Dynamics of Strategic Agents and Algorithms as PDEs

Franca Hoffmann

Computing and Mathematical Sciences
California Institute of Technology

CIRM, France

Aggregation-Diffusion Equations & Collective Behavior: Analysis, Numerics and Applications

Caltech

collaboration with:

Lauren Conger, Eric Mazumdar, Lillian Ratliff

NeurIPS 2023

https://arxiv.org/abs/2307.01166

social media

RESEARCH ARTICLE | SOCIAL SCIENCES

Exposure to opposing views on social media can increase political polarization

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, +6, and Alexander Volfovsky Authors Info & Affiliations

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved August 9, 2018 (received for review March 20, 2018)

August 28, 2018 | 115 (37) 9216-9221 | https://doi.org/10.1073/pnas.1804840115

Agents and Algorithms

social media

RESEARCH ARTICLE | SOCIAL SCIENCES |

Exposure to opposing views on social media can increase political polarization

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, +6, and Alexander Volfovsky Authors Info & Affiliations

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved August 9, 2018 (received for review March 20, 2018)

August 28, 2018 | 115 (37) 9216-9221 | https://doi.org/10.1073/pnas.1804840115

finance

The New York Times

'Dumb Money' Is on GameStop, and It's Beating Wall Street at Its Own Game

GameStop shares have soared 1,700 percent as millions of small investors, egged on by social media, employ a classic Wall Street tactic to put the squeeze — on Wall Street.

Agents and Algorithms

social media

RESEARCH ARTICLE | SOCIAL SCIENCES |

Exposure to opposing views on social media can increase political polarization

Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, +6, and Alexander Volfovsky Authors Info & Affiliations

Edited by Peter S. Bearman, Columbia University, New York, NY, and approved August 9, 2018 (received for review March 20, 2018)

August 28, 2018 | 115 (37) 9216-9221 | https://doi.org/10.1073/pnas.1804840115

finance

The New York Times

'Dumb Money' Is on GameStop, and It's Beating Wall Street at Its Own Game

GameStop shares have soared 1,700 percent as millions of small investors, egged on by social media, employ a classic Wall Street tactic to put the squeeze — on Wall Street.

gig economy

Uber, Lyft drivers coordinate to manipulate surge pricing at Virginia airport over pay concerns: Report

By Soo Youn, ABC News

Saturday, May 18, 2019

Distribution Shift

Strategic agents interact with an algorithm:

- maximizing utility, minimizing loss (depends on learning algorithm).

Agents adjust their attributes

- according to their objectives;
- random disturbances;
- due to interactions and pressures to become more or less similar;
- due to exogenous forces.

Consequence: distribution shifts.

Distribution Shift

Strategic agents interact with an algorithm:

- maximizing utility, minimizing loss (depends on learning algorithm).

Agents adjust their attributes

- according to their objectives;
- random disturbances;
- due to interactions and pressures to become more or less similar;
- due to exogenous forces.

Consequence: distribution shifts.

Idea: formulate the game theoretic set-up in the language of PDE analysis.

Modeling distribution shifts

- Simpler models miss out on important features such as polarization.

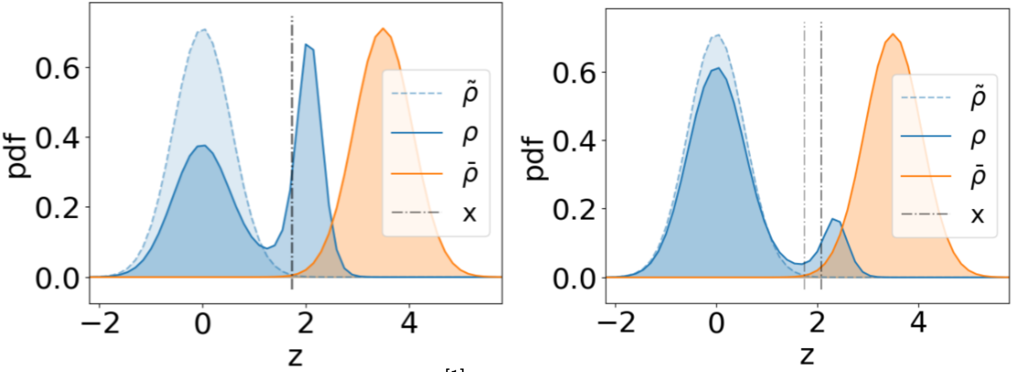


Figure: mean shift[1] (left) vs PDE model (right)

