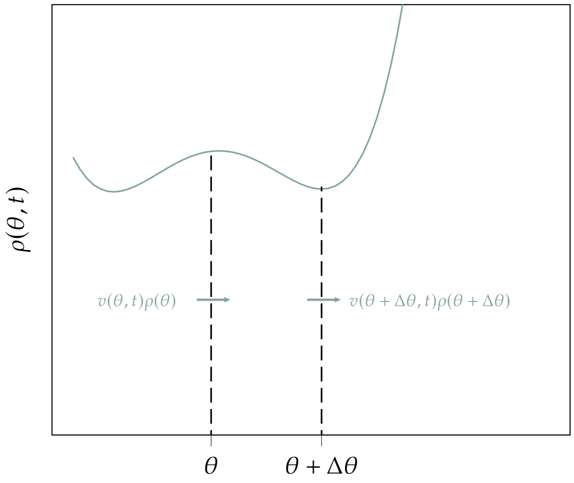

$\rho(\theta,t)$
$v(\theta,t)\rho(\theta)$
$v(\theta+\Delta\theta,t)\rho(\theta+\Delta\theta)$
θ
$\theta+\Delta\theta$

Figure 2.3 A visual representation of eq. (2.2), the continuity equation. Over a small change in time, the change in the density in the interval $(\theta, \theta+\Delta\theta)$ consists of density which enters the region from the left and the density which leaves the region from the right. By taking the difference, we find that $\frac{\partial\rho}{\partial t}\Delta\theta = -\Delta(pv)$, where the negative sign is due to density *leaving* on the right. In the limit as $\Delta\theta \longrightarrow 0$, we obtain the continuity equation.

N increases. However, if we take $N \to \infty$, our system (maybe surprisingly) can be simplified! As shown in fig. 2.2, instead of keeping track of each individual firefly, we'll reduce the size of our system by keeping track of only one quantity: the density of fireflies. Then, treating the interactions between agents as interactions between particles in a fluid, we can discover how our "fluid" population flows over the space of opinions.

Our new differential equation keeps track of a *density* of agents as a function of their opinions, and is governed by the following continuity equation, also illustrated in fig. 2.3:

$$\frac{\partial}{\partial t}\rho(x,t) + \frac{\partial}{\partial x}(v(x,t)\rho(x,t)) = 0, \tag{2.2}$$

where ρ is the density and v is the velocity of our fluid. Note that in this equation, ρ changes in time, which is not true of zealots. Thus, we'll use $\rho(x,t)$ to represent the density of *persuadable* agents, and $\rho_{\mathcal{Z}}(x)$ to represent the (constant in time) density of zealot agents.

Now, since eq. (2.1) describes the way opinions change, we expect that this equation corresponds to $v(x,t)$ in the continuum limit. However, note that $v(x,t)$ depends on the opinion value x, while the sum in eq. (2.1a) depends on the node index i. So in order to find $v(x,t)$, we'll start by reparameterizing our sum:

$$\dot{x}(t) = \sum_{\text{distinct opinions } y} (n(y_i,t) + n_{\mathcal{Z}}(y_i))(y_i - x)\frac{w(x,y_i)}{\sum_{y_i}(n(y_i,t) + n_{\mathcal{Z}}(y_i))w(x,y_i)}$$

where in order to account for the possibility of multiple nodes having the same opinion, we've introduced $n(y_i,t)$ and $n_{\mathcal{Z}}(y_i)$, which respectively counts the number of persuadable and zealot agents with opinion y_i. It's now tempting to take the limit as the number of agents N approaches infinity, but we can't quite do this yet, as that would also take $n(y_i,t) \to \infty$. So to make sure things stay finite, we will divide by N in both the numerator and denominator and replace $p(y_i,t) = n(y_i,t)/N$ (and similarly for zealots) so that:

$$\dot{x}(t) = \frac{\sum_y (p(y_i,t) + p_{\mathcal{Z}}(y_i))(y_i - x)w(x,y_i)}{\sum_y (p(y_i,t) + p_{\mathcal{Z}}(y_i))w(x,y_i)}.$$

When we made this replacement, instead of keeping track of a count $n(y_i,t)$, we instead keep track of the count divided by the total, i.e. a proportion, where $\sum_{y_i} p(y_i,t) = 1$. Furthermore, if we imagine a continuous underlying distribution of opinions, we can interpret each $p(y_i,t)$

as "lumping together" the opinions from y_i to $y_{i+1} = y_i + \Delta y$, i.e. that $p(y_i, t) = \int_{y_i}^{y_i+\Delta y} \rho(y', t)\,\mathrm{d}y' \approx \rho(y_i, t)\Delta y$, as shown in fig. 2.2. Thus, by taking the limit $\Delta y \to 0$, we obtain $p(y, t) = \rho(y, t)\,\mathrm{d}y$.

Now, making this replacement in the limit $\Delta y \to 0$ and taking sums to integrals:

$$v(x, t) = \frac{\int_{-\infty}^{\infty} (\rho(y, t) + \rho_{\mathcal{Z}}(y))(y - x) w(x, y)\,\mathrm{d}y}{\int_{-\infty}^{\infty} (\rho(y, t) + \rho_{\mathcal{Z}}(y)) w(x, y)\,\mathrm{d}y}. \tag{2.3}$$

Together with eq. (2.2), we have replaced our N ordinary differential equations in eq. (2.1a) with one partial differential equation (the continuity equation) which governs the motion of our system! Altogether, the continuity equation eq. (2.2) imposes the constraint that total "mass"[2] is conserved, while eq. (2.3) tells us the way ρ evolves, which we've derived directly from the discrete case.

Now we're set—all we need to do is just solve this one equation. Unfortunately for us, partial differential equations which are solvable analytically (i.e. by-hand) are quite elusive. Most physical systems (likely including this one) do not admit analytic solutions. In these situations, mathematicians must rely on numerical methods to evolve their system. Since I'm more partial to the analysis of our system (rather than the numerics), for the rest of this thesis we'll explore the many ways we can compare analytic results from the mean-field SBCM with numerical simulations of the (discrete) SBCM.

2.1.1 Recovering the Discrete SBCM

Now that we have our equation eq. (2.2), the first order of business is to check that this is indeed a suitable approximation of our system. To do this, we'll show that the discrete case is recoverable from this new equation[3].

In order to show that our new system contains the same dynamics as the SBCM, we'll show that we can recover the discrete SBCM by allowing ρ to be comprised of a sum of δ-distributions[4], following a similar approach to Bernoff and Topaz (2011). This allows us to get as close as possible to a

[2] i.e. $\int_{-\infty}^{\infty} \rho(x, t)\,\mathrm{d}x$

[3] If you got a little bit lost in the details of the previous section, don't stress—this section will hopefully give a bit more intuition and context for why eq. (2.2) and eq. (2.3) are the mean-field limit we're looking for.

[4] For the unacquainted, please visit appendix A.

discrete system while still working with continuous “functions”. Without further ado, suppose that ρ is a finite sum of δ-distributions, i.e.

$$\rho(x,t) = \sum_{i=1}^{N} \delta(x - x_i(t))$$

where $N \in \mathbb{N}$ and $x_i(t)$ is a real-valued function of t. Then, we can plug into eq. (2.3) and simplify:

$$\begin{aligned} v(x,t) &= \frac{\int_{-\infty}^{\infty}(\rho(y,t) + \rho_{\mathcal{Z}}(y))(y-x)w(x,y)\,\mathrm{d}y}{\int_{-\infty}^{\infty}(\rho(y,t) + \rho_{\mathcal{Z}}(y))w(x,y)\,\mathrm{d}y} \\ &= \frac{\int_{-\infty}^{\infty} \rho_{\mathcal{Z}}(y) + \sum_{i=1}^{N} \delta(y - x_i(t))(y-x)w(x,y)\,\mathrm{d}y}{\int_{-\infty}^{\infty} \rho_{\mathcal{Z}}(y) + \sum_{i=1}^{N} \delta(y - x_i(t))w(x,y)\,\mathrm{d}y} \\ &= \frac{\sum_{i=1}^{N}(x_i(t) - x)w(x, x_i(t)) + Z_1(x)}{\sum_{i=1}^{N} w(x, x_i(t)) + Z_2(x)}. \end{aligned}$$

where we’ve extracted $\rho_{\mathcal{Z}}$ in the last step:

$$\begin{aligned} Z_1(x) &= \int_{-\infty}^{\infty} \rho_{\mathcal{Z}}(y)(y-x)w(x,y)\,\mathrm{d}y \\ Z_2(x) &= \int_{-\infty}^{\infty} \rho_{\mathcal{Z}}(y)w(x,y)\,\mathrm{d}y \end{aligned}$$

Now, since it’s a bit pesky to work with δ distributions, we’ll integrate eq. (2.2) with respect to x so that:

$$\begin{aligned} C &= \frac{\partial}{\partial t}\int \rho(x,t)\,\mathrm{d}x + (v(x,t)\rho(x,t)) \\ &= -\sum_{i=1}^{N} \dot{x}_i(t)\mathbb{1}[x \geq x_i(t)] + \int v(x,t)\sum_{i=1}^{N} \delta(x - x_i(t))\,\mathrm{d}x \\ &= -\sum_{i=1}^{N} \dot{x}_i(t)\mathbb{1}[x \geq x_i(t)] + \sum_{i=1}^{N} v(x_i(t),t)\mathbb{1}[x \geq x_i(t)] \end{aligned}$$

so that after rearranging and equating coefficients:

$$\dot{x}_i(t) = v(x_i(t),t) = \frac{\sum_{j=1}^{N}(x_j(t) - x_i(t))w(x_i(t), x_j(t)) + Z_1(x)}{\sum_{j=1}^{N} w(x_i(t), x_j(t)) + Z_2(x)},$$

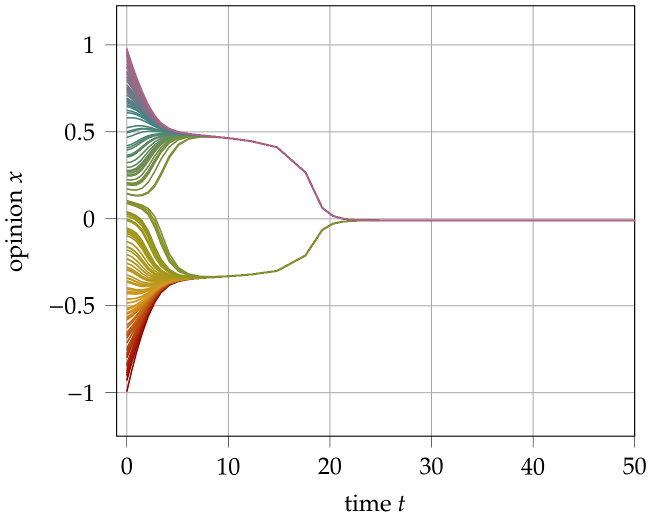


Figure 2.4 A plot with $n = 150$, $\gamma = 10$, $\delta = 0.03$, and no zealots. Note that despite the initial polarization, they both converged to a unique steady state in the absence of zealots. Each of the 150 agents are colored by their initial opinion.

