Dynamic Games and Applications
https://doi.org/10.1007/s13235-024-00592-5

# Cluster Formation in Iterated Mean Field Games

**P. Jameson Graber[1] · Elizabeth Matter[1] · Rafael Morales[1] · Lindsay North[1]**

Accepted: 26 August 2024


**Abstract**

We study a simple first-order mean field game in which the coupling with the mean field is only in the final time and gives an incentive for players to congregate. For a short enough time horizon, the equilibrium is unique. We consider the process of *iterating* the game, taking the final population distribution as the initial distribution in the next iteration. Restricting to one dimension, we take this to be a model of coalition building for a population distributed over some spectrum of opinions. Our main result states that, given a final coupling of the form $G(x, m) = \int \varphi(x - z)\mathrm{d}m(z)$ where $\varphi$ is a smooth, even, non-positive function of compact support, then as the number of iterations goes to infinity the population tends to cluster into discrete groups, which are spread out as a function of the size of the support of $\varphi$. We discuss the potential implications of this result for real-world opinion dynamics and political systems.

**Keywords** Mean field games · Dynamical systems · Long time behavior · Polarization

## 1 Introduction

Mean field games describe the strategic Nash equilibrium behavior of large (continuum) populations [3, 5–7, 14]. A typical mean field game of first order can be described by a system of PDE as follows:

$$\begin{aligned} -\partial_s u + H(x, \nabla_x u) &= F(x, m), \\ \partial_s m - \nabla_x \cdot \big(D_p H(x, \nabla_x u) m\big) &= 0, \\ m(x, 0) = m_0(x), \quad u(x, t) &= G(x, m(x, t)). \end{aligned} \tag{1.1}$$

Research supported by the National Science Foundation under NSF Grant DMS-2045027.

✉ Elizabeth Matter
Ellie_Carr3@baylor.edu

P. Jameson Graber
Jameson_Graber@baylor.edu

Rafael Morales
Rafael_Morales2@baylor.edu

Lindsay North
Lindsay_North1@baylor.edu

1 Department of Mathematics, Baylor University, Sid Richardson Building, 1410 S. 4th Street, Waco, TX 76706, US

Published online: 09 October 2024

The *Lasry-Lions monotonicity condition* [14] can be defined as

$$\int (F(x, m_1) - F(x, m_2))\,\mathrm{d}(m_1 - m_2)(x) \geq 0 \quad \forall m_1, m_2 \in \mathscr{P}(\mathbb{R}^d). \tag{1.2}$$

It is well-known that when $F$ and $G$ satisfy the Lasry–Lions monotonicity condition, players have an incentive to spread out from one another, and for this reason one can expect uniqueness of the equilibrium on an arbitrarily long time horizon $t$. If we assume, to the contrary, that $F$ and $G$ exhibit an *anti*-monotonicity property, then players have an incentive to *congregate*, i.e. form clusters. It is this latter case which interests us in the present study. In such a case, we do not in general expect uniqueness of the equilibrium, unless the time horizon $t$ is sufficiently small. Cf. [2, 4, 8, 11] for some general results on existence, non-existence, and non-uniqueness under this anti-monotonicity assumption.

As we seek to model cluster formation, a lack of uniqueness for the equilibrium presents a problem. One way around this is to assume the game is played over a sufficiently small time horizon, and then *iterate*, that is, take the final distribution of players to be our new initial distribution, and then repeat this process over many iterations. Such an iterative process can be taken as a model for population dynamics resulting from a string of small strategic decisions on the part of relatively myopic individuals. In the limit as the time horizon converges to zero, it has been shown in [12] that there is a connection between such a process and the "best-reply strategy," in which players choose their moves based on a gradient descent along their own cost function, via what is called "model predictive control." See also [1]. In this paper, our goal is to study the explicit long-term behavior of this iterative process. Assuming the cost to player provides an incentive to congregate, we expect that after sufficiently many iterations, we will see players accumulate in tighter and tighter clusters. Our results provide some conditions under which we can prove that clustering occurs as well as numerical simulations showing the location of the clusters for specific examples.

We are going to keep the focus on the clustering phenomenon described above and not on abstract results for mean field games. For this reason we posit a simple game by making the following assumptions. We will take $F = 0$, so that only the final cost depends on the distribution. In addition, we will assume $H(x, p) = \frac{1}{2}|p|^2$. Then the game becomes equivalent to the following fixed point problem. Define

$$J_t(x, y, m) = \frac{|x - y|^2}{2t} + G(x, m). \tag{1.3}$$

Let $y_t^m(x) := \operatorname{argmin} J_t(x, \cdot, m)$. We will specify below conditions under which $y_t^m(x)$ is a well-defined (single-valued) function. Define $F_{m_0}^t : \mathscr{P}(\mathbb{R}^d) \to \mathscr{P}(\mathbb{R}^d)$ by

$$F_{m_0}^t(m) := y_t^m \sharp m_0, \tag{1.4}$$

where $m_0 \in \mathscr{P}(\mathbb{R}^d)$ is the given initial measure. Then $m$ is a Nash equilibrium if and only if $m = F_{m_0}^t(m)$. The interpretation of this game is that individuals are willing to make a small move from their present state if and only if it will land them in a region of sufficiently high population density to justify the move.

By taking $t$ sufficiently small, we will be able to ensure that $F_{m_0}^t$ has a unique fixed point in the Wasserstein space $\mathscr{P}_1(\mathbb{R}^d)$, defined below. We now define the *equilibrium map* $E^t : \mathscr{P}_1(\mathbb{R}^d) \to \mathscr{P}_1(\mathbb{R}^d)$ to be the map which, for any given initial measure $m_0$, outputs the equilibrium measure $m = F_{m_0}^t(m)$. We then wish to consider the following dynamical system:

$$\begin{aligned} &m_0 \quad \text{is given,} \\ &m_{k+1} = E^t(m_k), \quad k = 0, 1, 2, \ldots \end{aligned} \tag{1.5}$$

The main theoretical contributions of this paper are as follows:

(1) We give a straightforward algorithm to reliably compute $E^t(m)$ by *discretizing the measure*, i.e. by putting an empirical measure in its place.
(2) We study the asymptotic behavior of the dynamical system (1.5). We prove that for dimension $d = 1$ and couplings of the form $G(x, m) = \int \varphi(x - y)\mathrm{d}m(y)$, where $\varphi$ is an even, non-positive, smooth function of compact support, we can explicitly locate the fixed points of $E^t$ and show that they are asymptotically stable.

The motivation behind this analysis is an interest in the implications for human behavior. Humans instinctually like to form groups. People form groups of different sizes for numerous reasons, but specifically, people join social groups to access grouping utility that is otherwise unavailable to lone individuals. Social groups also benefit when new members join. A larger platform could provide the necessary leverage to accomplish a group's goals. Moreover, group membership transforms individuals by providing them with a group identity and inviting members to take part in a shared belief system, and individuals affect their groups by adopting organizational roles and relating to other members. How people group and their groups' actions can completely change a population's status quo and reshape society at large. Cf. [9, 10, 13, 16]. Given the impact of grouping behavior on society, how does a population arrange itself considering only the individual proclivity to group?

The results of this paper show that a population may "cluster" into multiple discrete groups, despite having no aversion whatsoever to one another. Specifically, assume that $G(x, m) = \int \varphi(x-y)\mathrm{d}m(y)$ is as in point (2) above and that the initial population distribution is sufficiently spread out over an interval compared to the radius of the support of $\varphi$. Then after a sufficient number of iterations of the game, players will myopically cluster into small clusters sufficiently separated from each other so that they will not see any incentive to move toward each other. For example, consider a population whose ideological positions on an issue are distributed over some interval (say, between "left" and "right"). If the initial distribution is sufficiently spread out, then over time the incentive to congregate will in fact induce "polarization," i.e. the formation of two distinct groups with no incentive to move toward one another. This occurs despite the fact that no individual has any preference for moving one direction over another; they are driven solely by the incentive to find themselves in a crowded region. Somewhat counter-intuitively, the desire to maximize crowd size results in multiple insular groups rather than all players congregating in the middle. We find this to be a thought-provoking result that may have implications for social science.

In the remainder of this introduction, we introduce some notation and assumptions on the data which will hold throughout this study. Then in Sect. 2 we provide a simple proof that the problem is well-posed and give an algorithm to explicitly compute solutions when the initial condition is a given discrete (empirical) measure. In Sect. 3 we study the dynamical system (1.5) over the set of discrete measures; this section contains our key theoretical results. Section 4 discusses numerical simulations that illustrate our main results. Finally, we give some concluding remarks in Sect. 5.

### 1.1 Notation and Assumptions

We denote by $\mathcal{P}_1(\mathbb{R}^d)$ the set of all Borel probability measures $m$ on $\mathbb{R}^d$ such that the first moment $\int |x| \,\mathrm{d}m(x)$ is finite. It will be endowed with the Wasserstein metric

$$W_1(\mu, \nu) = \inf \left\{ \int |x - y| \,\mathrm{d}\pi(x, y) : \pi \in \Pi(\mu, \nu) \right\} \tag{1.6}$$

where $\Pi(\mu, \nu)$ denotes the set of all couplings between $\mu$ and $\nu$. By Kantorovitch duality, we also have the characterization

$$W_1(\mu, \nu) = \sup\left\{\int \phi \mathrm{d}(\mu - \nu) : \|\nabla\phi\|_\infty \leq 1\right\}. \tag{1.7}$$

When $A$ and $B$ are symmetric matrices, we write $A \geq B$ to mean that $A - B$ is positive semi-definite, i.e. all of its eigenvalues are non-negative.

The basic assumptions on the data are given here:

**Assumption 1.1** $G : \mathbb{R}^d \times \mathcal{P}_1(\mathbb{R}^d) \to \mathbb{R}$ is continuous in both variables, twice differentiable with respect to the variable $x \in \mathbb{R}^d$. $D_x G(x, m)$ is $L_1$-Lipschitz with respect to the measure variable in the $W_1$ metric, i.e.

$$|D_x G(x, m) - D_x G(x, \tilde{m})| \leq L_1 W_1(m, \tilde{m}) \quad \forall x \in \mathbb{R}^d,\ \forall m, \tilde{m} \in \mathcal{P}_1(\mathbb{R}^d). \tag{1.8}$$

Both $D_x G$ and $D^2_{xx} G$ are bounded. In particular, there exists $\lambda_1(G) \geq 0$ such that $D^2_{xx} G(x, m) + \lambda_1(G) I \geq 0$ for all $x, m$.

## 2 Computation of Solutions

The purpose of this section is to propose an algorithm to compute $E^t(m)$, based on discretization of the measure. To begin with, however, we prove that our problem is indeed well-posed. We then provide a simple algorithm for computing solutions when $m$ is an empirical measure, i.e. a sum of Dirac masses.