[1] John P. Miller, Juan C. Perdomo, and Tijana Zrnic. Outside the Echo Chamber: Optimizing the Performative Risk. en. In *Proceedings of the 38th International Conference on Machine Learning*. PMLR, 2021.

PDE model for distribution shift

SDE for individual agents

$$\mathrm{d}Z_t^{(i)} = -\nabla V(Z_t^{(i)}, x)\mathrm{d}t - \frac{1}{N}\sum_{j=1}^{N} \nabla W(Z_t^{(i)}, Z_t^{(j)})\mathrm{d}t + \sqrt{2\alpha}\mathrm{d}B_t^{(i)}$$

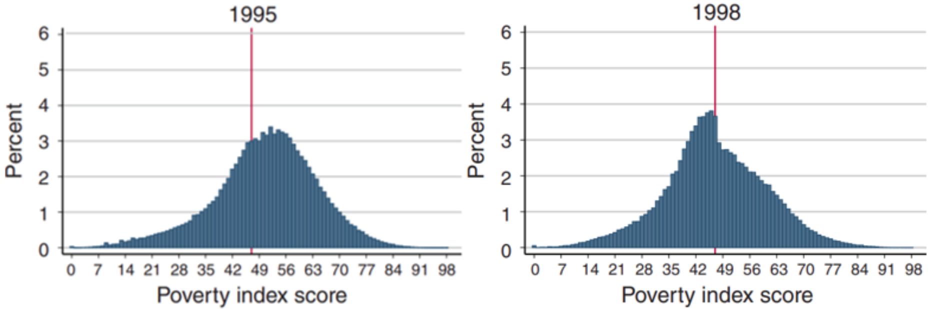
PDE for distribution of agents

$$\partial\rho = \nabla \cdot \left(\rho\nabla\left(V(z,x) + W * \rho + \alpha\log\rho\right)\right) = -\nabla_{W_2} F(\rho, x)$$

- W_2 gradient flow for $F(\rho, x) := \int V(z,x)\,\mathrm{d}\rho(z) + \frac{1}{2}\int \rho W * \rho + \alpha\int \rho\log\rho$,
- $V(z,x) = -U(z,x) - \alpha\log\tilde{\rho}(z)$,
- $U(z,x)$ utility maximized by agents,
- penalization to deviate from reference measure $\tilde{\rho}(z)$.

Does the model describe real-world data?

Study of Census Data in Colombia from 1995 to 2003[2]

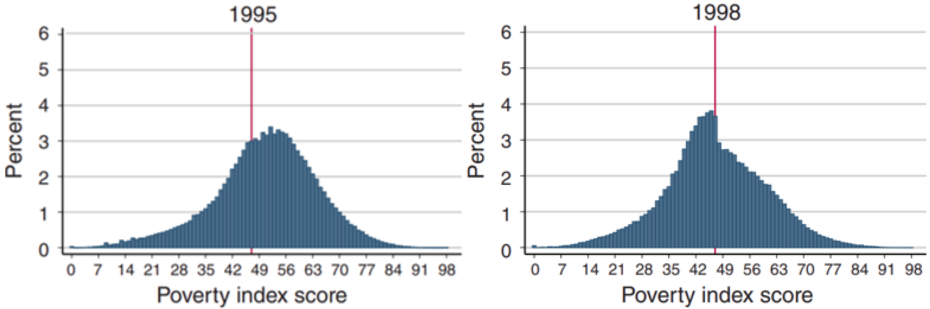


- Local officials misreported data in order to obtain lower poverty index scores for their constituents.
- Households with a poverty index score below a given threshold receive government aid.
- Release of algorithm for the poverty index score and threshold (1997) shifts distribution of scores.