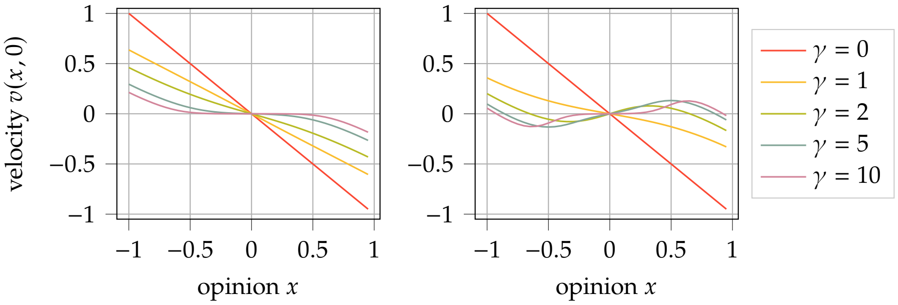
which is the same equation (eq. (2.1a)) as in the discrete case!

Thus, we've checked that our mean-field approximation indeed preserves the dynamics of the original SBCM, and we can now rest-assured that the forthcoming analysis is well-founded.

2.2 E is for Expectations and Simulations

Before we get into the mathematical weeds, let's develop some intuition for how our system should behave with discrete simulations. We'll focus on two cases: first, when there are no zealots, and then, when there are two zealots.

What do we expect to happen in the no zealot case? First, note that whenever agents interact, they tend toward a similar opinion as those they interact with. Thus, without zealots to pull agents apart, we might expect that after a long time, our system arrives at a consensus. A simulation in the no-zealot case is shown in fig. 2.4. We can see that despite the initial polarization, the agents eventually end up at a consensus. To see why this happens mathematically, we can look at the way our particles move. Note



a. No zealots.

b. Two zealots (±1).

Figure 2.5 The velocity profile $v(x)$ from eq. (2.3) with $\rho(y) = \frac{1}{2}\mathbb{1}[-1 \le x \le 1]$, $\delta = 0.2$, and a range of γ. In fig. 2.5a, there are no zealots and in fig. 2.5b there are two zealots with opinions ± 1. Since $v(-1) > 0$ and $v(1) < 0$, the agents will tend to move toward more similar opinions.

that when $\rho(x)$ is a uniform distribution centered at zero,

$$\operatorname{sgn}(v(x,t)) = -\operatorname{sgn}(x) \quad \text{where} \quad \operatorname{sgn}(x) = \begin{cases} -1 & x < 0 \\ 0 & x = 0 \\ 1 & x > 0 \end{cases}$$

so that agents in our system are always pushed toward each other. Similar to what we showed for the Abelson model, we can show the following for the discrete SBCM.

Lemma 2.2.1. *In an SBCM (on a complete graph) with no zealots and at least two persuadable nodes, the only steady state is a consensus state.*

Proof. To show this, we'll show the contrapositive: that whenever we do not have consensus, there exists some agent i with $\dot{x}_i \neq 0$ (so that we do not have a steady state). First, let $X = \{x_i\}$ be the set of unique opinions expressed by our population. When there is no consensus, $\min(X) \neq \max(X)$. Now, let a be an agent with opinion $\min(X)$, so that $x_a \le x_j$ for every agent j. In particular, since we do not have a consensus, there exists some agent b such that $x_a < x_b$. Thus, from eq. (2.1a), its opinion changes based on the following equation:

$$\dot{x}_a(t) = \frac{\sum_{j: x_j \ge x_a} (x_j - x_a) w(x_a, x_j)}{\sum_j w(x_a, x_j)},$$

and considering only signs, since w is always positive:

$$\begin{aligned}\operatorname{sgn}\dot{x}_a(t) &= \operatorname{sgn}\left(\frac{\sum_j (x_j - x_a)w(x_a, x_j)}{\sum_j w(x_a, x_j)}\right)\\ &= \operatorname{sgn}\left(\sum_{j:x_j>x_a}(x_j - x_a)w(x_a, x_j) + \sum_{j:x_j=x_a}(x_j - x_a)w(x_a, x_j)\right)\\ &= \operatorname{sgn}\left(\sum_{j:x_j>x_a}(x_j - x_a)w(x_a, x_j)\right) \geq \operatorname{sgn}((x_b - x_a)w(x_a, x_b)) = 1.\end{aligned}$$

Thus, $\dot{x}_a \neq 0$ and we do not have a steady state. □

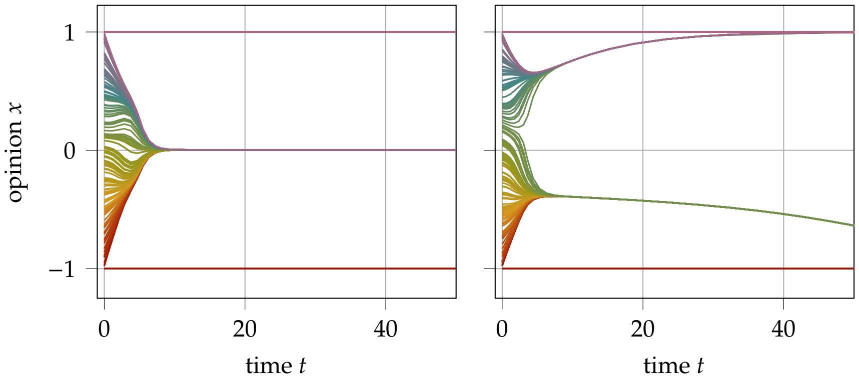
This argument relied on the fact that extremal agents always move toward more moderate opinions—we showed that any agent with the minimum opinion will move toward a higher opinion (in a system with at least two distinct opinions) since $\dot{x}_a > 0$, and a similar argument can be used to show that any agent b with the maximum opinion will always evolve toward a lower opinion[5] (i.e. $\dot{x}_b < 0$).

This also holds in a system with zealots! In fig. 2.6, we can see two simulations of a system of 150 agents with two distinct zealot opinions (±1) and five zealots at each opinion. In fig. 2.6a, our agents act (essentially) as if the zealots did not exist, while in fig. 2.6b, our agents seem to be converging toward the two zealot opinions. When a system contains zealots, if a persuadable agent a has an extremal opinon $|x_a| \geq 1$, then similarly $\dot{x}_a \neq 0$ and we do not have a steady state.

Note also that in a system with two zealots, the steady state cannot have persuadable agents who agree *exactly* with a zealot, (as the other zealot will cause $\dot{x} \neq 0$). Thus, the steady state system will have at least three distinct opinions. In the case of fig. 2.6b, we will end up with four "factions": the five zealots at +1, some persuadable agents at $x_a = 1 - \delta$, some persuadable agents at $x_b = -1 + \varepsilon$, and the zealots at −1, where δ and ε are very small and positive real numbers.

When we move to the continuous version, we can see that in fig. 2.5b, a similar phenomenon happens, where v intersects the x-axis at two additional points when γ increases. At the maximum and minimum values ±1, we can

[5] e.g. by negating all opinions and showing (as we did above) that the minimum always increases



a. A “consensus” state.

b. A polarized state.

Figure 2.6 Plots with $n = 150$, $\gamma = 10$, $\delta = 0.03$, and two zealot opinions ± 1 and five zealots at each opinion. These two simulations differ only by their initial conditions, which were randomly sampled from a uniform distribution over the interval $[-1, 1]$.

see that $\operatorname{sgn}(v(\pm 1, 0)) = \operatorname{sgn}(\mp 1)$. In particular, for a maximal opinion p, i.e. p is the least upper bound (*supremum*)[6] of the *support*[7] of $\rho + \rho_{\mathcal{Z}}$,

$$\begin{aligned}\operatorname{sgn}(v(p,t)) &= \operatorname{sgn}\left(\frac{\int_{-\infty}^{\infty} \rho(y,t)(y-x)w(x,y)\,\mathrm{d}y}{\int_{-\infty}^{\infty} \rho(y,t)w(x,y)\,\mathrm{d}y}\right)\\ &= \operatorname{sgn}\left(\int_{-\infty}^{p} \rho(y,t)(y-x)w(x,y)\,\mathrm{d}y + \int_{p}^{\infty} \rho(y,t)(y-x)w(x,y)\,\mathrm{d}y\right)\\ &= \operatorname{sgn}\left(\int_{-\infty}^{p} \rho(y,t)(y-x)w(x,y)\,\mathrm{d}y\right) = -1\end{aligned}$$

where in the first and last steps we used that $\rho(x,t) \geq 0$ and $w(x,y) \geq 0$, and in the third step we used $\rho(y,t) = 0$ for $y > p$ (as $p = \sup(\operatorname{supp}(\rho))$[8]).

[6] More concretely, this is a mathematical generalization of a maximum. For example, consider a function like $f(x) = -e^x$ (which is monotonically decreasing and asymptotes to the x axis as $x \longrightarrow -\infty$). For all intents and purposes, 0 is the “maximum” of the image of $f(x)$ even though 0 is not in the image of $f(x)$. Since a maximum of a set must belong in the set, we call this least upper bound the *supremum*.