### 2.1 Well-posedness

**Theorem 2.1** *Assume $0 < t^* < \frac{1}{\lambda_1(G)+L_1}$, where $L_1$ and $\lambda_1(G)$ are defined in Assumption 1.1. If $0 < t \leq t^*$, then*

*(1) $J_t(x, \cdot, m)$ has a unique minimizer for every $x$ and $m$, hence $y_t^m(x)$ and $F^t_{m_0}(m)$ are well-defined;*
*(2) $F^t_{m_0}$ is a contraction on $\mathcal{P}_1(\mathbb{R}^d)$ with Lipschitz constant $\alpha = \frac{tL_1}{1-t\lambda_1(G)} < 1$ and therefore has a unique fixed point, so that $E^t$ is well-defined;*
*(3) $E^t$ is continuous with the respect to the Wasserstein metric, i.e. if $\{m_{0,n}\}$ is a sequence in $\mathcal{P}_1(\mathbb{R}^d)$ such that $W_1(m_{0,n}, m_0) \to 0$, then $W_1\left(E^t(m_{0,n}), E^t(m_0)\right) \to 0$.*

***Proof*** *Step 1*

To find the minimizer for $J_t(x, \cdot, m)$ we start by taking the derivative with respect to $y$ and setting it equal to 0

$$D_y[J_t(x, y, m)] = \frac{y - x}{t} + D_y G(y, m) = 0 \tag{2.1}$$

So we see there is a critical point:

$$y + t D_y G(y, m) = x \tag{2.2}$$

Then taking the 2nd derivative with respect to $y$ and using 1.1, we see

$$D^2_{yy}[J_t(x, y, m)] = \frac{1}{t} I + D^2_{yy} G(y, m) \geq \left(\frac{1}{t} - \lambda_1(G)\right) I \tag{2.3}$$

where $I$ is the identity matrix. Thus $D^2_{yy}[J_t(x, y, m)]$ is positive definite provided that $0 < t \leq t^* < \frac{1}{\lambda_1(G)}$, and since $\frac{1}{\lambda_1(G)+L_1} \leq \frac{1}{\lambda_1(G)}$ this is satisfied. Then by the second derivative test, the critical point in 2.2 is the unique minimizer.

*Step 2*

Let $\varphi$ be Lipschitz with $\|\varphi\|_{\text{Lip}} \leq 1$

$$\int \varphi \mathrm{d}(y_t^m\#m_0 - y_t^{\tilde{m}}\#m_0) = \int [\varphi(y_t^m(x)) - \varphi(y_t^{\tilde{m}}(x))]\mathrm{d}m_0(x) \leq \int |y_t^m(x) - y_t^{\tilde{m}}(x)|\mathrm{d}m_0(x). \tag{2.4}$$

Using (2.2), we get

$$\begin{aligned} y_t^m - y_t^{\tilde{m}} &= x - tD_yG(y_t^m, m) - x + tD_yG(y_t^{\tilde{m}}, \tilde{m}) \\ &= -t[D_yG(y_t^m, m) - D_yG(y_t^m, \tilde{m}) + D_yG(y_t^m, \tilde{m}) - D_yG(y_t^{\tilde{m}}, \tilde{m})] \end{aligned} \tag{2.5}$$

Taking the dot product of each side with $y_t^m - y_t^{\tilde{m}}$ yields

$$\begin{aligned} |y_t^m - y_t^{\tilde{m}}|^2 = -t[&(y_t^m - y_t^{\tilde{m}}) \cdot (D_yG(y_t^m, m) - D_yG(y_t^m, \tilde{m})) \\ &+ (y_t^m - y_t^{\tilde{m}}) \cdot (D_yG(y_t^m, \tilde{m}) - D_yG(y_t^{\tilde{m}}, \tilde{m}))] \end{aligned} \tag{2.6}$$

Since $(y_t^m - y_t^{\tilde{m}}) \cdot (D_yG(y_t^m, \tilde{m}) - D_yG(y_t^{\tilde{m}}, \tilde{m})) \geq -\lambda_1(G)\left|y_t^m - y_t^{\tilde{m}}\right|^2$,

$$(1 - t\lambda_1(G))|y_t^m - y_t^{\tilde{m}}|^2 \leq |t(y_t^m - y_t^{\tilde{m}}) \cdot (D_yG(y_t^m, m) - D_yG(y_t^m, \tilde{m}))| \tag{2.7}$$

Now we have

$$|y_t^m - y_t^{\tilde{m}}| \leq \frac{t}{1 - t\lambda_1(G)}|D_yG(y_t^m, m) - D_yG(y_t^m, \tilde{m})| \leq \frac{tL_1}{1 - t\lambda_1(G)}W_1(m, \tilde{m}) \tag{2.8}$$

where $L_1$ is the Lipschitz constant of $D_yG$ with respect to $m$ in the $W_1$ metric. Therefore,

$$\int |y_t^m(x) - y_t^{\tilde{m}}(x)|\mathrm{d}m_0(x) \leq \frac{tL_1}{1 - t\lambda_1(G)}W_1(m, \tilde{m}) \tag{2.9}$$

So,

$$W_1(y_t^m\#m_0, y_t^{\tilde{m}}\#m_0) = \sup_{||\varphi||_{\text{Lip}\leq 1}} \int \varphi \mathrm{d}(y_t^m\#m_0 - y_t^{\tilde{m}}\#m_0) \leq \frac{tL_1}{1 - t\lambda_1(G)}W_1(m, \tilde{m}) \tag{2.10}$$

We deduce that for $t \leq t^* < \frac{1}{\lambda_1(G)+L_1}$, $F^t_{m_0}$ is a contraction on $\mathcal{P}_1(\mathbb{R}^d)$

*Step 3*

Set $\mu_0 := E^t(m_0)$ and $\mu_n := E^t(m_{0,n})$. By using the inverse function theorem on Equation (2.2), we deduce that for $0 < t \leq t^* < \frac{1}{\lambda_1(G)+L_1}$ we have $\nabla y_t^{\mu_n}(x) = (I + tD^2_{yy}G(y_t^{\mu_n}(x), \mu_n))^{-1}$. Since $D^2_{yy}G$ is bounded we get $\left\|\nabla y_t^{\mu_n}\right\| < C_0$ such that $C_0$ does not depend on $n$, but only on $t$ and the bound on $D^2_{yy}G$.

Let $\phi$ such that $\|\nabla\phi\|_\infty \leq 1$. Then

$$\begin{aligned} \int \phi d(\mu_0 - \mu_n) &= \int \left(\phi \circ y_t^{\mu_0} - \phi \circ y^{\mu_n}\right)\mathrm{d}m_0 + \int \phi \circ y_t^{\mu_n}\mathrm{d}(m_0 - m_n) \\ &\leq \int \left\|y_t^{\mu_0} - y_t^{\mu_n}\right\|\mathrm{d}m_0 + \int \phi \circ y_t^{\mu_n}\mathrm{d}(m_0 - m_n). \end{aligned} \tag{2.11}$$

Since $\left\|\nabla\phi \circ y_t^{\mu_n}\right\|_\infty \leq C_0$, by taking the supremum with respect to $\phi$ and applying (2.8) we have

$$W_1(\mu_0, \mu_n) \leq \frac{tL_1}{1 - t\lambda_1(G)}W_1(\mu_0, \mu_n) + C_0W_1(m_0, m_n). \tag{2.12}$$

Since $t \le t^*$ and $\frac{t^* L_1}{1-t^*\lambda_1(G)} < 1$, we let $n \to \infty$ to conclude that $\mu_n$ converges to $\mu_0$ respect to the $W_1$ metric. □

## 2.2 Discretization

We will now discretize the problem by considering only empirical distributions as initial measures. We then give an algorithm, based on standard Picard iteration, to compute the equilibrium to any specified degree of accuracy. As empirical measures are dense in $\mathcal{P}_1(\mathbb{R}^d)$, by part 3 of Theorem 2.1, our algorithm can in fact approximate the equilibrium for arbitrary initial measures to any specified degree of accuracy.

Let us now suppose that the initial measure is an empirical distribution, i.e. $m_0 = \sum_{j=1}^N a_j \delta_{x_j}$ for some points $x_1, \dots, x_N \in \mathbb{R}^d$ and some non-negative numbers $a_j$ that sum to 1. Then

$$F_{m_0}^t(m) = y_t^m \sharp m_0 = \sum_{j=1}^N a_j \delta_{y_t^m(x_j)}. \tag{2.13}$$

It follows that any equilibrium $m$ must itself be an empirical measure of the form $m = \sum_{j=1}^N a_j \delta_{z_j}$ for some points $z_1, \dots, z_N \in \mathbb{R}^d$. For any vector $\mathbf{z} = (z_1, \dots, z_N) \in \mathbb{R}^{dN}$, define

$$\begin{aligned} \tilde{y}_t^{\mathbf{z}}(x) &= y_t^m(x) \quad \text{where} \quad m = \sum_{j=1}^N a_j \delta_{z_j}, \\ \tilde{F}_{\mathbf{x}}^t(\mathbf{z}) &= \tilde{y}_t^{\mathbf{z}}(\mathbf{x}) := (\tilde{y}_t^{\mathbf{z}}(x_1), \dots, \tilde{y}_t^{\mathbf{z}}(x_N)). \end{aligned} \tag{2.14}$$

To find an equilibrium, it is enough to find a fixed point of $\tilde{F}_{\mathbf{x}}^t$. To see this, note that if $\mathbf{z} = \tilde{F}_{\mathbf{x}}^t(\mathbf{z})$, then we have

$$F_{m_0}^t\left(\sum_{j=1}^N a_j \delta_{z_j}\right) = \sum_{j=1}^N a_j \delta_{\tilde{y}_t^{\mathbf{z}}(x_j)} = \sum_{j=1}^N a_j \delta_{z_j}. \tag{2.15}$$

It is not hard to see that under the same hypotheses as in Theorem 2.1, $\tilde{F}_{\mathbf{x}}^t$ is a contraction with Lipschitz constant $\alpha = \frac{tL_1}{1-t\lambda_1(G)}$ with respect to the 1-norm $\|\mathbf{x}\|_1 = \sum_{j=1}^n |x_j|$, and it therefore has a unique fixed point, which we will denote

$$\tilde{E}_N^t(\mathbf{x}) := \mathbf{z} = \tilde{F}_{\mathbf{x}}^t(\mathbf{z}). \tag{2.16}$$

By classical theory, one can define

$$\begin{aligned} \mathbf{z}_0 &= \mathbf{x}, \\ \mathbf{z}_{k+1} &= \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) \end{aligned} \tag{2.17}$$

and get $\mathbf{z}_k \to \mathbf{z} = \tilde{F}_{\mathbf{x}}^t(\mathbf{z})$. The error estimate is

$$\|\mathbf{z}_k - \mathbf{z}\| \le \frac{\alpha^n}{1-\alpha} \left\| \tilde{F}_{\mathbf{x}}^t(\mathbf{x}) - \mathbf{x} \right\|, \tag{2.18}$$

where $\alpha \in (0, 1)$ is the Lipschitz constant for $\tilde{F}_{\mathbf{x}}^t$.