[2] Adriana CAMACHO and Emily CONOVER. Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy*, 2011.

Does the model describe real-world data?

Study of Census Data in Colombia from 1995 to 2003[3]



- Classification problem: an algorithm aims to separate poverty index scores into ones which qualify for aid and ones that do not. The algorithm cannot observe the true labels for each data point.
- Each family aims to be classified as qualifying for aid, regardless of their true label.

---

[3] Adriana Camacho and Emily Conover. Manipulation of Social Program Eligibility. *American Economic Journal: Economic Policy*, 2011.

Does the model describe real-world data?

$$\partial_t \rho = \nabla \cdot \left(\rho \nabla \left(V(z,x) + \frac{1}{10} \log \rho \right) \right), \qquad p(z,x) = (1 + \exp(-2(z-x)))^{-1},$$

$$V(z,x) = 1 - q(z,x) + \int q(z,x)\bar{\rho}(z) - \frac{1}{10} \log \tilde{\rho} + \frac{6z}{100}, \qquad q(z,x) = 1 - p(z,x)$$

- $d = 1$, $\rho_0 = \tilde{\rho} = \mathcal{N}(54, 10)$, $\bar{\rho} = \rho^{(95)} - \rho_0$.
- $q(z,x)$ is the probability that the classifier assigns a label of "qualified" to a family with attributes z and classifier parameters x.
- Families aim to maximize their probability of such a classification.
- The term $\frac{6z}{100}$ models a preference for a lower poverty index score, regardless of the classifier parameters.

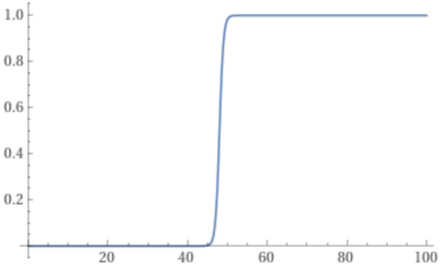
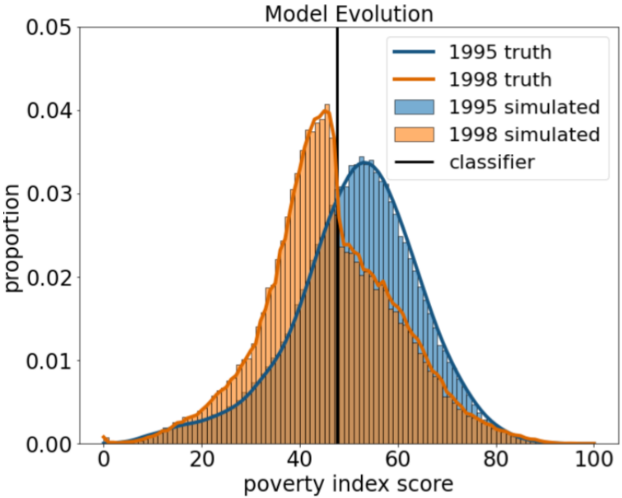


Figure: $p(z, 48)$

Does the model describe real-world data?



Interplay between retraining and distribution shift

Retraining of algorithm: $x(t)$ is minimizing loss $L(z,x)$

$$\dot{x} = -\nabla_x \left[\mathbb{E}_{z\sim\rho} L(z,x)\right] \quad \Longrightarrow \quad \partial_t \mu = \nabla \cdot \left(\mu \nabla \int L(z,x) \mathrm{d}\rho(t,z)\right)$$

- **Aligned Objectives:**

$$\partial_t \rho = -\nabla_{W_2,\rho} G_a(\rho,\mu)\,, \qquad \partial_t \mu = -\nabla_{W_2,\mu} G_a(\rho,\mu)\,,$$

- **Competing Objectives:**

$$\partial_t \rho = +\nabla_{W_2,\rho} G_c(\rho,\mu)\,, \qquad \partial_t \mu = -\nabla_{W_2,\mu} G_c(\rho,\mu)\,,$$

$\longrightarrow$ bridge between mathematical biology and game theory!