[7] The support of a function is the set of elements in the function’s domain where the function is nonzero, i.e. $\operatorname{supp}(f) = \{x \mid f(x) \neq 0\}$. Since this is a (possibly infinite) set of elements, we aren’t guaranteed that it has a maximum (or minimum), so we use the supremum (or infimum, see footnote 9).

[8] “sup(supp(suppp(supppp(suρpppp(suppppppp(ρ))))))” — Clay Adams ’24 (abridged)

In general, *extremal* opinions, i.e. the least upper bound (*supremum*) or greatest lower bound (*infimum*)[9] of the support of $\rho + \rho_{\mathcal{Z}}$, will always evolve toward more moderate opinions. Though this argument does not account for occasions where the suprema don't exist, it provides intuition for the analogues we can make between the discrete and continuous cases. To make this more concrete, we'll need tools from the next section.

Finally, since opinions are always being pushed toward each other (as ω_γ is always positive), we also expect that the mean-field SBCM will converge to a finite number of opinion "factions" which we can write as δ-distributions. However, as we'll see in more detail in the next chapter, since ω_γ is bounded, this only happens in the limit as $t \longrightarrow \infty$, so we likely won't see δ-distributions emerge after some finite time t from a $\rho_0(x)$ which does not contain δ-distributions. Physically, this corresponds to communities which have more-or-less divided into factions, allowing for slight variations of opinion within each faction. Each faction will never reach *complete* agreement, but they will approach it as time goes on.

2.3 Findings

We started this chapter by introducing the mean-field limit as a way to reframe our model as a fluid flowing over opinion space rather than as discrete interacting agents. Mathematically, we derived it for our system by taking the limit as the number of agents $N \longrightarrow \infty$, resulting in the continuity equation eq. (2.2) which imposes conservation of mass, and eq. (2.3) which tells us how our "fluid" moves. Then, we verified that this model is a suitable approximation of the discrete system by extracting the discrete SBCM from the mean-field SBCM with a sum of δ-distributions.

After deriving our model, we took inspiration from simulations of the no-zealot case fig. 2.4 and the two-zealot case fig. 2.6 to motivate two conjectures:

1. The first one, which we showed is true for the discrete case, is that the range of opinions expressed by our population is always shrinking. We also used this to show that when there are no zealots and at least two agents, we always expect for the system to end up in a consensus state.

[9] Similar to the supremum from footnote 6, there is also a mathematical generalization for the minimum of a set, called the infimum.

In the mean-field SBCM, we can ask: **is the support of ρ is always shrinking?** We expect the answer to be yes, and motivated this by considering the sign of v at the infimum and supremum of the support of ρ, which showed us that v is always pushing extremal opinions toward more moderate ones.

2. Secondly, **are δ-distributions steady states of our system?** We argued that since the support is always shrinking and opinions are being pushed toward each other by ω_γ, we might expect for our system to approach δ-function steady-states. However, unless our initial condition $\rho_0(x)$ contains a δ-distribution, we don't expect for our system to coalesce into one in finite time.

In order to (at least partially) answer these questions (and more!) mathematically, we'll introduce some tools from PDEs and wack (and get wacked by) some weedy math in the next chapter.

Chapter 3

The FGHs of Wading into Wasserstein Weeds

To answer our questions from the previous chapter, we'll start by considering a simplified version of our model, where we set the denominator of our velocity flow (eq. (2.3)) to be a constant:

$$v(x,t) = \int_{-\infty}^{\infty} (\rho(y,t) + \rho_{\mathcal{Z}}(y,t))(y-x)w(x,y)\,\mathrm{d}y. \tag{3.1}$$

where, as before, $\rho(y,t)$ is the density of the persuadable nodes and $\rho_{\mathcal{Z}}(y)$ is the density of zealot nodes. In a way, this mixes the SBCM with the Abelson model, which features a constant denominator equal to the mass of the system (N in the discrete case) due to its constant weight function. With this approximation, we can apply tools from PDEs and obtain results which give insight into how the full mean-field SBCM behaves.

Before we dive in, we'll give a bit of a motivating example. Suppose we have a particle moving as a result of some potential. Since the particle is always moving in the direction that minimizes its energy (as the force on the particle $F(x) = -\nabla U(x)$ is the negative gradient of the potential energy U), we can write this as:

$$\dot{x}(t) = -\nabla U(x(t)), \tag{3.2}$$

which is the prototypical *gradient flow* equation. That is, starting at some $x(0) = x_0$, how can $x(t)$ move so that it minimizes U as quickly as possible?

How does this relate to our continuity equation? It turns out that this is related to the content of the SIAM Journal of Mathematical Analysis's *most* *downloaded article*[1]. Jordan et al. (1998) found that after making the suitable changes, techniques from the study of gradient flow equations can be used to study equations like continuity equations[2]. In particular, they can help us find answers to our questions at the end of the previous chapter!

3.1 F is for Formulating the Gradient Flow SBCM

Now, how can we reformulate our model as a gradient flow model? For one, in the gradient flow equation eq. (3.2), there's a $\dot{x}$ rather than a $\dot{\rho}$, so that instead of minimizing x, an element of $\mathbb{R}$, we'd like to minimize ρ, a fluid density in a function space.

Before we describe this function space, we'll build some intuition for which energy potential we're hoping to minimize by thinking about how our agents interact. Besides the Kuramoto model, a variety of models of biological systems contain pairwise interactions between agents, for example in chemotaxis (e.g. Keller and Segel (1970)), bird flocks (e.g. Cucker and Smale (2007) and Ha and Liu (2009)), or locust swarms (e.g. Bernoff and Topaz (2011)). In general, the mean-field approximation of these models are all special cases of the following equation:

$$\frac{\partial}{\partial t}\rho(x,t) = \frac{\partial}{\partial x}(\rho(x,t)V(x,t))$$

where:

$$V(x,t) = \frac{\partial}{\partial x}\left(a(\rho) - \int_{\mathbb{R}} G_\gamma(x-y)\rho(y,t)\,\mathrm{d}y + F(x)\right). \tag{3.3}$$

This velocity is constructed as the gradient of a potential energy which combines a diffusive term $a(\rho)$ describing the particles' tendency to spread out, an interaction term $\int G_\gamma(x-y)\rho(y,t)\,\mathrm{d}y$ with an *interaction potential* $G_\gamma(x,y)$, and a forcing term $F(x)$ representing an external potential which takes into account any external forces on our system of particles.

Now, how does the mean-field SBCM fit into the picture?

Theorem 3.1.1. *The adjusted mean-field SBCM is a gradient flow problem*!

Proof. In our case, we don't have any diffusion, as the only things influencing agents' opinions are their interactions, but we do have forcing due to the zealots. Thus, setting $a(\rho(x,t)) = 0$, we obtain $V(x,t) =$

[1] In its 56 volumes of existence.

[2] In particular, they connected gradient flow problems with a broader class of equations called Fokker–Planck equations.

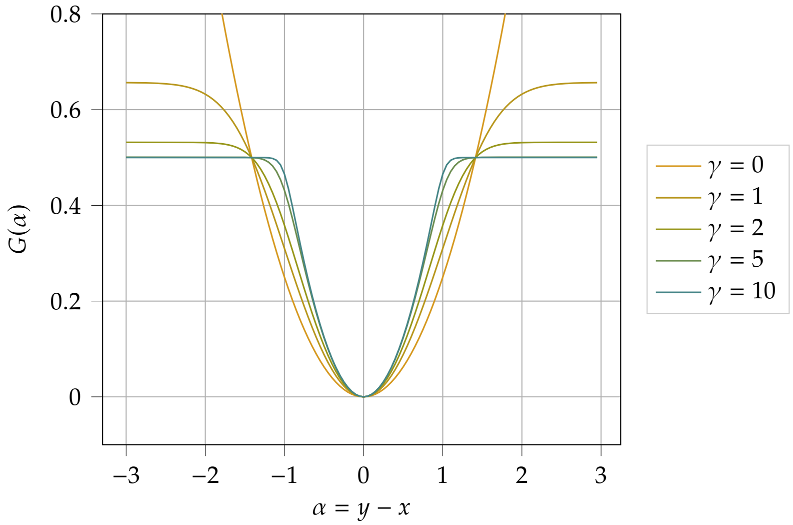


Figure 3.1 A plot of $G_\gamma(\alpha)$ for $\delta = 1$ and various values of γ.

$\frac{\partial}{\partial x}\left(\int G_\gamma(x-y)\rho(y,t)\,\mathrm{d}y + F(x)\right)$. Now, when $G_\gamma(x-y)$ is bounded above, by the dominated convergence theorem[3] we can pass the derivative under the integral and obtain:

$$V(x,t) = -\int_{\mathbb{R}} \frac{\partial}{\partial x} G_\gamma(x-y)\rho(y,t)\,\mathrm{d}y + \frac{\partial F}{\partial x}.$$

Now, since we want $V(x,t) = \int_{\mathbb{R}}(\rho(y,t) + \rho_{\mathcal{Z}}(y))(y-x)w_\gamma(x,y)$, we need:

$$\frac{\partial G_\gamma(x-y)}{\partial x} = G'(x-y) = -(x-y)\frac{1}{1+e^{\gamma((x-y)^2-\delta^2)}}, \tag{3.4}$$

$$\frac{\partial F}{\partial x} = \int_{\mathbb{R}} \rho_{\mathcal{Z}}(y)(y-x)w_\gamma(x,y). \tag{3.5}$$

$G_\gamma(\alpha) = -\int \frac{\alpha}{1+e^{\gamma(\alpha^2-\delta^2)}}\,\mathrm{d}\alpha$[4] is visualized in fig. 3.1. Also, note that since $G'(\alpha)$ is an odd function, any integral with bounds symmetric about the origin must vanish:

$$\int_{-c}^{c} G'(x)\,\mathrm{d}x = 0.$$

[3] See (again): measure theory!

[4] Which is also called an *incomplete Fermi-Dirac integral*, often used to find the expected value for the energy of a system of fermions (particles which obey the Pauli exclusion principle, e.g. electrons!)