***Remark 2.2*** Although the contraction mapping theorem allows us to take any initial condition we like, we choose $\mathbf{z}_0 = \mathbf{x}$ for the following common sense reason. Since the time horizon is small, players cannot move much from their initial distribution, so the final distribution is sure to be close to $\mathbf{x}$; hence a good initial guess is $\mathbf{x}$ itself.

In principle, this simple algorithm is enough to solve our problem. However, we cannot compute $\tilde{F}_{\mathbf{x}}^t(\mathbf{z})$ explicitly. We will now point out that it is enough to approximate $\tilde{F}_{\mathbf{x}}^t(\mathbf{z})$ sufficiently well.

**Lemma 2.3** *Suppose $\{\mathbf{z}_k\}$ is a sequence such that*

$$\left\| \mathbf{z}_{k+1} - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) \right\| \leq \varepsilon_k, \tag{2.19}$$

*where $\varepsilon_k \leq \alpha^k$ and $\alpha < 1$ is the Lipschitz constant for $\tilde{F}_{\mathbf{x}}^t$. Then $\mathbf{z}_k \to \mathbf{z}$, where $\mathbf{z}$ is the unique fixed point of $\tilde{F}_{\mathbf{x}}^t$. We have an error estimate:*

$$\|\mathbf{z}_k - \mathbf{z}\| \leq \alpha^k \left( \frac{2\,(k - \alpha k + 1)}{(1-\alpha)^2} + \frac{\|\mathbf{z}_1 - \mathbf{z}_0\|}{1-\alpha} \right). \tag{2.20}$$

***Remark 2.4*** We can take any norm in Lemma 2.3, not just the Euclidean norm. In particular, it may be convenient to take the 1-norm $\|\mathbf{x}\| = \sum_{j=1}^n |x_j|$.

***Proof*** We have $\left\| \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_{k+1}) - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) \right\| \leq \alpha \|\mathbf{z}_{k+1} - \mathbf{z}_k\|$ for every $k$. Applying (2.19) and the triangle inequality,

$$\begin{aligned}\|\mathbf{z}_{k+1} - \mathbf{z}_k\| &\leq \left\| \mathbf{z}_{k+1} - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) \right\| + \left\| \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_{k-1}) \right\| + \left\| \mathbf{z}_k - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_{k-1}) \right\| \\ &\leq \varepsilon_k + \alpha \|\mathbf{z}_k - \mathbf{z}_{k-1}\| + \varepsilon_{k-1}.\end{aligned}$$

By iteration we get

$$\|\mathbf{z}_{k+1} - \mathbf{z}_k\| \leq \alpha^k \|\mathbf{z}_1 - \mathbf{z}_0\| + \sum_{j=0}^{k} \alpha^{k-j} \left( \varepsilon_{j+1} + \varepsilon_j \right). \tag{2.21}$$

Now taking $m, n \in \mathbb{N}$ with $m > n$, we use (2.21) to estimate

$$\begin{aligned}\|\mathbf{z}_m - \mathbf{z}_n\| \leq \sum_{k=n}^{m-1} \|\mathbf{z}_{k+1} - \mathbf{z}_k\| &\leq \sum_{k=n}^{m-1} \alpha^k \|\mathbf{z}_1 - \mathbf{z}_0\| + \sum_{k=n}^{m-1} \sum_{j=0}^{k} \alpha^{k-j} \left( \varepsilon_{j+1} + \varepsilon_j \right) \\ &\leq \frac{\alpha^n}{1-\alpha} \|\mathbf{z}_1 - \mathbf{z}_0\| + \sum_{k=n}^{m-1} \sum_{j=0}^{k} \alpha^{k-j} \left( \varepsilon_{j+1} + \varepsilon_j \right).\end{aligned}$$

Now we focus on the double sum. Using the assumption $\varepsilon_k \leq \alpha^k$, we get

$$\sum_{k=n}^{m-1} \sum_{j=0}^{k} \alpha^{k-j} \left( \varepsilon_{j+1} + \varepsilon_j \right) \leq \sum_{k=n}^{m-1} (k+1) \left( \alpha^{k+1} + \alpha^k \right) \leq 2 \sum_{k=n}^{\infty} (k+1)\, \alpha^k.$$

For a closed form we compute

$$2 \sum_{k=n}^{\infty} (k+1)\, \alpha^k = 2 \frac{\mathrm{d}}{\mathrm{d}\alpha} \left[ \sum_{k=n}^{\infty} \alpha^{k+1} \right] = 2 \frac{\mathrm{d}}{\mathrm{d}\alpha} \left[ \frac{\alpha^{n+1}}{1-\alpha} \right] = \frac{2\alpha^n\,(n - \alpha n + 1)}{(1-\alpha)^2}.$$

We then deduce

$$\|\mathbf{z}_m - \mathbf{z}_n\| \leq \alpha^n \left( \frac{2\,(n - \alpha n + 1)}{(1-\alpha)^2} + \frac{\|\mathbf{z}_1 - \mathbf{z}_0\|}{1-\alpha} \right).$$

Therefore $\{\mathbf{z}_k\}$ is Cauchy and thus converges. The error estimate (2.20) is derived by taking $m \to \infty$ and replacing $n$ with $k$. □

Thanks to Lemma 2.3, we can now specify a fully formed algorithm to compute the equilibrium:

- Set $\mathbf{z}_0 = \mathbf{x}$.
- Given $\mathbf{z}_k$, use Newton's method to approximate $\tilde{y}_t^{\mathbf{z}_k}(x_j)$ for $j = 1, \ldots, N$, until we obtain a vector $\mathbf{z}_{k+1}$ satisfying

$$\left\| \mathbf{z}_{k+1} - \tilde{F}_{\mathbf{x}}^t(\mathbf{z}_k) \right\| \leq \alpha^k. \tag{2.22}$$

Then $\mathbf{z}_k$ will converge to the equilibrium, with error estimate given by Lemma 2.3. We remark that, since empirical measures are dense in the Wasserstein space and the equilibrium map $E^t(m)$ is continuous, this algorithm can also be used to approximate $E^t(m)$ for any $m \in \mathscr{P}_1(\mathbb{R}^d)$ within an arbitrary specified margin of error.

## 3 Asymptotic Behavior of the Equilibrium Map

Recall that $E^t(m_0)$ is defined to be the equilibrium measure, i.e. the final distribution, given an initial distribution $m_0$. In this section we study the dynamical system (1.5). We will focus in this paper on determining stability for the dynamical system (1.5) for initial measures of the form $m_0 = \sum_{j=1}^n \frac{1}{n}\delta_{x_j}$. Such empirical measures are also dense in the Wasserstein space $\mathscr{P}_1(\mathbb{R}^d)$. As discussed in Sect. 2.2, for such initial measures we can replace $E^t(m)$ with $\tilde{E}_n^t(\mathbf{x})$, so it is equivalent to study the classical dynamical system

$$\begin{aligned} &\mathbf{x}_0 \in (\mathbb{R}^d)^n \quad \text{is given,} \\ &\mathbf{x}_{k+1} = \tilde{E}_n^t(\mathbf{x}_k), \quad k = 0, 1, 2, \ldots \end{aligned} \tag{3.1}$$

We determine the asymptotic behavior by identifying

- the fixed points of $\tilde{E}_n^t$, and
- the spectrum of the derivative of $\tilde{E}_n^t$ at fixed points.

Our results in this section will hold for a particular case, namely when $d = 1$ and the cost function $G$ takes the form

$$G(y, \mu) = \int \varphi(y - z)\mathrm{d}\mu(z),$$

where $\varphi$ is a smooth function. However, we will start with some more general properties of $\tilde{E}_n^t$ before restricting to this special case.

### 3.1 General Properties of $D\tilde{E}_n^t$

**Proposition 3.1** *Recall the definition $y_t^\mu(x) = \operatorname{argmin} J_t(x, \cdot, \mu)$. Let $0 < t \leq t^*$, where $t^*$ is as in Theorem 2.1. Then $Dy_t^\mu(x)$ is a positive definite symmetric matrix for each $x$.*

***Proof*** Let $y(x) = y_t^\mu(x)$. Then

$$\frac{y(x) - x}{t} + D_y G(y(x), \mu) = 0. \tag{3.2}$$

Taking the derivative with respect to $x$ yields

$$\frac{Dy(x)-I}{t}+D^2_{yy}G(y(x),\mu)Dy(x)=0 \tag{3.3}$$

$$(I+tD^2_{yy}G(y(x),\mu))Dy(x)=I \tag{3.4}$$

Note: $D^2_{yy}G(y(x),\mu)$ is a symmetric matrix and $I+tD^2_{yy}G(y(x),\mu)$ is as well. Moreover, by Assumption 1.1 and the condition $t \le t^* < \frac{1}{\lambda_1(G)}$, we see that $I+tD^2_{yy}G(y(x),\mu)$ is positive definite, hence its inverse $Dy(x)$ is positive definite as well. □

Now we may use this result to determined the fixed points of the dynamical system.

**Proposition 3.2** *Let*

$$\mu=\frac{1}{n}\sum_{i=1}^{n}\delta_{x_i} \tag{3.5}$$

*where $x_i \in \mathbb{R}$ for $1 \le i \le n$. Suppose $y\#\mu=\mu$, and $y$ is strictly monotone increasing, then $y(x)=x$ for all $x \in \mathrm{supp}(\mu)$.*

***Proof*** Since $\mu$ is an empirical measure, $\mathrm{supp}(\mu)=\{x_i\}_{i=1}^n$. Also, since $y\#\mu=\mu$,

$$\mathrm{supp}(y\#\mu)=\mathrm{supp}(\mu)=\{x_i\}_{i=1}^n \tag{3.6}$$

Therefore for each $1 \le i \le n$, $y(x_i)=x_j$ for some $1 \le j \le n$. Without loss of generality we may assume $x_1 < x_2 < \cdots < x_n$. Suppose $y(x_1)=x_i$ for some $i>1$. Then since $y$ is increasing, $y(x_j)>x_i$ for all $j>1$. Then by the pigeonhole principle, this implies $y$ is not injective. This contradicts that $y$ is strictly monotone increasing, so $y(x_1)=x_1$. This argument can be repeated to show $y(x_i)=x_i$ for each $1 \le i \le n$. Therefore $y(x)=x$ for all $x \in \mathrm{supp}(\mu)$. □

Now we know fixed points of $\tilde{E}^t_n$ correspond to measures $\mu=\frac{1}{n}\sum_{j=1}^n x_j$ such that $y^\mu_t=I$ on the support of $\mu$. Our next goal is to linearize $\tilde{E}^t_n$ around fixed points to check if the system is stable. We do this by computing $D\tilde{E}^t_n(\mathbf{x})$ and showing that the eigenvalues of $D\tilde{E}^t_n(\mathbf{x})$ are small if $\mathbf{x}$ is a fixed point.