Aligned Objectives

Coupled gradient flow for $\gamma = (\rho, \mu)$ with respect to $\bar{W}(\gamma, \gamma')^2 = W_2(\rho, \rho')^2 + W_2(\mu, \mu')^2$

$$\partial_t \gamma = -\nabla_{\bar{W}} G_a(\gamma)$$

Energy functional

$$\begin{aligned} G_a(\rho, \mu) = &\iint f_1(z, x) \mathrm{d}\rho(z) \mathrm{d}\mu(x) + \iint f_2(z, x) \mathrm{d}\bar{\rho}(z) \mathrm{d}\mu(x) \\ &+ \frac{\beta}{2} \int \|x - x_0\|^2 \mathrm{d}\mu(x) + \frac{1}{2} \int \rho W * \rho + \alpha \int \rho \log\left(\frac{\rho}{\tilde{\rho}}\right) . \end{aligned}$$

Example: classification of 1D features

- $\rho(t, z) =$ true label 0; $\bar{\rho}(z) =$ true label 1 (fixed);
- $\mu(t, x) = \delta_{x(t)}$ algorithm threshold below which agents are classified as having label 0;
- $\tilde{\rho}(z) =$ fixed reference distribution: agents (true label 0) are penalized with strength α to deviate from $\tilde{\rho}$;
- $x_0 =$ fixed algorithm threshold : the algorithm is penalized with strength β to deviate from x_0.

Aligned Objectives

Coupled gradient flow for $\gamma = (\rho, \mu)$ with respect to $\bar{W}(\gamma, \gamma')^2 = W_2(\rho, \rho')^2 + W_2(\mu, \mu')^2$

$$\partial_t \rho = \alpha \Delta \rho + \operatorname{div}\left(\rho \nabla_z \left(\int f_1 \mathrm{d}\mu - \alpha \log \tilde{\rho} + W * \rho\right)\right),$$

$$\partial_t \mu = \operatorname{div}\left(\mu \nabla_x \left(\int f_1 \mathrm{d}\rho + \int f_2 \mathrm{d}\bar{\rho} + \frac{\beta}{2}\|x - x_0\|^2\right)\right).$$

Energy functional

$$G_a(\rho, \mu) = \iint f_1(z, x)\mathrm{d}\rho(z)\mathrm{d}\mu(x) + \iint f_2(z, x)\mathrm{d}\bar{\rho}(z)\mathrm{d}\mu(x)$$
$$+ \frac{\beta}{2}\int \|x - x_0\|^2 \mathrm{d}\mu(x) + \frac{1}{2}\int \rho W * \rho + \alpha \int \rho \log\left(\frac{\rho}{\tilde{\rho}}\right).$$

Example: recommender systems with 1D features

- $\rho(t, z) =$ true label 0; $\bar{\rho}(z) =$ true label 1 (fixed);
- $\mu(t, x) = \delta_{x(t)}$ algorithm threshold below which agents are classified as having label 0;
- $\tilde{\rho}(z) =$ fixed reference distribution: agents (true label 0) are penalized with strength α to deviate from $\tilde{\rho}$;
- $x_0 =$ fixed algorithm threshold : the algorithm is penalized with strength β to deviate from x_0.

Aligned Objectives

Coupled gradient flow for $\gamma = (\rho, \mu)$ with respect to $\overline{W}(\gamma, \gamma')^2 = W_2(\rho, \rho')^2 + W_2(\mu, \mu')^2$

$$\partial_t \gamma = -\nabla_{\overline{W}} G_a(\gamma)$$

Energy functional

$$\begin{aligned} G_a(\rho, \mu) = & \iint f_1(z, x) \mathrm{d}\rho(z) \mathrm{d}\mu(x) + \iint f_2(z, x) \mathrm{d}\bar{\rho}(z) \mathrm{d}\mu(x) \\ & + \frac{\beta}{2} \int \|x - x_0\|^2 \mathrm{d}\mu(x) + \frac{1}{2} \int \rho W * \rho + \alpha \int \rho \log\left(\frac{\rho}{\tilde{\rho}}\right) . \end{aligned}$$

Example: recommender systems with 1D features

- $f_1(z, x) =$ cost if agents (label 0) have attributes z and algorithm has parameters x;
- $f_2(z, x) =$ cost if agents (label 1) have attributes z and algorithm has parameters x;
- $W(z) =$ interaction potential for agents (label 0).