Using this, we can show that $G_\gamma(\alpha)$ is even:

$$\begin{aligned}G_\gamma(\alpha) &= -\int_0^\alpha \frac{\alpha'}{1+e^{\gamma(\alpha'^2-\delta^2)}}\,\mathrm{d}\alpha' + C = \int_\alpha^0 \frac{\alpha'}{1+e^{\gamma(\alpha'^2-\delta^2)}}\,\mathrm{d}\alpha' + C\\ &= -\int_\alpha^{-\alpha} \frac{\alpha'}{1+e^{\gamma(\alpha'^2-\delta^2)}}\,\mathrm{d}\alpha' + \int_\alpha^0 \frac{\alpha'}{1+e^{\gamma(\alpha'^2-\delta^2)}}\,\mathrm{d}\alpha' + C\\ &= -\int_0^{-\alpha} \frac{\alpha'}{1+e^{\gamma(\alpha'^2-\delta^2)}}\,\mathrm{d}\alpha' + C = G_\gamma(-\alpha) \qquad (3.6)\end{aligned}$$

as desired. Since $\int_{\mathbb{R}} C\rho(y,t)\,\mathrm{d}y = C\int_{\mathbb{R}} \rho(y,t)\,\mathrm{d}y = C$ for any time t by our assumption that the total mass is a (unit) constant, it will vanish when we take a derivative with respect to x in eq. (3.3). Thus, we can set $C = 0$. It turns out also that G has an analytical solution:

$$G_\gamma(\alpha) = \begin{cases}\alpha^2/4 & \gamma = 0,\\ \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma(\alpha^2-\delta^2)}+e^{\gamma\alpha^2}}{e^{\gamma(\alpha^2-\delta^2)}+1}\right) & \text{otherwise.}\end{cases}$$

Note also that from this expression, we can see that when $\gamma \to \infty$ and $\alpha < \delta$, $e^{\gamma(\alpha^2-\delta^2)} \to 0$ so that:

$$\lim_{\gamma\to\infty} G_\gamma(\alpha) = \lim_{\gamma\to\infty} \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma(\alpha^2-\delta^2)}+e^{\gamma\alpha^2}}{e^{\gamma(\alpha^2-\delta^2)}+1}\right) = \frac{1}{2\gamma}\ln\left(e^{\gamma\alpha^2}\right) = \frac{\alpha^2}{2}.$$

When $\alpha > \delta$, $e^{-\gamma\delta^2} \to 0$ and $e^{\gamma(\alpha^2-\delta^2)} \to 0$ so that by doing a bit of rearranging (and multiplying by $e^{-\gamma(\alpha^2-\delta^2)}/e^{-\gamma(\alpha^2-\delta^2)}$):

$$\lim_{\gamma\to\infty} G_\gamma(\alpha) = \lim_{\gamma\to\infty} \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma\alpha^2}\left(1+e^{-\gamma\delta^2}\right)e^{-\gamma(\alpha^2-\delta^2)}}{1+e^{-\gamma(\alpha^2-\delta^2)}}\right) = \frac{1}{2\gamma}\ln\left(e^{\gamma\delta^2}\right) = \frac{\delta^2}{2}.$$

Finally, when $\alpha = \delta$, $e^{\gamma(\alpha^2-\delta^2)} = 1$ so that:

$$\lim_{\gamma\to\infty} G_\gamma(\delta) = \lim_{\gamma\to\infty} \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma\delta^2}+1}{1+1}\right) = \lim_{\gamma\to\infty} \frac{1}{2\gamma}\left(\ln\left(e^{\gamma\delta^2}\right) - \ln(2)\right) = \frac{\delta^2}{2}.$$

Thus, in total, we've found that:

$$\lim_{\gamma\to\infty} G_\gamma(\alpha) = \begin{cases}\alpha^2/2 & |\alpha| < \delta,\\ \delta^2/2 & |\alpha| \geq \delta.\end{cases}$$

as expected, since

$$\lim_{\gamma\to\infty} w_\gamma(x,y) = \begin{cases} 1 & |y-x| < \delta \\ 0 & |y-x| \geq \delta \end{cases}.$$

Thus, we can see that whenever $\gamma \neq 0$, G is bounded:

$$\begin{aligned} G_\gamma(\alpha) &= \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma(\alpha^2-\delta^2)} + e^{\gamma\alpha^2}}{e^{\gamma(\alpha^2-\delta^2)} + 1}\right) \\ &< \frac{1}{2\gamma}\ln\left(\frac{e^{\gamma(\alpha^2-\delta^2)} + e^{\gamma\alpha^2}}{e^{\gamma(\alpha^2-\delta^2)}}\right) = \frac{1}{2\gamma}\ln\left(1 + e^{\gamma\delta^2}\right). \end{aligned}$$

so that the dominated convergence theorem applies! On the other hand if $\gamma = 0$, the dominated converge theorem applies whenever our domain is bounded (for example to $x \in [-c, c]$), since $G(x) = x^2/4 \in [0, c^2/4]$ is bounded.

With G and F together, we've derived an expression for our gradient flow equation! □

Now, all that's left is to figure out which (function) space we're working with.

3.2 G is for Grappling with the Wasserstein Metric

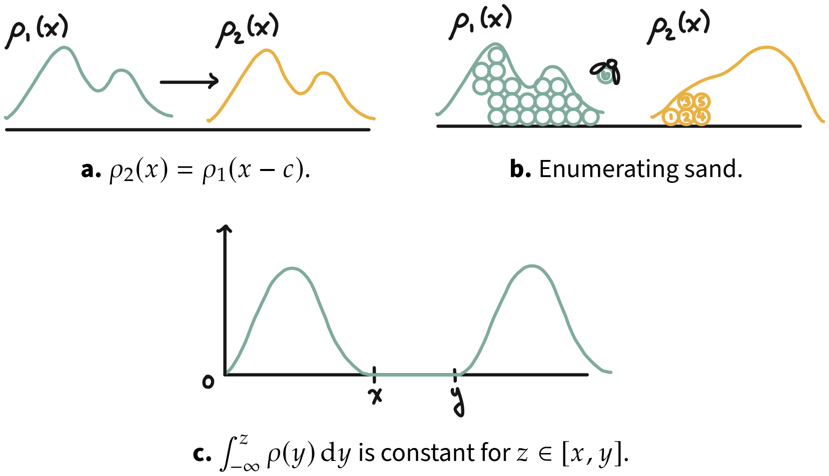
In order to construct this new space, we can take a look at a few key properties it needs to satisfy in order to be able to support a gradient flow. In order to write down the typical gradient flow equation eq. (3.2), we used the following properties of (functions on) real numbers:

- We can take derivatives of (absolutely) continuous curves:

$$|\dot{x}(t)| = \lim_{h\to 0} \frac{\|x(t+h) - x(t)\|}{|h|}$$

- For a differentiable function $F : \mathbb{R}^d \to \mathbb{R}$, we can define a gradient:

$$\begin{aligned} |\nabla F(x(t))| &= \limsup_{y\to x(t)} \frac{|F(y) - F(x(t))|}{\|y - x(t)\|} \\ &= \lim_{\varepsilon\to 0}\left(\sup\left\{\frac{|F(y) - F(x(t))|}{\|y - x(t)\|} : y \in B(x(t), \varepsilon)\backslash\{x(t)\}\right\}\right) \end{aligned} \tag{3.7}$$



a. $\rho_2(x) = \rho_1(x - c)$.

b. Enumerating sand.

c. $\int_{-\infty}^{z} \rho(y)\,\mathrm{d}y$ is constant for $z \in [x, y]$.

Figure 3.2 Visualizations of the notion of distance between two density distributions using sand piles.

where sup is the supremum (see footnote 6), and $B(a, r)$ is the ball of radius r centered at a. As usual, this gradient gives the direction of steepest descent.

Is there a space we can work with which makes our continuity equation a gradient flow problem? In this space, we'd like $\rho(x)$ to be the "objects" we're working with, so that at a specified time t, we can obtain $\rho(x, t)$. Let X be the space of ρ-like objects. Since the objects in X are similar to ρ in some sense, we'll require for any $\rho \in X$ that:

- ρ is always nonnegative so that $\rho : \mathbb{R} \to \mathbb{R}_{\geq 0}$.
- the mass of ρ stays constant, i.e. $\int_{\mathbb{R}} \rho(x)\,\mathrm{d}x = 1$.

Then, in order to take derivatives, we need a notion of distance between two density functions. How can we define this? Suppose we have two density functions ρ_1 and ρ_2. And since our continuity equation has a physical interpretation, let's imagine $\rho_1(x)$ and $\rho_2(x)$ physically, e.g. as two piles with equal amounts of sand. To gain some intuition, let's consider the following special case.

Example 3.2.1 (Densities with a constant shift.)**.** If $\rho_2(x) = \rho_1(x - c)$ for some constant c as in fig. 3.2a, in order to reshape pile 1 to look like pile

2 (corresponding to ρ_1 and ρ_2 respectively), we just have to move $\rho_1(x)$ a distance of c. Perhaps we'd like to define $d(\rho_1(x), \rho_1(x-c)) = |c|$. As a sanity check, we can see that $d(\rho_1(x), \rho_1(x)) = 0$ (which we always expect from a metric).