Let us introduce the notation

$$\mu_{\mathbf{x}}=\frac{1}{n}\sum_{j=1}^{n}x_j \quad \forall \mathbf{x}=(x_1,\ldots,x_n).$$

Then we define $g(y,\mathbf{x})=G(y,\mu_{\mathbf{x}})$. Now $\mu_{\mathbf{y}}$ is the equilibrium for an initial measure $\mu_{\mathbf{x}}$ if and only if for each $1 \le j \le n$, $y_j$ is the optimal move for $x_j$ given $\mu_{\mathbf{y}}$ as a final measure, i.e.

$$y_j+tD_yg(y_j(\mathbf{x}),\mathbf{y})=x_j. \tag{3.7}$$

Hence we can define

$$\mathbf{y}(\mathbf{x})=(y_1(\mathbf{x}),\ldots,y_n(\mathbf{x}))=\tilde{E}^t_n(\mathbf{x}) \tag{3.8}$$

implicitly through (3.7).

We want to compute $D\tilde{E}^t_n(\mathbf{x})=D\mathbf{y}(\mathbf{x})$. To do this we take the implicit partial derivative of (3.7) with respect to $x_k$ for $1 \le k \le n$

$$\frac{\partial}{\partial x_k}[y_j(\mathbf{x})+tD_yg(y_j(\mathbf{x}),\mathbf{y}(\mathbf{x}))=x_j] \tag{3.9}$$

to get

$$\frac{\partial y_j}{\partial x_k}+tD^2_{yy}g(y_j(\mathbf{x}),\mathbf{y}(\mathbf{x}))\frac{\partial y_j}{\partial x_k}+t\sum_{i=1}^{n}D^2_{yx_i}g(y_j(\mathbf{x}),\mathbf{y}(\mathbf{x}))\cdot\frac{\partial y_i}{\partial x_k}=\frac{\partial x_j}{\partial x_k}=\delta_{j,k}. \quad (3.10)$$

We get a system of equations which is presented in matrix form as $AD\tilde{E}^t_n(\mathbf{x})=I$ where $A=(A_{i,j})^n_{i,j=1}$,

$$A_{i,j}=\delta_{i,j}(1+tD^2_{yy}g(y_j(\mathbf{x}),\mathbf{y}(\mathbf{x})))+tD^2_{yx_i}g(y_j(\mathbf{x}),\mathbf{y}(\mathbf{x})). \quad (3.11)$$

Next we will determine the spectrum of $A$ and thereby obtain the spectrum of $D\tilde{E}^t_n(\mathbf{x})$.

### 3.2 Spectral Properties

For simplicity, we will look specifically at the case

$$G(y,\mu)=\int\varphi(y-z)\mathrm{d}\mu(z).$$

In this case,

$$g(y,\mathbf{y})=G\left(y,\frac{1}{n}\sum_{k=1}^{n}\delta_{y_k}\right)=\frac{1}{n}\sum_{k=1}^{n}\varphi(y-y_k), \quad (3.12)$$

where $\varphi$ is $C^2$. Then we compute

$$D^2_{yy}g(y_j,\mathbf{y})=\frac{1}{n}\sum_{k=1}^{n}\varphi''(y_j-y_k) \quad (3.13)$$

and

$$D^2_{yx_i}g(y_j,\mathbf{y})=-\frac{1}{n}\varphi''(y_j-y_k). \quad (3.14)$$

We see $A$ can be written in the following form

$$\begin{aligned}A&=I+\frac{t}{n}\begin{pmatrix}\sum_{k=1,k\neq 1}^{n}\varphi''(y_1-y_k) & -\varphi''(y_2-y_1) & \cdots & -\varphi''(y_n-y_1)\\ -\varphi''(y_1-y_2) & \sum_{k=1,k\neq 2}^{n}\varphi''(y_2-y_k) & \cdots & -\varphi''(y_n-y_1)\\ \vdots & \vdots & \ddots & \vdots\\ -\varphi''(y_1-y_n) & -\varphi''(y_2-y_n) & \cdots & \sum_{k=1,k\neq n}^{n}\varphi''(y_n-y_k)\end{pmatrix}\\ &=: I+\frac{t}{n}B\end{aligned} \quad (3.15)$$

**Assumption 3.3** $\varphi\in C^2$ is even, has a minimum at $x=0$, is increasing on the interval $(0,r)$, and $\varphi(x)=0$ for $x\geq r$.

Note: Since we work with points $\mathbf{x}\in\mathbb{R}^{\mathbf{n}}$ which describe an empirical measure, without loss of generality, we may arrange the components of $\mathbf{x}$ so that $x_1\leq x_2\leq\cdots\leq x_n$.

**Theorem 3.4** *Suppose $\varphi$ meets Assumption 3.3. Then $\mathbf{x}=(x_1,x_2,\ldots,x_n)$ (where each $x_i\in\mathbb{R}$) is a fixed point of $\tilde{E}^t_n$ if and only if for any $1\leq j,k\leq n$, $x_j=x_k$ or $|x_j-x_k|\geq r$.*

***Proof*** Now if $\mathbf{x}$ is a fixed point of $\tilde{E}_n^t$, then for each $1 \le j \le n$

$$y_j + \frac{t}{n}\sum_{k=1}^{n}\varphi'(y_j - y_k) = x_j, \tag{3.16}$$

and by Proposition 3.2 $y_j = x_j$, so for $1 \le j \le n$

$$\sum_{k=1}^{n}\varphi'(x_j - x_k) = 0. \tag{3.17}$$

For $j = 1$, $\sum_{k=1}^{n}\varphi'(x_k - x_1) = 0$. Now if $x_1 = x_k$ for any $k$, $\varphi'(x_k - x_1) = 0$. Suppose $k_1 + 1$ is the first $k$ such that $x_k - x_1 > 0$, then

$$\sum_{k=k_1+1}^{n}\varphi'(x_j - x_k) = 0 \tag{3.18}$$

and since $\varphi'(x_k - x_1) \ge 0$ for $k > k_1$, we need $|x_k - x_1| \ge r$ for $k > k_1$. So $x_1 = x_k$ for $k \le k_1$ and $|x_k - x_1| \ge r$ for $k > k_1$.

Now suppose $x_j$ is such that $|x_j - x_k| \ge r$ for $k < j$, then

$$\sum_{k=1}^{j}\varphi'(x_k - x_j) = 0. \tag{3.19}$$

So we still need

$$\sum_{k=j+1}^{n}\varphi'(x_k - x_j) = 0. \tag{3.20}$$

If $x_j = x_k$ for any $k > j$, $\varphi'(x_k - x_j) = 0$, and if $k_j + 1$ is the first $k$ such that $x_k - x_j > 0$,

$$\sum_{k=k_j+1}^{n}\varphi'(x_k - x_j) = 0. \tag{3.21}$$

Therefore $\varphi'(x_k - x_j) = 0$ for $k > k_j$, so $|x_k - x_j| \ge r$ for $k > k_j$. Thus if $\mathbf{x}$ is a fixed point of $\tilde{E}_n^t$, for all $1 \le j, k \le n$ either $x_j = x_k$ or $\left|x_j - x_k\right| \ge r$.

Now suppose for any $1 \le j, k \le n$ either $x_k = x_j$ or $|x_k - x_j| \ge r$. Then if $x_j = x_k$, $\varphi'(x_j - x_k) = \varphi'(0) = 0$, and if $|x_k - x_j| \ge r$, $\varphi'(x_k - x_j) = 0$. So

$$\sum_{k=1}^{n}\varphi'(x_k - x_j) = 0 \tag{3.22}$$

for all $1 \le j \le n$. Thus $\mathbf{x}$ is a fixed point for $\tilde{E}_n^t$. □

Let $\mathbf{x}$ be a fixed point of $\tilde{E}_n^t$. By Theorem 3.4, there exist indices $k_m$ such that:

$$\begin{aligned}
x_{k_1} &= x_k, & 1 = k_0 &\le k \le k_1\\
x_{k_2} &= x_k, & k_1 &< k \le k_2\\
&\vdots & &\vdots\\
x_{k_i} &= x_k, & k_{i-1} &< k \le k_i = n
\end{aligned}$$

and $x_{k_m} - x_{k_{m-1}} \ge r$ for $2 \le m \le i$.

**Definition 3.5** We define $\mathbf{x}$ to be *sufficiently spread out* if we have the strict inequality $x_{k_m} - x_{k_{m-1}} > r$ for $2 \leq m \leq i$.

Recall, we are evaluating $D\tilde{E}_n^t(\mathbf{x})$ for $\mathbf{x}$ a fixed point of $\tilde{E}_n^t$. So it suffices to look at the spectrum of $B$ evaluated at fixed points. According to Theorem 3.4, if $x_k \neq x_j$, then $|x_j - x_k| > r$ and $\varphi''(x_j - x_k) = 0$. But if $x_k = x_j$, then $\varphi''(x_j - x_k) = \varphi''(0) = 1$.

$$B = \begin{pmatrix} \sum_{k=1,k\neq 1}^n \varphi''(x_1 - x_k) & -\varphi''(x_2 - x_1) & \cdots & -\varphi''(x_n - x_1) \\ -\varphi''(x_1 - x_2) & \sum_{k=1,k\neq 2}^n \varphi''(x_2 - x_k) & \cdots & -\varphi''(x_n - x_1) \\ \vdots & \vdots & \ddots & \vdots \\ -\varphi''(x_1 - x_n) & -\varphi''(x_2 - x_n) & \cdots & \sum_{k=1,k\neq n}^n \varphi''(x_n - x_k) \end{pmatrix} \tag{3.23}$$

We see $B$ has a block diagonal form:

$$B = \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_i \end{pmatrix} \tag{3.24}$$

where for $1 \leq m \leq i$, $B_m$ is the $(k_m - k_{m-1}) \times (k_m - k_{m-1})$ matrix given by

$$B_m = \begin{pmatrix} (k_m - k_{m-1} - 1) & -1 & \cdots & -1 \\ -1 & (k_m - k_{m-1} - 1) & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & (k_m - k_{m-1} - 1) \end{pmatrix}. \tag{3.25}$$