Then, more generally, we can define $d(\rho_1, \rho_2)$ as the amount of work it takes to move the grains of sand in pile 1 so that it resembles pile 2[5]. Now, how might we enumerate these grains of sand? If we imagine building our piles of sand from left to right, bottom to top (as in fig. 3.2b), we always know that the ith grain of sand has $i-1$ grains of sand behind it, i.e. index each grain of sand by the mass of sand we've already placed. To formalize this, let's index our grains of sand by a real number $z \in [0,1]$. Then, we might try to define a function $u(z)$ which gives the x-position of the zth grain of sand with:

$$\int_{-\infty}^{u(z)} \rho(y)\,\mathrm{d}y = z$$

However, we need to be careful in this definition, because as in fig. 3.2c, it might be possible that there are multiple x-values for which $\int_{-\infty}^{x} \rho(y)\,\mathrm{d}y = z$. In order to make $u(z)$ well-defined, we'll just define $u(z)$ to be the *smallest* (infimum, see footnote 9) x-value for which the above equation is satisfied, so that

$$u(z) = \inf\left\{x \in \mathbb{R} : \int_{-\infty}^{x} \rho(y)\,\mathrm{d}y > z\right\}. \tag{3.8}$$

We call this the *pseudo-inverse*[6]. Now, integrating over the contribution from

[5] By default, we might define $d(\rho_1, \rho_2)$ to be the distance between the heights of the two functions (for those who are familiar, this is the L^1 norm), i.e.

$$d(\rho_1, \rho_2) = \int_{\mathbb{R}} |\rho_2(x) - \rho_1(x)|\,\mathrm{d}x.$$

However, this distance doesn't contain a physical interpretation in the same way as the one we derive. In particular, if $\rho(x) = 0$ for $|x| > R$, we know that $\rho(x+2R)$ is zero wherever $\rho(x)$ is nonzero, and $d(\rho(x), \rho(x+2R)) = 2\int \mathbb{R}\rho(x)\,\mathrm{d}x = 2$, which does not depend on the transport distance $2R$. Very roughly, since our problem involves the movement of our density, we'd like for this information to also factor in to our distance metric.

[6] For those who have seen a bit of statistics, it's the sort-of inverse of the cumulative distribution function of our density distribution $\Phi(x) = \int_{-\infty}^{x} \rho(y)\,\mathrm{d}y$ where $(u \circ \Phi)(x) = x$ as long as $\Phi(x)$ is bijective. Since we require that ρ integrates to one, $\Phi(x)$ is always surjective, so when Φ is not bijective it is not injective. These cases are accounted for by the inf in the definition of $u(z)$ which guarantees that $u(\Phi(x))$ gives the unique "smallest" x mapped to $\Phi(x)$.

each grain of sand, we find that:

$$d(\rho_1, \rho_2) = \int_{[0,1]} |u_2(z) - u_1(z)| \, \mathrm{d}z. \tag{3.9}$$

Originally formulated by Kantorovich and Rubenstein, this metric is often called the Kantorovich-Rubenstein distance, which is part of a more general class of metrics called *Wasserstein metrics*.

To verify that this distance metric matches our intuition, consider again the case that $\rho_2(x) = \rho_1(x - c)$. In this case, we can see that by changing our variable of integration:

$$\begin{aligned}
u_2(z) &= \inf\left\{x \in \mathbb{R} : \int_{-\infty}^{x} \rho_1(y - c) \, \mathrm{d}y > z\right\} \\
&= \inf\left\{x \in \mathbb{R} : \int_{-\infty}^{x-c} \rho_1(y') \, \mathrm{d}y' > z\right\} \\
&= \inf\left\{(x - c) \in \mathbb{R} : \int_{-\infty}^{x-c} \rho_1(y') \, \mathrm{d}y' > z\right\} + c \\
&= \inf\left\{x' \in \mathbb{R} : \int_{-\infty}^{x'} \rho_1(y') \, \mathrm{d}y' > z\right\} + c \\
&= u_1(z) + c.
\end{aligned}$$

Thus,

$$d(\rho_1, \rho_1(x - c)) = \int_{[0,1]} |u_1(z) + c - u_1(z)| \, \mathrm{d}z = |c|,$$

as desired :)

3.2.1 A Little Context

As it turns out, this problem of "how do we minimize the cost of transporting piles of sand" inspired a field of math (aptly) called optimal transport (to read more, see Villani (2009b)). This particular metric has also found its way into probability by reframing density funtions as probability distributions: real-valued probability measure $\mu \in \mathcal{P}(\mathbb{R})$, we can see that both conditions

- μ is always nonnegative so that $\mu : \mathbb{R} \to \mathbb{R}_{\geq 0}$, and
- the total probability is always 1, i.e. $\int_{\mathbb{R}} \mu(x) \, \mathrm{d}x = 1$

are satisfied. Then, we can interpret the pseudo-inverse is the "inverse" of the cumulative distribution function:

$$\Phi(x) = \int_{-\infty}^{x} \mu(y)\,\mathrm{d}y.$$

In this space, the Wasserstein metric is also a bit more complicated (see Burger et al. (2008) for the gory details), and what I called the Wasserstein metric above is actually a simplifed version of it. First, like the typical norm endowed on function spaces

$$\|f(x)\|_{L^p(\mathbb{R})} = \int_{\mathbb{R}} |f(x)|^p \,\mathrm{d}x,$$

the Wasserstein metric is actually a class of metrics also indexed by p so that in the one-dimensional case, i.e. for $\mu_1, \mu_2 \in \mathcal{P}(\mathbb{R}^1)$,

$$w_p(\mu_1, \mu_2) = d_p(\mu_1, \mu_2) = \|u_2(x) - u_1(x)\|_{L^p(\mathbb{R})}, \tag{3.10}$$

where w_1 is the Kantorovich-Rubenstein distance[7]. In higher dimensions, i.e. in $\mathcal{P}(\mathbb{R}^d)$ where $d > 1$, we can't always simplify down to the form given in eq. (3.9), and instead must consider something considerably funkier (also colled the Monge-Kantorovich problem, which we won't get into, but again, Burger et al. (2008) is a great resource if you're interested).

Now, I was originally also going to give a bit of history in this subsection, but I realized that I cannot do better than the Fields medalist, politician, and man who inspired many of my haircuts, Cédric Villani. If you're interested in bakeries and the lore behind the many characters (including Monge, Kantorovich, Wasserstein, and Rubenstein) involved in this chapter, take a look at Villani (2009a).

3.3 H is for the (Gradient Flow) Hammer

Now that we've reformulated our system as a gradient flow problem, we can use it as a tool to answer our questions from the previous chapter. Returning to eq. (2.3) which describes how a particular opinion x_i changes, we can use our new way of enumerating opinions as sand grains using our pseudoinverse $u(z)$ from eq. (3.8). Thus, by carefully replacing x's with

[7] In particular, the norm induced associated with the w_1 metric.

$u(z)$'s using the correspondence from eq. (3.10) we obtain:

$$\begin{aligned}\dot{x} &= -\int_{\mathbb{R}} \rho(y) G'(x-y)\,\mathrm{d}y + \int_{\mathbb{R}} \rho_{\mathcal{Z}}(y) G'(x-y)\,\mathrm{d}y \\ \frac{\partial u(z)}{\partial t} &= -\int_0^1 G'(u(z,t)-u(\zeta,t))\,\mathrm{d}\zeta + F'(u(z,t)) \\ \frac{\partial u(z)}{\partial t} &= -\int_0^1 G'(u(z,t)-u(\zeta,t))\,\mathrm{d}\zeta - \int_0^1 G'(u(z,t)-u_{\mathcal{Z}}(\zeta))\,\mathrm{d}\zeta.\end{aligned}$$

Now, we'll show that a solution exists (and is unique) by using a neat tool from PDEs called the contraction mapping theorem (or Banach fixed-point theorem). First, we'll pretend that this is an ODE and integrate both sides with respect to t in hopes of solving for $u(z)$:

$$u(z,t) = u(z,0) - \int_0^t \int_0^1 G'(u(z,s)-u(\zeta,s)) + G'(u(z,s)-u_{\mathcal{Z}}(\zeta))\,\mathrm{d}\zeta\,\mathrm{d}s.$$

Now, following Burger et al. (2008) we'll describe the right side as the operation of a linear operator $\mathcal{T}$ on u so that

$$(\mathcal{T}u)(z,t) = u(z,0) - \int_0^t \int_0^1 G'(u(z,s)-u(\zeta,s)) + G'(u(z,s)-u_{\mathcal{Z}}(\zeta))\,\mathrm{d}\zeta\,\mathrm{d}s \tag{3.11}$$

and we're looking for solutions where:

$$\mathcal{T}u = u.$$

As the contraction mapping principle alludes, if we can show that $\mathcal{T}$ is a contraction map, i.e. that $\|\mathcal{T}u - \mathcal{T}v\| \le \lambda\|u-v\|$ where $\lambda \in [0,1)$, then $\mathcal{T}$ *must* map exactly one unique point to itself. Intuitively, we can imagine that if we are on the surface of the Earth and we pull out a map of the Earth, at least one point on the map is exactly on top of the corresponding point on Earth. Thus, we'll proceed by showing that $\mathcal{T}$ is a contraction mapping in the space of bounded functions on $[0,1]\times[0,T]$ which we'll call $X = L^\infty([0,1]\times[0,T])$ for some fixed T:

$$\begin{aligned}\|\mathcal{T}u - \mathcal{T}v\|_X = \Bigg\| u_0(z) - u_0(z) - \int_0^T \int_0^1 &(u(z,s)-u(\zeta,s))w_\gamma(u(z,s),u(\zeta,s)) \\ &- (v(z,s)-v(\zeta,s))w_\gamma(v(z,s),v(\zeta,s)) \\ &+ (u(z,s)-u_{\mathcal{Z}}(\zeta,s))w_\gamma(u(z,s),u_{\mathcal{Z}}(\zeta,s) \\ &-(v(z,s)-u_{\mathcal{Z}}(\zeta,s))w_\gamma(v(z,s),v(\zeta,s))\,\mathrm{d}\zeta\,\mathrm{d}s\Bigg\|_X .\end{aligned}$$

Note that I've set $u_{\mathcal{Z}}$ and $u_0(z)$ to be the same for v, since the zealot distribution and intial condition are part of the setup of our system. Now, since $\int_0^T f(t) \leq T \sup_{t\in[0,T]}\{f(t)\}$ and $w_\gamma \leq 1$ we have an upper bound:

$$\begin{aligned}\|\mathcal{T}u - \mathcal{T}v\|_X &\leq \left\| T \int_0^1 2(u(z,s) - v(z,s)) - (u(\zeta,s) - v(\zeta,s))\,\mathrm{d}\zeta \right\|_X \\ &\leq 3T\|u - v\|_X.\end{aligned}$$

Thus, for $T < \frac{1}{3}$, $\mathcal{T}$ is a contraction mapping as desired, which gives the existence of a unique solution up to time T. However, we can extend this unique solution by setting an new initial condition for $t = T$ and repeating the process. Thus, we've completed the first step of showing that a solution exists and is unique!