**Lemma 3.6** *Let $C_n$ be an $n \times n$ matrix with the following form:*

$$C_n = \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix}. \tag{3.26}$$

*Then if $n \geq 2$, $C_n$ is a positive matrix.*

***Proof*** We proceed by induction on $n$.

Base case: $n = 2$

$$C_2 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \tag{3.27}$$

Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$ be an arbitrary real vector. Then

$$\mathbf{v}^T C_2 \mathbf{v} = (v_1 - v_2)^2 \geq 0. \tag{3.28}$$

Now supposing the hypothesis holds for $C_{n-1}$ for $n > 2$, we want to prove it holds for $C_n$. Let $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$. Now,

$$\begin{aligned} \mathbf{v}^T C_n \mathbf{v} &= \begin{pmatrix} v_1 & \cdots & v_{n-1} \end{pmatrix} C_{n-1} \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} - 2v_n \sum_{k=1}^{n-1} v_k + \sum_{k=1}^{n-1} v_k^2 + \sum_{k=1}^{n-1} v_n^2 \\ &= \begin{pmatrix} v_1 & \cdots & v_{n-1} \end{pmatrix} C_{n-1} \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} + \sum_{k=1}^{n-1} (v_k - v_n)^2 \end{aligned} \tag{3.29}$$

By our inductive hypothesis,

$$\begin{pmatrix} v_1 & \cdots & v_{n-1} \end{pmatrix} C_{n-1} \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} \geq 0.$$

Additionally, $(v_k - v_n)^2 \geq 0$ for all $1 \leq k \leq n-1$, so we can conclude $C_n$ is a positive matrix. □

**Corollary 3.7** *If $n \geq 2$, the eigenspace for $C_n$ for the eigenvalue $\lambda = 0$ is one dimensional and spanned by* $\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$.

***Proof***

$$C_n \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} n-1 & -1 & \cdots & -1 \\ -1 & n-1 & \cdots & -1 \\ \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & n-1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \mathbf{0} \tag{3.30}$$

So our vector is an eigenvector for the eigenvalue $\lambda = 0$. Now we'd like to show that it spans the eigenspace. Without loss of generality, suppose $v_n \neq v_k$ for some $1 \leq k \leq n-1$ where $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix}$. Then from (3.29) we have

$$\mathbf{v}^T C_n \mathbf{v} = \begin{pmatrix} v_1 & \cdots & v_{n-1} \end{pmatrix} C_{n-1} \begin{pmatrix} v_1 \\ \vdots \\ v_{n-1} \end{pmatrix} + \sum_{k=1}^{n-1} (v_k - v_n)^2 \tag{3.31}$$

As before, the first piece is nonnegative and since $v_n \neq v_k$ for some $k$, the sum is strictly greater than 0 so any $\mathbf{v}$ not in the span of $\begin{pmatrix}1\\1\\\vdots\\1\end{pmatrix}$ is not an eigenvector for $\lambda = 0$. □

So eigenvalues for $B$ are greater than or equal to 0. Recalling $D\tilde{E}_n^t(\mathbf{x}) = A^{-1}$ and $A = I + \frac{t}{n}B$, eigenvalues for $D\tilde{E}_n^t(\mathbf{x})$ are $\lambda \leq 1$.

**Definition 3.8** Consider the dynamical system

$$\begin{aligned} \mathbf{y}_{k+1} &= F(\mathbf{y}_k) \\ \mathbf{y}_0 &\in \mathbb{R}^n \end{aligned} \tag{3.32}$$

where $F : \mathbb{R}^n \to \mathbb{R}^n$ and $D \subset \mathbb{R}^n$. An fixed point $\mathbf{y}^*$ of $F$ is said to be **asymptotically stable** if there exists an $\varepsilon > 0$ such that whenever $|\mathbf{y}_0 - \mathbf{y}^*| < \varepsilon$, then the trajectory $\mathbf{y}_k$ converges to $\mathbf{y}^*$ as $k \to \infty$.

Throughout the rest of this section we will make use of the following theorem to show asymptotic stability of fixed points of $\tilde{E}_n^t$.

**Theorem 3.9** *[15, Theorem 1.3.7] If $F \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^n)$, then a fixed point $\mathbf{y}^*$ of the dynamical system (3.32) is asymptotically stable if the eigenvalues of $DF(\mathbf{y}^*)$ lie strictly inside the unit circle.*

From our spectral analysis, we see $D\tilde{E}_n^t(\mathbf{x}^*)$ has eigenvalues $\lambda \leq 1$. To be able to apply Theorem 3.9, we use the following lemmas to restrict to a subspace $W$ that is both invariant under $\tilde{E}_n^t$ and where $D\tilde{E}_n^t(\mathbf{x}^*)$ has eigenvalues strictly less than 1.

**Lemma 3.10** *If $g(y_j + \alpha, \mathbf{y} + \alpha(1, \ldots, 1)) = g(y_j, \mathbf{y})$ for any $\alpha \in \mathbb{R}$ and $y \in \mathbb{R}^n$, then $\tilde{E}_n^t(\mathbf{x} + \alpha(1, \ldots, 1)) = \tilde{E}_n^t(\mathbf{x}) + \alpha(1, \ldots, 1)$.*

***Proof*** Recall: $\tilde{E}_n^t(\mathbf{x}) = \mathbf{y}$ if and only if for all $1 \leq j \leq n$

$$y_j(\mathbf{x}) + tD_y g(y_j(\mathbf{x}), \mathbf{y}(\mathbf{x})) = x_j. \tag{3.33}$$

So if $\tilde{E}_n^t(\mathbf{x} + \alpha(1, \ldots, 1)) = \mathbf{z}$ then

$$z_j + tD_y g(z_j, \mathbf{z}) = x_j + \alpha \tag{3.34}$$

Equivalently,

$$z_j - \alpha + tD_y g(z_j, \mathbf{z}) = x_j \tag{3.35}$$

Now since $g(y + \alpha, \mathbf{y} + \alpha(1, \ldots, 1)) = g(y, \mathbf{y})$

$$z_j - \alpha + tD_y g(z_j - \alpha, \mathbf{z} - \alpha(1, \ldots, 1)) = x_j \tag{3.36}$$

This implies $\tilde{E}_n^t(\mathbf{x}) = \mathbf{z} - \alpha(1, \ldots, 1)$, so $\tilde{E}_n^t(\mathbf{x} + \alpha(1, \ldots, 1)) = \tilde{E}_n^t(\mathbf{x}) + \alpha(1, \ldots, 1)$ as desired. □

***Remark 3.11*** Lemma 3.10 essentially says that if the cost function $g$ is invariant under translations, then the translation operation commutes with the equilibrium map. More precisely, if we move all the players' initial positions by $\alpha$ and then find the equilibrium measure, we get the same result as when we first find the equilibrium measure and then translate it by $\alpha$. Note that, by Assumption 3.3, $g(y, \mathbf{y}) = \frac{1}{n}\sum_{k=1}^{n} \varphi(y - y_k)$ meets the criteria of Lemma 3.10.

**Lemma 3.12** *Let $g(y,\mathbf{y}) = \frac{1}{n}\sum_{k=1}^{n}\varphi(y-y_k)$ where $\varphi$ satisfies Assumption 3.3. If $\mathbf{x}\in W = \big[\operatorname{span}\{(1,\ldots,1)\}\big]^{\perp}$, then so is $\tilde{E}_n^t(\mathbf{x})$.*

***Proof*** Let $\mathbf{y} = \tilde{E}_n^t(\mathbf{x})$ and $\mathbf{x}\in W$. We are going to show that $\sum_{j=1}^{n} y_j = \sum_{j=1}^{n} x_j = 0$, from which the claim follows as a particular case. Recall that

$$y_j + \frac{t}{n}\sum_{k=1}^{n}\varphi'(y_j - y_k) = x_j \tag{3.37}$$

for each $j$, so

$$\sum_{j=1}^{n} y_j + \frac{t}{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi'(y_j - y_k) = \sum_{j=1}^{n} x_j. \tag{3.38}$$

It is enough to show that

$$\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi'(y_j - y_k) = 0. \tag{3.39}$$

Recall from Assumption 3.3 that $\varphi'$ is odd, and in particular $\varphi'(0) = 0$. Rewrite the double sum as follows:

$$\begin{aligned}
\sum_{j=1}^{n}\sum_{k=1}^{n}\varphi'(y_j - y_k) &= \sum_{j=1}^{n}\sum_{k=1}^{j-1}\varphi'(y_j - y_k) + \sum_{j=1}^{n}\sum_{k=j+1}^{n}\varphi'(y_j - y_k)\\
&= \sum_{j=1}^{n}\sum_{k=1}^{j-1}\varphi'(y_j - y_k) + \sum_{k=1}^{n}\sum_{j=1}^{k-1}\varphi'(y_j - y_k)\\
&= \sum_{j=1}^{n}\sum_{k=1}^{j-1}\varphi'(y_j - y_k) - \sum_{k=1}^{n}\sum_{j=1}^{k-1}\varphi'(y_k - y_j) = 0,
\end{aligned} \tag{3.40}$$

and the proof is complete. ◻

Let

$$W = [\operatorname{span}\{(1,\ldots,1)\}]^{\perp} \tag{3.41}$$

Now, let $\mathbf{x}_0$ be a point in our system. Then $\mathbf{x}_0$ has the following orthogonal decomposition

$$\mathbf{x}_0 = \mathbf{w}_0 + \alpha(1,\ldots,1) \tag{3.42}$$

where $\alpha\in\mathbb{R}$ and $\mathbf{w}\in W$. (One way to look at this decomposition is that the component $\mathbf{w}_0\in W$ corresponds to the initial distribution translated so that its barycenter lies at the origin, while $\alpha/n$ is equal to the original barycenter.) By Lemmas 3.10 and 3.12, the dynamical system (3.1) is equivalent to one restricted to $W$:

$$\begin{aligned}
\mathbf{w}_0 &\in W \quad \text{is given},\\
\mathbf{w}_{k+1} &= \tilde{E}_n^t(\mathbf{w}_k), \quad k = 0,1,2,\ldots
\end{aligned} \tag{3.43}$$

since for all $k$ we must have $\mathbf{x}_k = \mathbf{w}_k + \alpha(1,\ldots,1)$ and $\mathbf{w}_k\in W$. It is therefore sufficient to study the asymptotic stability of fixed points of $\tilde{E}_n^t$ that lie in $W$. We begin with the origin, which corresponds to a simple Dirac mass.

**Theorem 3.13** *Suppose*

$$\begin{aligned} \mathbf{x}_{k+1} &= \tilde{E}_n^t(\mathbf{x}_k) \\ \mathbf{x}_0 &\in W \end{aligned} \tag{3.44}$$

*where* $\tilde{E}_n^t : W \to W$ *is as defined in* (3.1) *and* $\varphi$ *meets assumption 3.3. Then the fixed point* $\mathbf{x}^* = (0, \dots, 0)$ *of the system is asymptotically stable.*

***Proof*** We just need to show (3.44) satisfies the assumptions of Theorem 3.9. $\tilde{E}_n^t \in \mathcal{C}^2(\mathbb{R}^n, \mathbb{R}^n)$ by a routine application of the Implicit Function Theorem, and therefore by restricting to $W$ we have $\tilde{E}_n^t \in \mathcal{C}^2(W, W)$, where $W$ is isomorphic to $\mathbb{R}^{n-1}$. From Corollary 3.7, we deduce that the eigenspace of $D\tilde{E}_n^t(\mathbf{x}^*)$ corresponding the eigenvalue $\lambda = 1$ is $W^\perp$. So $D\tilde{E}_n^t(\mathbf{x}^*)$ as a linear map from $W$ to $W$ has eigenvalues $\lambda < 1$, thus the fixed point $\mathbf{x}^*$ is asymptotically stable. □

**Corollary 3.14** *Suppose*

$$\begin{aligned} \mathbf{x}_{k+1} &= \tilde{E}_n^t(\mathbf{x}_k) \\ \mathbf{x}_0 &\in \mathbb{R}^n \end{aligned} \tag{3.45}$$

*where* $\tilde{E}_n^t : \mathbb{R}^n \to \mathbb{R}^n$ *is as defined in* (3.1) *and* $\varphi$ *meets assumption 3.3. Then a fixed point* $\mathbf{x}^* \in \operatorname{span}\{(1, \dots, 1)\}$ *of the system is asymptotically stable.*

***Proof*** For any $\mathbf{x} \in \mathbb{R}^n$, $\mathbf{x} = \mathbf{w} + \alpha(1, \dots, 1)$ for $\mathbf{w} \in W$ and $\alpha \in \mathbb{R}$. We see $\alpha(1, \dots, 1)$ simply produces a translation of $\tilde{E}_n^t$ by $\alpha$, which does not affect asymptotic stability. By Theorem 3.13, the fixed point $\mathbf{x}^* = (0, \dots, 0)$ is asymptotically stable for $\tilde{E}_n^t$ restricted to $W$, so any translation $\mathbf{x}^* \in \operatorname{span}\{(1, \dots, 1)\}$ is also stable on $\tilde{E}_n^t : \mathbb{R}^n \to \mathbb{R}^n$. □

Of course, there are more fixed points of $\tilde{E}_n^t$ than just the ones that correspond to a single dirac mass. To prove stability for these general fixed points, we use the following lemma to split up the game, inspect stability of the fixed points of each smaller game, and use these results to prove stability for the whole game.

**Lemma 3.15** *Suppose* $\mathbf{x}$ *is sufficiently close to a fixed point* $\mathbf{x}^*$*, where* $\mathbf{x}^*$ *is sufficiently spread out (Definition 3.5), so that*

$$x_1 \le \dots \le x_{k_1} \le x_{k_1+1} \le \dots \le x_{k_2} \le \dots \le x_{k_i}$$

*are such that* $x_{k_m+1} - x_{k_m} > r$ *for* $1 \le m \le i - 1$*. Then*

$$\tilde{E}_n^t(\mathbf{x}) = (\tilde{E}_{k_1}^{t_1}(x_1, \dots, x_{k_1}), \tilde{E}_{k_2-k_1}^{t_2}(x_{k_1+1}, \dots, x_{k_2}), \dots, \tilde{E}_{k_i-k_{i-1}}^{t_i}(x_{k_{i-1}+1}, \dots, x_{k_i})) \tag{3.46}$$

*where* $t_m = \frac{t(k_{m+1}-k_m)}{n}$.

***Proof*** Let us consider $\tilde{E}_{k_{m+1}-k_m}^{t_m}(x_{k_m+1}, \dots, x_{k_{m+1}})$ for $0 \le m \le i - 1$ and $k_i = n$.

$$\tilde{E}_{k_{m+1}-k_m}^{t_m}(x_{k_m+1}, \dots, x_{k_{m+1}}) = (y_{k_m+1}, \dots, y_{k_{m+1}}) \tag{3.47}$$

if and only if, for each $k_m + 1 \le j \le k_{m+1}$

$$y_j + \frac{t_m}{k_{m+1} - k_m} \sum_{k=k_m+1}^{k_{m+1}} \varphi'(y_j - y_k) = x_j. \tag{3.48}$$

Now we show $y_{k_m+1} - y_{k_m} > r$. Plug $j = k_m + 1$ into (3.48) to get

$$y_{k_m+1} + \frac{t_m}{k_{m+1} - k_m} \sum_{k=k_m+1}^{k_{m+1}} \varphi'(y_{k_m+1} - y_k) = x_{k_m+1}. \tag{3.49}$$

Now $\varphi'(y_{k_m+1} - y_k) \leq 0$ for each $k$ in this summation, since $\varphi'$ is odd and $y_{k_m+1} \leq \cdots \leq y_{k_{m+1}}$. Therefore the sum is non-positive, and we deduce $y_{k_m+1} \geq x_{k_m+1}$. By similar reasoning (with $m$ replaced by $m-1$ in (3.48) and plugging in $j = k_m$), we get $y_{k_m} \leq x_{k_m}$. Therefore,

$$y_{k_m+1} - y_{k_m} \geq x_{k_m+1} - x_m > r. \tag{3.50}$$

Thus $\varphi'(y_j - y_k) = 0$ for $k \geq k_{m+1} + 1$ and $k \leq k_m$. So we can write

$$y_j + \frac{t_m}{k_{m+1} - k_m} \sum_{k=1}^{n} \varphi'(y_j - y_k) = x_j. \tag{3.51}$$

Now taking $\frac{t_m}{k_{m+1}-k_m} = \frac{t}{n}$ yields

$$y_j + \frac{t}{n} \sum_{k=1}^{n} \varphi'(y_j - y_k) = x_j. \tag{3.52}$$

and this is true if and only if $\tilde{E}_n^t(x_1, \ldots, x_n) = (y_1, \ldots, y_n)$. □

This allows us to look at asymptotic behavior for each $\tilde{E}^{t_m}_{(j_m - j_{m-1})}$, which allows us to prove the main result of this section:

**Theorem 3.16** *Suppose*

$$\begin{aligned} \mathbf{x}_{k+1} &= \tilde{E}_n^t(\mathbf{x}_k) \\ \mathbf{x}_0 &\in \mathbb{R}^n \end{aligned} \tag{3.53}$$

*where $\tilde{E}_n^t : \mathbb{R}^n \to \mathbb{R}^n$ is as defined in (3.1) and $\varphi$ meets Assumption 3.3 Then every sufficiently spread out (see Definition 3.5) fixed point $\mathbf{x}^*$ of the dynamical system is asymptotically stable.*

***Proof*** Let $\mathbf{x}^*$ be sufficiently spread out fixed point of $\tilde{E}_n^t$. Then it has the following form

$$x_1^* = \cdots = x_{k_1}^* < x_{k_1+1}^* = \cdots = x_{k_2}^* < \cdots < x_{k_i-1}^* = x_{k_i}^*$$

with $x_{k_m+1}^* - x_{k_m}^* > r$. In other words, $\mathbf{x}^* = \left(\mathbf{x}^{1*}, \ldots, \mathbf{x}^{i*}\right)$, where $\mathbf{x}^{m*} = x_{k_m}^*(1, \ldots, 1)$. Using Lemma 3.15, for $|\mathbf{x} - \mathbf{x}^*| < \varepsilon$ with $\varepsilon > 0$ sufficiently small,

$$\tilde{E}_n^t(\mathbf{x}) = (\tilde{E}_{k_1}^{t_1}(x_1, \ldots, x_{k_1}), \tilde{E}_{k_2-k_1}^{t_2}(x_{k_1+1}, \ldots, x_{k_2}), \ldots, \tilde{E}_{k_i-k_{i-1}}^{t_i}(x_{k_{i-1}+1}, \ldots, x_{k_i})) \tag{3.54}$$

Then by Corollary 3.14, a fixed point $\mathbf{x}^{m*} \in \operatorname{span}\{(1, \ldots, 1)\}$ of the dynamical system with $\tilde{E}^{t_m}_{(j_m - j_{m-1})}$ is asymptotically stable. So $\mathbf{x}^* = (\mathbf{x}^{1*}, \mathbf{x}^{2*}, \ldots, \mathbf{x}^{i*})$ is asymptotically stable for the dynamical system with $\tilde{E}_n^t$. □

## 4 Numerical Simulations

In this section we report on some numerical simulations, based on the algorithm outlined at the end of Sect. 2, that confirm the main result of Sect. 3. The structure of the algorithm is as follows.

(1) For an initial measure $m = \sum_{j=1}^{n} a_j \delta_{x_j}$, let $\mathbf{x}_0 = (x_1, \ldots, x_n)$. Let $\varepsilon$ be a fixed small error (we took $\varepsilon = 10^{-6}$) and let $N$ be the number of iterations (e.g. in the experiments below, $N$ is taken in the range [1500, 13000]). Let $t > 0$ be a fixed small time horizon such that the map $E_n^t$ is well-defined, and let $\alpha = \frac{tL_1}{1-t\lambda_1(G)}$ be a Lipschitz constant of $\tilde{F}_{\mathbf{x}}^t$, as guaranteed by Theorem 2.1.

(2) For $k = 1, \ldots, N$ calculate $\mathbf{x}_k$ as follows:

 (a) Set $\mathbf{z}_0 = \mathbf{x}_{k-1}$.

 (b) For $\ell = 1, 2, \ldots$ define $\mathbf{z}_\ell$ as follows:

  (i) For $j = 1, \ldots, n$, use Newton's method to compute an approximation $z_{\ell,j} \approx \tilde{y}_t^{\mathbf{z}_{\ell-1}}(z_{0,j})$, where $z_{\ell,j}$ is the $j$th component of $\mathbf{z}_\ell$, so that the error is at most $\frac{1}{n}\alpha^\ell$.

  (ii) The vector $\mathbf{z}_\ell$ so defined necessarily satisfies $\left\| \mathbf{z}_\ell - \tilde{F}_{\mathbf{z}_0}^t(\mathbf{z}_{\ell-1}) \right\| \le \alpha^\ell$.