In order to make sure that this solution is indeed still a pseudo-inverse, we need to make sure that it is monotonic. Again following Burger et al. (2008), we can keep track of the slope $\lambda(h,t) = u(z+h,t) - u(z,t)$ where $z \in [0,1)$ and $h \in (0, 1-z]$. Then, note that since $\lambda(h,0)/h \geq 0$:

$$\begin{aligned}\frac{\partial}{\partial t}\frac{\lambda(h,t)}{h} = -\frac{1}{h}\int_0^1 &(u(z+h,t) - u(\zeta,t))w_\gamma(u(z+h,t), u(\zeta,t)) \\ &- (u(z,t) - u(\zeta,t))w_\gamma(u(z,t), u(\zeta,t)) \\ &+ (u(z+h,t) - u_{\mathcal{Z}}(\zeta,t))w_\gamma(u(z+h,t), u_{\mathcal{Z}}(\zeta,t)) \\ &- (u(z,t) - u_{\mathcal{Z}}(\zeta,t))w_\gamma(u(z,t), u_{\mathcal{Z}}(\zeta,t))\,\mathrm{d}\zeta \\ \geq -\frac{1}{h}&(\lambda(h,t) + \lambda(h,t)) = -2\frac{\lambda(h,t)}{h}\end{aligned}$$

following the same bounding argument from before. Thus, we can see that by rearranging and recognizing a product rule:

$$\frac{\partial}{\partial t}\frac{\lambda(h,t)}{h} + 2\frac{\lambda(h,t)}{h} = \frac{\partial}{\partial t}\left(\frac{\lambda(h,t)}{h}e^{2t}\right) \geq 0.$$

As a result, we can see that $\frac{\partial}{\partial t}\frac{\lambda(h,t)}{h}$ is always positive and since $\lambda(h,0)/h \geq 0$, we know that $\lambda(h,t)/h \geq 0$ for any $t > 0$, as desired.

Finally, to show that our system will not converge to δ-distributions in finite time, we'll start by assuming that our initial condition is *Lipschitz continuous*, which is to say that our u_0 must be more than uniformly continuous—there must be some bound $K \geq 0$ such that:

$$|u_0(x) - u_0(y)| \leq K|x - y|$$

for all $x, y \in [0,1]$. Note that this does not allow ρ to be composed exclusively of sums of δ-distributions, as this would force $u_0(x)$ to contain a discontinuity. Now, we can see that (using similar reasoning as above):

$$\frac{\partial}{\partial t}|u(x,t) - u(y,t)| \leq |u_0(x) - u_0(y)| + 2|u(x,t) - u(y,t)|$$

so that:

$$|u(x,t) - u(y,t)| \leq K|x - y|(1+t)e^{2t}$$

so that for any fixed t, $u(x,t)$ is also Lipschitz continuous, so we can count there being no δ-distributions in $\rho(x,t)$ in finite time.

Additionally, we can count on δ distributions being stationary states, as whenever $\rho = \delta(x-c)$, we can see that $u(x) = c$ and as long as $u_{\mathcal{Z}} = c$, we also have:

$$\frac{\partial u(x,t)}{\partial t} = -\int_0^1 (c-c)w_\gamma(c,c) - \int (c - u_{\mathcal{Z}})w_\gamma(c,u_{\mathcal{Z}}) = 0,$$

as desired. Additionally, we can see that whenever we don't have a zealot distribution, the only distributional steady states are ones which are constant.

3.4 In Closing

We started this section by making a simplification to our model in eq. (3.1) which allowed us to reinterpret our model so that we can apply tools from a field of math called gradient flows. Then, after constructing the space we are working in and building intuition for objects like the Wasserstein metric and the pseudo-inverse, we followed Burger et al. (2008) in deriving a new differential equation to work with. By analyzing the system with our new tools, we confirm that our system admits the existence of a unique (well-behaved) solution, that steady states can look like sums of δ-distributions, and that in the case where there are no zealots, the only δ-distribution steady state is a consensus state (with the entire population coalesced in one δ-function).

Chapter 4

The Z of Conclusions and Future Work

We made it (woohoo!!)! To celebrate, here's another "connections quandry" (solution in the footnote)[1]:

SBCM	MEAN FIELD THEORY	GRADIENT FLOWS	WASSERSTEIN METRIC

To summarize, in chapter 1 we were introduced to graphs, the Abelson and HK models, and their connection to the sigmoidal bounded confidence model (SBCM). Then in chapter 2 we introduced the mean-field approximation and derived the mean-field SBCM. Using simulations, and mathematical arguments in the discrete case, we developed some intuition for how our system should behave, and what its steady states should look like. Finally in chapter 3, after making a simplification to the mean-field SBCM, we reframe it as a gradient flow problem, and we showed that a simplified version of our model indeed meets the expectations from chapter 2 and more!

Before I start talking about future work, I'd like to share that in writing this section, one quote kept coming to mind, which I first heard in a conversation between Terence Tao and Steven Strogatz on Strogatz's podcast *The Joy of Why* in an episode titled "What makes for 'Good' Mathematics?". There, Tao paraphrases from a Math Overflow reply by Minhyong Kim, which reads:

> [It]'s almost as though definite mathematical results are money in the bank. After you've built up some savings, you can afford

[1] The thing which unites these four phrases is, of course, my thesis!

> to spend a bit by philosophizing. But then, you can't let the balance get too low because people will start looking at you in funny, suspicious ways.

Tao later posted on Mathstodon that he "sometimes [wonders] if we should have more spaces to encourage mathematical speculation". So though I have certainly not accumulated "mathematical currency" in any way, I'll follow in this spirit and do a bit (a lot) of speculating.

4.1 Z is for the Zillion Questions I'd Like to Explore

First, it would be insightful to continue our exploration of the gradient-flow version of our model. For example, we've shown that δ distributions are indeed steady states, but are they the *only* steady states? Additionally, it would be interesting to investigate the stability of our steady states—in the two zealot case, for which γ, δ is it more "natural" for our agents to become polarized rather than coming to a consensus?

4.1.1 Code

On the more computational side, it might be interesting to run more discrete SBCM simulations with varying proportions of zealots and with different numbers of zealot opinions. Does the parity of the number of zealot opinions affect the steady states of the system? How does the proportion of zealots to persuadable nodes affect the rate of convergence to a steady state or other properties of the system?

Additionally, it would be very useful to be able to numerically calculate solutions to the mean-field SBCM for experimentation purposes. I briefly tried to do this with Julia but did not get very far—perhaps it would be more possible to code up in Mathematica, then port over to python/Julia. With a mean-field SBCM simulator, it would be much easier to test hypotheses and check our mathematical analysis.

Also, it could be helpful overall to make sure that our simulations aren't too greatly affected by possible floating point errors. Though $w_\gamma(x, y)$ is always positive whenever $x \neq y$, it can get very small whenever $|x - y| > \delta$, and these small values may be neglected in the simulation process. If the mean-field SBCM is sensitive to perturbations, these floating point errors could effect relatively large changes in the model's behavior.

4.1.2 Generalizations

Though we abandoned the graph structure quite early on, it would be exciting to bring it back and see what it adds to the mix. We think we have a good idea of what the mean-field limit of the SBCM (including graph structure) looks like—following the explanation in Lovász (2012) and taking inspiration from Chiba and Medvedev (2018), we would introduce an infinite parameterization of a graph called a *graphon* and an extra parameter keeping track of each node's "index" so that:

$$v(x, \zeta, t) = \frac{\int_0^1 \int_{\mathbb{R}} \rho(y, \xi, t)(y - x) w_\gamma(x, y) W(\xi, \zeta) \, \mathrm{d}\xi \, \mathrm{d}y}{\int_0^1 \int_{\mathbb{R}} \rho(y, \xi, t) w_\gamma(x, y) W(\xi, \zeta) \, \mathrm{d}\xi \, \mathrm{d}y}$$

where $W(\xi, \zeta)$ is our graphon, and $\xi \in [0, 1]$ is our node index. It might be interesting to try to replicate analysis from the discrete SBCM in Brooks et al. (2023) for the mean-field case, and determine ways in which they differ.

I'm also quite intrigued by the many connections to physics which exist in this model. First, as I mentioned, a sigmoid is a Fermi–Dirac distribution. By interpreting $(x - y)^2 \mapsto \varepsilon$ as an energy, $\delta^2 \mapsto \mu$ as our particle's chemical potential, and $\gamma \mapsto 1/(k_B T)$ as the thermal energy in our system,

$$w_\gamma(x, y) = \frac{1}{1 + e^{\gamma((x-y)^2 - \delta^2)}} \mapsto \frac{1}{1 + e^{(\varepsilon - \mu)/k_B T}} = \bar{n}_\varepsilon$$

which gives the *number density* of fermions at a particular energy ε. Thus $(y - x) w_\gamma(x, y) = \sqrt{\varepsilon} \bar{n}_\varepsilon$. Funnily enough the density of states of particles in three spatial dimensions $D(\varepsilon) \sim \sqrt{\varepsilon}$ scales the same way as $\sqrt{\varepsilon}$ so that integrating gives the total number of particles in three dimensions (up to a constant factor). Similarly, $\int w_\gamma(x, y)$ gives (up to a constant) the total number of fermions in a two-dimensional system. We could also interpret this as finding the expected value of $\sqrt{\varepsilon}$ in a 2D fermion gas.

One more physical connection is the ability to keep track of phase transitions. Brooks et al. (2023) included some analysis which plotted the number of linearly stable steady states against γ and δ for a variety of graph structures, which I feel would be interesting to do here as (yet another) a way to visualize our system. In statistical physics, when we are looking for a phase transition, we often do so by keeping track of some *order parameter* which indicates using its behavior (e.g. how fast it grows) the phase we're currently in. We could explore the possiblity of analyzing this phase parameter to see what it has to say about the mean-field SBCM.