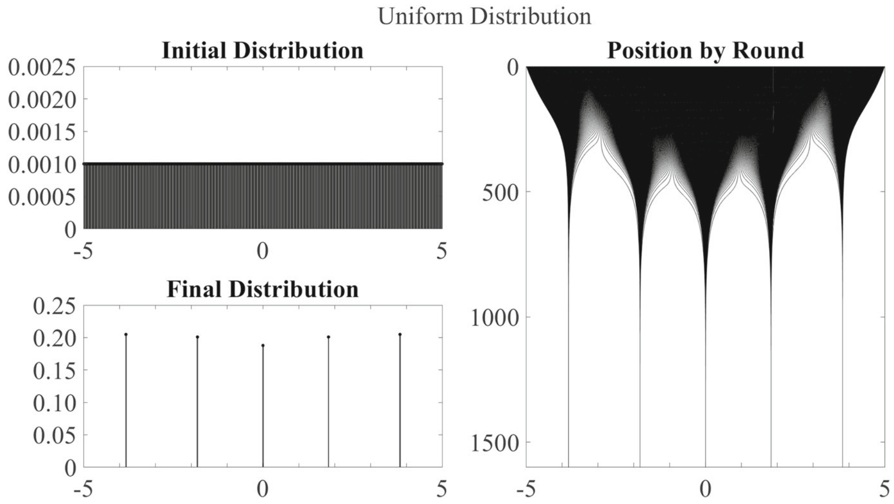
 (c) Repeat until $\alpha^\ell \left( \frac{2(\ell - \alpha\ell + 1)}{(1-\alpha)^2} + \frac{\|\mathbf{z}_1 - \mathbf{z}_0\|}{1-\alpha} \right) \le \varepsilon$, then set $\mathbf{x}_k = \mathbf{z}_\ell$.

 By Lemma 2.3 we necessarily have $\left\| \mathbf{x}_k - \tilde{E}_n^t(\mathbf{x}_{k-1}) \right\| \le \varepsilon$.

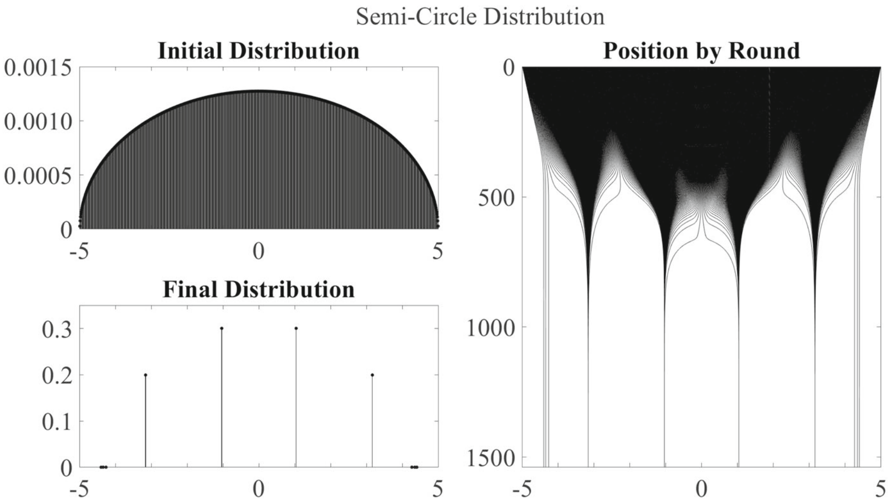
We examined four different empirical population measures distributed across a vector of 1000 evenly spaced points between [−4.995, 4.995]. The coupling function was given to be $G(x, m) = \int \varphi(x - z)\mathrm{d}m(z)$ where $-\varphi$ is a standard "bump function" of the form

$$-\varphi(x) = \begin{cases} \exp\left(-\frac{1}{1-x^2}\right), & -1 < x < 1, \\ 0, & \text{otherwise}, \end{cases} \tag{4.1}$$

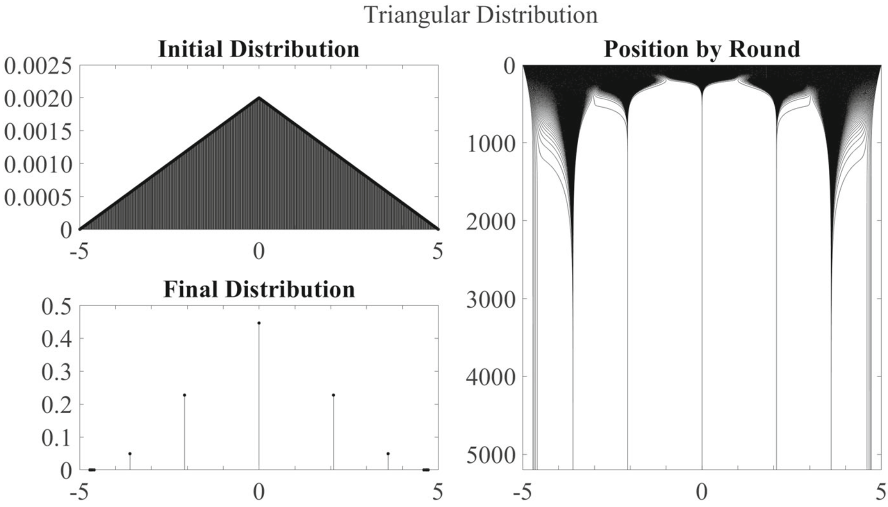
which satisfies Assumption 3.3. The radius of the support of $\varphi$ is 1, and so any fixed point of the dynamical system is "sufficiently spread out" by Definition 3.5 as long as the clusters have distance greater than 1 between them.



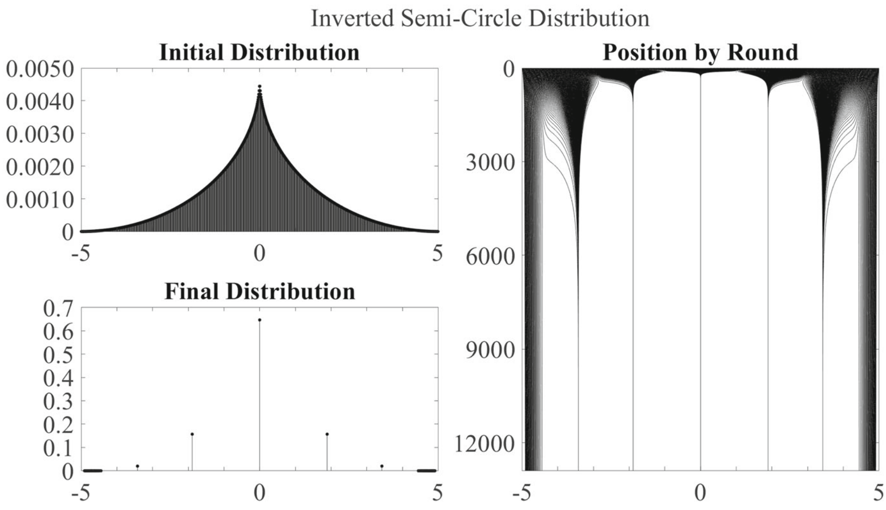
| Group | Coalescing point | Algorithm iterations to coalesce | Total % of population | Range of initial positions | Average initial position | Average total movement | Average total cost |
|---|---|---|---|---|---|---|---|
| 1 | −3.8225 | 1260 | 20.50 | [−4.9950, −2.9550] | −3.9750 | 0.5335 | −241.9120 |
| 2 | −1.8275 | 1417 | 20.10 | [−2.9450, −0.9450] | −1.9450 | 0.5134 | −225.6389 |
| 3 | 0.0000 | 1508 | 18.80 | [−0.9350, 0.9350] | 0.0000 | 0.4700 | −208.5507 |
| 4 | 1.8275 | 1417 | 20.10 | [0.9450, 2.9450] | 1.9450 | 0.5134 | −225.6389 |
| 5 | 3.8225 | 1260 | 20.50 | [2.9550, 4.9950] | 3.9750 | 0.5335 | −241.9120 |




| Group | Coalescing point | Algorithm iterations to coalesce | Total % of population | Range of initial positions | Average initial position | Average total movement | Average total cost |
|---|---|---|---|---|---|---|---|
| 1 | −3.1619 | 1535 | 19.94 | [−4.9650, −2.4650] | −3.4983 | 0.6729 | −227.6249 |
| 2 | −1.0379 | 1169 | 30.03 | [−2.4550, −0.0050] | −1.2026 | 0.6533 | −321.4321 |
| 3 | 1.0379 | 1169 | 30.03 | [0.0050, 2.4550] | 1.2026 | 0.6533 | -321.4321 |
| 4 | 3.1619 | 1535 | 19.94 | [2.4650, 4.9650] | 3.4983 | 0.6729 | -227.6249 |




| Group | Coalescing point | Algorithm iterations to coalesce | Total % of population | Range of initial positions | Average initial position | Average total movement | Average total cost |
|---|---|---|---|---|---|---|---|
| 1 | −3.5989 | 5131 | 4.93 | [−4.9550, −3.4350] | −3.9527 | 0.6052 | −206.4532 |
| 2 | −2.0759 | 1355 | 22.75 | [−3.4250, −1.2850] | −2.2094 | 0.6614 | −943.5018 |
| 3 | 0.0000 | 639 | 44.65 | [−1.2750, 1.2750] | 0.0000 | 0.6087 | −1869.5074 |
| 4 | 2.0759 | 1355 | 22.75 | [1.2850, 3.4250] | 2.2094 | 0.6614 | −943.5018 |
| 5 | 3.5989 | 5131 | 4.93 | [3.4350, 4.9550] | 3.9527 | 0.6052 | −206.4532 |



| Group | Coalescing point | Algorithm iterations to coalesce | Total % of population | Range of initial positions | Average initial position | Average total movement | Average total cost |
|---|---|---|---|---|---|---|---|
| 1 | −3.4234 | 12806 | 2.00 | [−4.6650, −3.1550] | −3.6028 | 0.4675 | −209.4471 |
| 2 | −1.8900 | 1604 | 15.68 | [−3.1450, −1.2850] | −1.9927 | 0.5479 | −1653.3058 |
| 3 | 0.0000 | 406 | 64.62 | [−1.2750, 1.2750] | 0.0000 | 0.5369 | −6841.9609 |
| 4 | 1.8900 | 1604 | 15.68 | [1.2850, 3.1450] | 1.9927 | 0.5479 | −1653.3058 |
| 5 | 3.4234 | 12806 | 2.00 | [3.1550, 4.6650] | 3.6028 | 0.4675 | −209.4471 |

In the graphs above, the initial and final distribution are graphed by plotting points $(x, y)$ where $x$ is the location and $y$ is the total proportion of the population holding position $x$. The "Position by Round" graph displays only the distribution of $x$-values at each round, starting from the top (the population density itself is not directly visible in this graph). In all four cases, the iterative process of each agent minimizing their own total cost function divided the majority of the population into clusters, but only the Uniform Distribution simulation saw every agent eventually belong to a cluster. The other three distributions ended the simulation with a small number of isolated agents at the extreme ends of each distribution in the lowest density regions (i.e. the simulation ended before reaching a stable fixed point). These isolated agents represented 0.0431% of the Semi-Circle Distribution, 0.0065% of the Triangular Distribution, and 0.0224% of the Inverted Semi-Circle Distribution. Isolated agents in these low density regions begin the iterative process with the same ingathering behavior as their interior neighbors and move towards the distribution's mean for many iterations. However as their inlying neighbors start moving inwards faster towards a nearby accumulating mass of other agents, the isolated agents get left behind, and the innermost isolated agents begin moving outwards towards the still ingathering, most extreme isolated agents. These low density regions of isolated agents may eventually coalesce into groups, but the meager benefit of gathering in such low density regions results in incredibly slow movement from residing agents. Consequently, waiting for these regions to coalesce would take more iterations than can be reasonably observed.