Finally, of course a natural extension of our gradient flow explorations is to try to do the same things with the full-fledged mean-field SBCM (complete with a nonconstant denominator). Of course, this is easier said than done—I spent much of this year trying to do just this, but made headway mostly in tangential directions. Some things we tried, like applying analysis by Crawford and Davies (1999) for the *generalized Kuramoto model* (see appendix B) also had complications arising from this varying denominator, as it relies on the Fourier transform of our velocity v, which we can find easily for convolutions (like the numerator of v) but not so easily for quotients of convolutions. However, this model feels so natural and physical that it feels like this system should be solvable, at least for a uniformly distributed initial condition.

That's all, folks![2]

[2] Time flies like an arrow; fruit flies like a banana.

Appendix A

A Brief Connection to δ-Distributions

Suppose we have a point particle located at the origin with mass m, as we often do when we set up a physics problem. How would we describe the density ρ of this particle as a function of x? In order to fully describe this point particle, we would like this ρ to have the following properties:

1. Since the particle has mass m, we need $\int_{-\infty}^{\infty} \rho(x)\,\mathrm{d}x = m$. In particular, since the particle is localized at $x = 0$, we expect:

$$\int \rho(x)\,\mathrm{d}x = \begin{cases} 0 & x < 0 \\ m/2 & x = 0 \\ m & x > 0. \end{cases}$$

2. Since the particle is a point particle, we also need $\rho(x) = 0$ for all $x \neq 0$.

However, these two requirements seem to contradict each other. Whenever we have a function that is nonzero at a finite number of points, the integral misses the "blips" in the function, and treats the function as if it were continuous. Intuitively, this happens because we're always integrating over some set with measurable width, while our blip occurs only at a point—a set with zero width[1].

[1] To learn more, check out measure theory (especially if you enjoyed/want to learn more about real analysis)! It's one of my favorite undergrad math courses :)

Thus, if $\rho(x) = 0$ for all $x \neq 0$, we have a problem, since this would imply that $\int_{-\infty}^{\infty} \rho(x)\,\mathrm{d}x = 0$[2]. As a result, we're forced to reconsider our assumptions. It's impossible for ρ to be a *function* which satisfies these parameters. To address this problem, let's try to construct a $\rho(x)$ which satisfies the three conditions above (following the construction in Howison (2020)).

First, let $\varphi : \mathbb{R} \to \mathbb{R}$ be a continuous function. By the definition of continuity, for any $\varepsilon > 0$, there exists some $\Delta > 0$ such that:

$$-\Delta < x < \Delta \implies \varphi(0) - \varepsilon < \varphi(x) < \varphi(0) + \varepsilon.$$

Additionally, since $\rho(x) = 0$ whenever $x \neq 0$:

$$\begin{aligned}\int_{-\infty}^{\infty} \rho(x)\varphi(x)\,\mathrm{d}x &= \int_{-\infty}^{-\Delta} \rho(x)\varphi(x)\,\mathrm{d}x + \int_{-\Delta}^{\Delta} \rho(x)\varphi(x)\,\mathrm{d}x \\ &\quad + \int_{\Delta}^{\infty} \rho(x)\varphi(x)\,\mathrm{d}x \\ &= \int_{-\Delta}^{\Delta} \rho(x)\varphi(x)\,\mathrm{d}x.\end{aligned}$$

Additionally, since[3]

$$\varphi(0) - \varepsilon \leq \min_{x\in[-\Delta,\Delta]}(\varphi(x)) \leq \varphi(x) \leq \max_{x\in[-\Delta,\Delta]}(\varphi(x)) \leq \varphi(0) + \varepsilon,$$

we can see that multiplying by our mass m:

$$\begin{aligned}(\varphi(0) - \varepsilon)m &= \int_{-\Delta}^{\Delta} \rho(x)(\varphi(0) - \varepsilon)\,\mathrm{d}x \\ &\leq \int_{-\Delta}^{\Delta} \rho(x)\varphi(x)\,\mathrm{d}x \leq \int_{-\Delta}^{\Delta} \rho(x)(\varphi(0) + \varepsilon)\,\mathrm{d}x = (\varphi(0) + \varepsilon)m\end{aligned}$$

Thus, in the limit $\varepsilon \to 0$, we obtain:

$$\int_{-\infty}^{\infty} \rho(x)\varphi(x)\,\mathrm{d}x = \int_{-\Delta}^{\Delta} \rho(x)\varphi(x)\,\mathrm{d}x = m\varphi(0).$$

[2] If you've taken a course in quantum physics, perhaps you've seen the δ-distribution defined with $\delta(x) = \begin{cases} \infty & x = 0 \\ 0 & \text{elsewhere} \end{cases}$ where the ∞ is "just the right amount" to make the integral evaluate to one.

[3] For those who are familiar with supremums and infimums (e.g. you've taken a course in real analysis), note that since $\varphi(x)$ is continuous and $[-\Delta, \Delta] \subset \mathbb{R}$ is closed, the supremum and infimum of $\varphi(x)$ are achieved on this interval, so it's okay to use max and min in place of sup and inf (respectively) here.

In this way, instead of a function, perhaps we can consider ρ a generalized function, which is defined by the evaluation of its integral with other functions. Letting $\rho(x) = m\delta(x)$ so that $\int_{-\infty}^{\infty} \delta(x)\,\mathrm{d}x = 1$, we can see that $\delta(x)$ is fully determined by its behavior in the following *functional*, a map from (real-valued) continuous functions to real numbers:

$$\varphi \mapsto \int_{-\infty}^{\infty} \delta(x)\varphi(x)\,\mathrm{d}x = \varphi(0).$$

We think of this as a *generalized function*, since for any function f, we can define a corresponding functional F_f with:

$$F_f(\varphi) = \int_{-\infty}^{\infty} f(x)\varphi(x)\,\mathrm{d}x.$$

In our case, we can think of δ as the *functional* defined with $\delta(\varphi) = \varphi(0)$.

In general, a *distribution* F must satisfy a couple of other properties, in addition to being a functional. In particular, in order for us to be able differentiate them just like we can for functions, their domain is the space of *test functions*, i.e. smooth (infinitely-differentiable) functions which have a bounded domain. Also, it must be linear so that $F(a\varphi + b\psi) = aF(\varphi) + bF(\psi)$ for $a, b \in \mathbb{R}$ and test functions φ, ψ. Finally, we require that they be continuous in the following sense: if $\varphi(x)$, $\varphi_n(x)$ are test functions where the kth derivatives $\varphi_n(x)^{(k)}$ converge (uniformly) to $\varphi(x)^{(k)}$ as $n \to \infty$, then $F(\varphi_n)$ converges to $F(\varphi)$. However, these mathematical details aren't important for our application, as we only deal with δ-distributions. However, do check out Howison (2020) if you're curious!

Appendix B

Fields of Fireflies and the Kuramoto Model

If you have not yet witnessed the synchronization of fireflies' flashes, you should watch this video. It's a very spectacular sight! If we want to explore what happens with different types/populations of fireflies, but don't have immediate access to a large amount of fireflies, we can instead model the system mathematically and explore the results when we change our initial conditions. In order to find a mathematical model, we need to consider what qualities we'd like it to have:

- First, we want to focus on the effects of the community of fireflies on only one firefly. By keeping track of the frequency at which each firefly flashes, and when each firefly flashes inside the period of oscillation (i.e. the "phase offset"), we can describe the change in when they flash (their flashing "phase") using an update rule of the form $x_{n+1} = f(x_n)$), so that the new flashing time is influenced by when the neighbors flash.

- Since we'd like to encourage the fireflies to flash at the same time, we'd like the update rule to depend on when the fireflies' neighbors flash. If one of firefly A's neighbors, firefly B, is already flashing in sync with firefly A, it likely won't try to change its flashing time based on firefly B's flashing time. On the other hand, if firefly B is flashing at a time which is offset from when firefly A flashes, then we might expect firefly A to be influenced by firefly B and try to move toward it, so that firefly B contributes to firefly A's update rule. Conveniently, there is a

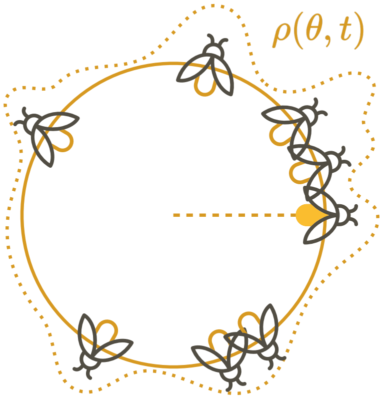


Figure B.1 A visualization of the $N \longrightarrow \infty$ limit. By taking the limit as the number of fireflies goes to infinity, it's no longer possible to keep track of individual fireflies. Instead, we use a density function $\rho(\theta, t)$ to describe the proportion of fireflies flashing near some frequency at a point in time.

function which does just this: $\sin(x)$ is large when $x = \pi/2$ and small when $x = 0$!

B.1 The Kuramoto Model

Yoshiki Kuramoto (1984) was inspired by similar processes that arise in nature, and proposed the Kuramoto model to mathematically describe them. His update rules is given by the following function:

$$\dot{\theta}_i(t) = \omega_i + \frac{1}{N} \sum_{j=1}^{N} \sin(\theta_j(t) - \theta_i(t)), \tag{B.1}$$

where N is the total number of fireflies, θ_i is the *(flashing) phase* of firefly i, and ω_i is the *(flashing) frequency* of firefly i. We can imagine each firefly in the collection flying in tiny circles with frequency ω_i, and flashing only when they reach $\theta = 0$ along their path, as shown in fig. B.1. In order for them to synchronize, we'd like for them to all flash at the same time, so that $\theta_j(t) - \theta_i(t) = 0$ (thus $\sin(\theta_j(t) - \theta_i(t)) = 0$). If firefly j is flashing ahead of firefly i, then $\theta_j(t) > \theta_i(t)$ so that $\sin(\theta_j(t) - \theta_i(t)) > 0$, and firefly i will speed up to catch up to firefly j.

Though this system looks quite complex (it's a system of coupled nonlinear differential equations), we can simplify it by using a neat trick. Recall that by Euler's identity, $e^{i\theta} = \cos(\theta) + i \sin(\theta)$. Then, treating each firefly as a complex number on the unit circle, we can consider the sum:

$$r(t)e^{i\psi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_i(t)} \tag{B.2}$$

Note that if the fireflies are synchronized, $r(t) = 1$ and $\psi(t) = \theta_i(t)$, and on the other hand, when the fireflies are all flashing randomly, we expect that the $e^{i\theta_i(t)}$ terms sum to zero so that $r(t) = 0$. Thus, $r(t) \in [0, 1]$ represents how aligned the fireflies are with one another, and $\psi(t)$ gives an average flashing phase for the fireflies. These quantities are called *order parameters*—they're often used in physics as indicators of phase transitions. For example, as ice melts, we move from a highly ordered state (where the molecules satisfy a repeating pattern) to a disordered state (where the molecules have much more freedom for movement). Similarly, in the Kuramoto model, there is an ordered phase where all fireflies are synchronized, and a disordered phase where the fireflies are flashing at random[1].

Using eq. (B.2), we can rewrite our original differential equation (eq. (B.1)) in the following way:

$$\dot{\theta}_i(t) = \omega_i + \mathrm{Im}\left(e^{i(\psi(t)-\theta_i(t))}\right) = \omega_i + \sin(\psi(t) - \theta_i(t))$$

where $\mathrm{Im}(a + bi) = b$ gives the imaginary part of a complex number. Now, our equation looks much simpler—the dependence on the other fireflies is all encoded in the ψ parameter. Though the dependence is hidden, it's still there—to further simplify (and solve the system), we turn to a powerful tool called the *mean field approximation*. The difficulty with graph structure dependence lies in the wide variety of different types of graphs. Especially in the finite case, it's often difficult to make use of symmetry to simplify our analysis. Luckily, physicists have been dealing with complications due to extra degrees of freedom for a long time—in statistical mechanics, the study of large aggregations of particles, physicists developed a tool called the *mean field approximation*, which reduces the complexity of our system by taking the number of particles to infinity and averaging over degrees of freedom. In particular, a field of inquiry which studies *Kuramoto models* (models describing the way certain species of fireflies synchronize their flashes), has used the mean field approximation in a similar context in order to determine its asymptotic behavior. Conveniently, the Kuramoto model is

[1] In fact, the Kuramoto model also describes plasmas, where phase transitions have physical meaning.

very conducive to being placed on a graph, and researchers have studied the mean field limit of the Kuramoto model, specifically on graphs!

B.2 The Mean Field Limit

The Kuramoto model, as we've left it, is still quite difficult to solve, and the number of coupled differential equations we need to solve only gets larger as N increases—a similar predicament to the one we faced for the SBCM. However, if we take $N \to \infty$, our system (maybe surprisingly) becomes exactly solveable! As we will see, this approximation actually reduces the number of differential equations we must solve—intuitively, we are now keeping track of an "average" of sorts, rather than each individual firefly.

Following Strogatz (2000), we first need to find analogues for each of our equations in the mean field limit. Let's first turn to our order parameter, $r(t)e^{i\psi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_i(t)}$. Now, in the limit as $N \to \infty$, our $\frac{1}{N}$ term becomes very small, and we can approximate our sum as an integral! If we return to our distribution of fireflies on the unit circle, and as the number of fireflies gets larger, it gets more and more unwieldy—instead, we can keep track of the density of fireflies $\rho(\theta, t)$. We can also impose the restriction that our density function is normalized, so that $\int_0^{2\pi} \rho(\theta, t)\,\mathrm{d}\theta = 1$ for any time t. Then, eq. (B.2) becomes:

$$r(t)e^{i\psi(t)} = \int_0^{2\pi} e^{i\theta}\rho(\theta, t)\,\mathrm{d}\theta,$$

where we are taking a weighted average of $e^{i\theta}$ for $\theta \in [0, 2\pi)$ based on the density $\rho(\theta, t)$. Now, for eq. (B.1), whenever $\theta(t)$ changes, we are changing the distribution of our density $\rho(\theta, t)$. By replacing $\theta(t) \to \rho(\theta, t)$, we have:

$$v(\theta, t) = \omega(\theta) + \int_0^{2\pi} \sin(\theta' - \theta)\rho(\theta', t)\,\mathrm{d}\theta', \tag{B.3}$$

and ρ itself satisfies the following *continuity equation*:

$$\frac{\partial}{\partial t}\rho(\theta, t) + \frac{\partial}{\partial \theta}(v(\theta, t)\rho(\theta, t)) = 0, \tag{B.4}$$

as shown in fig. 2.3. By taking the limit as $N \to \infty$, we've reduced the number of variables in our system by keeping track of the density of fireflies rather than each individual firefly. A few plots of the evolution of $\psi(t)$ with respect to various initial conditions is shown in fig. 2.3.

Additionally, the Kuramoto model has a convenient feature: in eq. (B.3), we can see that our update function depends on $\sin(\theta' - \theta)$. A pair of mathematicians, Crawford and Davies, noticed that conveniently, Fourier series decompositions depend only on sine functions, and asked the question: can the analysis of the Kuramoto model be extended by replacing $\sin(\theta' - \theta)$ with $f(\theta' - \theta)$ (we'll call this a "coupling function"), assuming f has a Fourier series decomposition? It turns out that this is possible! In Crawford and Davies (1999), they analyzed the asymptotic behavior of the Kuramoto model, generalized with an arbitrary coupling function, by examining the coefficients in the Fourier series expansion of $f(\theta' - \theta)$. Since the coupling function in the SBCM does not depend on sines and cosines, this has potential to be a point of connection between the Kuramoto model and the mean-field SBCM, but needs further investigation.

Bibliography

Abelson, Robert P. 1967. Mathematical Models in Social Psychology. In *Advances in Experimental Social Psychology*, vol. 3, 1–54. Elsevier. doi: 10.1016/S0065-2601(08)60341-X.

Alfano, Andrea. 2013. How Do Starling Flocks Create Those Mesmerizing Murmurations? https://www.allaboutbirds.org/news/how-do-starling-flocks-create-those-mesmerizing-murmurations/.

Bernoff, Andrew J., and Chad M. Topaz. 2011. A Primer of Swarm Equilibria. *SIAM Journal on Applied Dynamical Systems* 10(1):212–250. doi:10.1137/100804504.

Brooks, Heather Z., Philip S. Chodrow, and Mason A. Porter. 2023. Emergence of polarization in a sigmoidal bounded-confidence model of opinion dynamics. 2209.07004.

Burger, Martin, Marco Di Francesco, Martin Burger, and Marco Di Francesco. 2008. Large time behavior of nonlocal aggregation models with nonlinear diffusion. *Networks and Heterogeneous Media* 3(1556-1801_2008_4_749):749–785. doi:10.3934/nhm.2008.3.749.

Chiba, Hayato, and Georgi S. Medvedev. 2018. The mean field analysis of the Kuramoto model on graphs i. The mean field equation and transition point formulas. *Discrete and Continuous Dynamical Systems* 39(1):131–155. doi:10.3934/dcds.2019006.

Crawford, John D., and K. T. R. Davies. 1999. Synchronization of globally coupled phase oscillators: Singularities and scaling for general couplings. *Physica D: Nonlinear Phenomena* 125(1):1–46. doi:10.1016/S0167-2789(98)00235-8.

Cucker, Felipe, and Steve Smale. 2007. Emergent Behavior in Flocks. *IEEE Transactions on Automatic Control* 52(5):852–862. doi:10.1109/TAC.2007.895842.

Fischer, Peter, Julia K. Fischer, Nilüfer Aydin, and Dieter Frey. 2010. Physically Attractive Social Information Sources Lead to Increased Selective Exposure to Information. *Basic and Applied Social Psychology* 32(4):340–347. doi:10.1080/01973533.2010.519208.

Ha, Seung-Yeal, and Jian-Guo Liu. 2009. A simple proof of the Cucker-Smale flocking dynamics and mean-field limit. *Communications in Mathematical Sciences* 7(2):297–325. doi:10.4310/CMS.2009.v7.n2.a2.

Hegselmann, Rainer, and Ulrich Krause. 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation* 5(3).

Howison, Sam. 2020. Integral Transforms Lecture Notes.

Jordan, Richard, David Kinderlehrer, and Felix Otto. 1998. The Variational Formulation of the Fokker–Planck Equation. *SIAM Journal on Mathematical Analysis* 29(1):1–17. doi:10.1137/S0036141096303359.

Keller, Evelyn F., and Lee A. Segel. 1970. Initiation of slime mold aggregation viewed as an instability. *Journal of Theoretical Biology* 26(3):399–415. doi:10.1016/0022-5193(70)90092-5.

Kuramoto, Yoshiki. 1984. *Chemical Oscillations, Waves, and Turbulence*, *Springer Series in Synergetics*, vol. 19. Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-69689-3.

Lovász, László. 2012. *Large Networks and Graph Limits*, *Colloquium Publications*, vol. 60. American Mathematical Society. doi:10.1090/coll/060.

Noorazar, Hossein, Kevin R. Vixie, Arghavan Talebanpour, and Yunfeng Hu. 2020. From classical to modern opinion dynamics. *International Journal of Modern Physics C* 31(07):2050,101. doi:10.1142/S0129183120501016. 1909.12089.

Nugent, Andrew, Susana N. Gomes, and Marie-Therese Wolfram. 2023. On evolving network models and their influence on opinion formation. *Physica D: Nonlinear Phenomena* 456:133,914. doi:10.1016/j.physd.2023.133914.

Strogatz, Steven H. 2000. From Kuramoto to Crawford: Exploring the onset of synchronization in populations of coupled oscillators. *Physica D: Nonlinear Phenomena* 143(1):1–20. doi:10.1016/S0167-2789(00)00094-4.

Villani, Cédric. 2009a. The founding fathers of optimal transport. In *Optimal Transport: Old and New*, ed. Cédric Villani, 29–37. Berlin, Heidelberg: Springer. doi:10.1007/978-3-540-71050-9_3.

———. 2009b. *Optimal Transport*, *Grundlehren Der Mathematischen Wissenschaften*, vol. 338. Berlin, Heidelberg: Springer Berlin Heidelberg. doi:10.1007/978-3-540-71050-9.