Conversely, agents in the highest density regions tend to coalesce the fastest. Group 3 in the Inverted Semi-Circle Distribution simulation comprised 64.62% of the population and coalesced in a speedy 406 iterations. Altogether, the Uniform Distribution had completely coalesced into clusters after 1508 iterations with its tails coalescing first after 1260 iterations. This observed behavior of the highest density regions coalescing the fastest and lowest density regions moving too slow to coalesce completely aligns with the structure of the individual's total cost function. High density regions present a larger reward for gathering, so individual agents can justify more movement in a single iteration. The opposite is true for low density regions. As a result, the region centered about the mean contained the largest group at the end of the simulation in almost every case. The Uniform Distribution simulation notably deviates from this trend with the outermost groups containing 1.7% more of the population than the central group at the mean. This is a direct result of the distribution's fat tails. Agents in the tail ends of the distribution are markedly incentivized to get away from the empty regions next to them and generally move inwards quickly until enough of their inlying neighbors find their accumulating mass sufficiently attractive to meet them to form a single mass. In the Uniform Distribution simulation, this quick piling up of the already dense tails produces a large, attractive mass that outsizes the more inward accumulating masses.

Moreover, agent mobility varied across distribution and final group assignment. The average total movement across all iterations to coalescence for a single agent was 0.5135 in the Uniform Distribution simulation, 0.6609 in the Semi-Circle Distribution simulation, 0.6323 in the Triangular Distribution simulation, and 0.5374 in the Inverted Semi-Circle Distribution simulation. Due to their fatter tails, the Uniform and Semi-Circle Distribution simulations had decreasing average total movement from the outermost groups to the innermost group(s), which is a result of the outermost agents trying to rapidly get away from the empty region next to them. The least mobile groups in the Triangular and Inverted Semi-Circle Distribution simulations with thinner tails were the outermost groups, but similar to the Uniform and Semi-Circle Distribution simulations, the middling groups had a higher average total movement then the innermost group. Low population density in the outermost regions slows down the groups and negatively affects the groups' average total movement. Meanwhile, the middling regions may not have completely empty regions next to them, the outermost regions are comparatively empty to the innermost regions, producing the same effect on the middling and innermost groups as the Uniform and Semi-Circle Distribution simulations. Within groups, agents that happened to begin the simulations near the coalescing point for their group did not have to move very far to converge on their final position, while agents that happened to begin further away from the coalescing point for their group had to move significantly to converge on their final position.

Population density had the biggest impact on average total cost. Groups with the highest population density had the lowest average total cost accumulated across all rounds to coalescence. Like total movement for individual agents, the agents that happened to begin the simulations near the coalescing point for their group had lower total cost than the agents that happened to begin the simulations farther away from the coalescing point for their group.

Finally, in every case we examined, the final distribution was narrower than the initial distribution, but not every agent consistently moved towards the mean. 35.20% of agents in the Uniform Distribution, 41.04% of agents in the Semi-Circle Distribution, 23.26% of agents in the Triangular Distribution, and 16.96% of agents in the Inverted Semi-Circle Distribution ended up farther away from the mean of the distribution in their final position than in their initial position. However, the final coalescing point of every group is closer to the mean of each distribution than their average initial position. The narrowing of the population in addition to the difference between the final coalescing point and average initial position of each group indicates an overall mean-convergent behavior, but the existence of separate groups that do not all converge to the mean as well as the not insignificant portion of each population distribution moving away from the mean implies something interesting is going on. If the population in all four simulations has an overall ingathering behavior, why do distinct groups form? The observed effect of the iterative process of each agent minimizing their own total cost function is not wholesale "polarization" because the vast majority of agents end the simulation in more moderate positions than they started in. Rather in all four simulations, we can see the flight of the relative-moderate who is attracted to the mass of agents with more extreme positions that are willing to compromise and exhibit mean-converging behavior.

Altogether, a population distribution's density and concavity affects the formation of groups, especially at the tails.

## 5 Conclusion

In this paper we have studied a simple mean field game in which players have an incentive to congregate, and we have analyzed the resulting dynamical system formed by iterating the game. We have identified the fixed points and rigorously proved the asymptotic stability for this dynamical system. Our numerical simulations both demonstrate the validity of these results and also provide thought-provoking details about the dynamics. We now conclude this paper with some open questions for both theory and applications.

### 5.1 Equilibrium for a Long Time Horizon

One natural question is whether the stable fixed points of the dynamical system (3.1) correspond (at least approximately) to Nash equilibrium points for the game with a very long time horizon. A heuristic argument in favor of this conjecture is as follows. We take $G(y,\mu) = \int \varphi(y-z)\mathrm{d}\mu(z)$ and consider its discrete analog $g(y,\mathbf{z}) = \frac{1}{n}\sum_{k=1}^{n}\varphi(y-z_k)$. Recall that the cost to each player moving from position $x$ to $y$ is

$$\frac{|x-y|^2}{2t} + \frac{1}{n}\sum_{k=1}^{n}\varphi(y-z_k).$$

Taking $t \to \infty$, we see that $x$ is almost irrelevant. The optimal strategy is essentially to find a minimizer of $y \mapsto \sum_{k=1}^{n}\varphi(y-z_k)$. If $\mathbf{z} = (z_1,\dots,z_n)$ is to represent an equilibrium measure, then each $z_j$ must in fact be a minimizer of this function. Assume $-\varphi$ is a bump function as in (4.1). Glancing at the graph of $y \mapsto \sum_{k=1}^{n}\varphi(y-z_k)$ suggests $\mathbf{z}$ is an equilibrium if and only if the points $z_1,\dots,z_n$ form equal sized clusters $z_1 = \cdots = z_k < z_{k+1} = \cdots = z_{2k} < \cdots < z_{mk}$ such that $z_{(j+1)k} - z_{jk}$ is sufficiently large for each $j$. Hence $\mathbf{z}$ should also be a stable fixed point of the dynamical system (3.1) according to Theorem 3.4. Since we cannot expect the equilibrium to be unique for an arbitrary time horizon, this infinite time limit case might provide some meaningful way of selecting among multiple equilibria for large time horizon games. We intend to investigate this in future research.

### 5.2 Stability for Non-empirical Measures

A natural extension of this paper would be to consider the same question of stability analysis for non-empirical measures. That is, for $m$ a non-empirical measure, what is $\lim_{k\to\infty}\left(E^t\right)^k(m)$? It is not clear whether fixed points are stable, since the eigenvalues computed in Sect. 3 could possibly move toward 1 as $n \to \infty$. It is nevertheless tempting to conjecture that the limit $\lim_{k\to\infty}\left(E^t\right)^k(m)$ can in many cases be well-approximated by discretizing $m$.

### 5.3 Opinion Dynamics

One particularly compelling application for our model is where the population measure represents the distribution of people on a spectrum of opinions, e.g. a single-issue political spectrum. The iterated game could model a process in which people change their positions incrementally, trying to find communities with valuable social identities of like-minded peers, or perhaps trying to build coalitions to win political capital. On an optimistic note, we observe that in our simulations the entire population becomes more moderate as a whole, and most individuals elect to compromise to join less extreme communities. However, the population's separation into distinct, consolidated groups gives dissidents platforms with greater leverage than they began with, and the most extreme dissidents are left behind entirely with their voices completely drowned out by more popular groups. Our model shows that the individual proclivity to group can reshape a population at large, but the effects of the population's new shape on society are harder to determine.

It would be interesting to use game theoretic models to further investigate the impact of grouping on political or other social systems.

**Author Contributions** The authors have each contributed equally to the present manuscript.

**Funding** The authors gratefully acknowledge the support of the National Science Foundation through research grant DMS-2045027.

**Data Availability** Data sharing not applicable to this article as no external datasets were generated or analyzed during the current study.

## Declarations

**Conflict of interest** The authors declare that they do not have any Conflict of interest.

**Ethical Approval** Not applicable.

## References

1. Bardi M, Cardaliaguet P (2021) Convergence of some mean field games systems to aggregation and flocking models. Nonlinear Anal 204:112199
2. Bardi M, Fischer Markus (2019) On non-uniqueness and uniqueness of solutions in finite-horizon mean field games. ESAIM: Control Optim Calc Var 25:44
3. Bensoussan A, Frehse J, Yam P (2013) Mean field games and mean field type control theory. Springer
4. Bayraktar E, Zhang X (2020) On non-uniqueness in mean field games. Proceed Am Math Soc 148(9):4091–4106
5. Cardaliaguet P (2010) Notes on mean field games. From P.-L. Lions' lectures at College de France
6. Carmona R, Delarue F (2017) Probabilistic theory of mean field games: vol. I, mean field FBSDEs, control, and games. Springer Verlag, Berlin
7. Carmona R, Delarue F (2017) Probabilistic theory of mean field games: vol. II, mean field games with common noise and master equations. Springer Verlag, Berlin
8. Cirant Marco (2016) Stationary focusing mean-field games. Comm Partial Differ Equ 41(8):1324–1346
9. Coleman JS (1994) Found Soc Theory. Harvard University Press, Cambridge
10. Converse Philip E (1964) The nature of belief systems in mass publics. In: Apter David (ed) Ideology and Discontent. Free Press, New York
11. Cirant Marco, Tonon Daniela (2019) Time-dependent focusing mean-field games: the sub-critical case. J Dyn Diff Equat 31:49–79
12. Degond Pierre, Herty Michael, Liu Jian-Guo (2017) Mean-field games and model predictive control. Commun Math Sci 15(5):1403–1422
13. Forsyth Donelson R (2014) Group dynamics. Wadsworth Cengage Learning, Boston
14. Lasry Jean-Michel, Lions Pierre-Louis (2007) Mean field games. Japan J Math 2(1):229–260
15. Stuart Andrew, Humphries Anthony R (1998) Dynamical systems and numerical analysis, vol 2. Cambridge University Press, Cambridge
16. Van Dijk Teun A (2013) Ideology and discourse analysis. In: Van D, Teun A (eds) The Meaning of Ideology. Routledge, London, pp 110–135