Aligned Objectives: Assumptions[4]

Convexity of f_1 and f_2

The functions $f_1, f_2 \in C^2(\mathbb{R}^d \times \mathbb{R}^d; [0,\infty))$ satisfy for all $(z,x) \in \mathbb{R}^d \times \mathbb{R}^d$ the following:

- There exists constants $\lambda_1, \lambda_2 \geq 0$ such that Hess $(f_1) \succeq \lambda_1 \mathrm{Id}_{2d}$ and $\nabla_x^2 f_2 \succeq \lambda_2 \mathrm{Id}_d$;
- There exist constants $a_i > 0$ such that $x \cdot \nabla_x f_i(z,x) \geq -a_i$ for $i = 1,2$;

Reference Distribution

There exists $\tilde{\lambda} > 0$ s.t. $\tilde{\rho} \in \mathcal{P}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$ satisfies $\log \tilde{\rho} \in C^2(\mathbb{R}^d)$ and $\nabla_z^2 \log \tilde{\rho}(z) \preceq -\tilde{\lambda}\,\mathrm{Id}_d$.

Convex Interaction Kernel

The interaction kernel $W \in C^2(\mathbb{R}^d; [0,\infty))$ is convex, symmetric $W(-z) = W(z)$, and for some $D > 0$ satisfies

$$z \cdot \nabla_z W(z) \geq -D, \quad |\nabla_z W(z)| \leq D(1 + |z|) \quad \forall\, z \in \mathbb{R}^d.$$

[4] José A. Carrillo, Robert J. McCann, and Cédric Villani. Kinetic equilibration rates for granular media and related equations: entropy dissipation and mass transportation estimates. *Revista Matemática Iberoamericana*, 2003.

Aligned Objectives: Results

Consider solutions $\gamma_t := (\rho_t, \mu_t)$ to the dynamics

$$\partial_t \gamma = -\nabla_{\overline{W}} G_a(\gamma) \tag{2}$$

with initial conditions satisfying $\gamma_0 \in \mathcal{P}_2(\mathbb{R}^d) \times \mathcal{P}_2(\mathbb{R}^d)$ and $G_a(\gamma_0) < \infty$.

Theorem (Existence of steady states)

There exists a unique minimizer $\gamma_\infty = (\rho_\infty, \mu_\infty)$ of G_a, which is also a steady state for equation (2). Moreover, $\rho_\infty \in L^1(\mathbb{R}^d)$, has the same support as $\tilde{\rho}$, and its density is continuous.

Theorem (Convergence)

The solution γ_t converges exponentially fast in $G_a(\cdot \,|\, \gamma_\infty) = G_a(\cdot) - G_a(\gamma_\infty)$ and $\overline{W}$,

$$G_a(\gamma_t \,|\, \gamma_\infty) \le e^{-2\lambda_a t} G_a(\gamma_0 \,|\, \gamma_\infty) \quad \textit{and} \quad \overline{W}(\gamma_t, \gamma_\infty) \le c e^{-\lambda_a t} \quad \textit{for all } t \ge 0\,,$$

where $c > 0$ is a constant only depending on γ_0, γ_∞ and the parameter λ_a,

$$\lambda_a := \lambda_1 + \min(\lambda_2 + \beta, \alpha\tilde{\lambda}) > 0\,.$$

Aligned Objectives: Proof Sketch

- G_a is lower semi-continuous with respect to the weak-$*$ topology.
- G_a uniformly displacement convex with constant $\lambda_a > 0$.
- Existence of minimizer of G_a: direct method in the calculus of variations.
- **HWI inequality:** For the dissipation functional $D_a(\gamma) := \iint |\nabla_{zx}\delta_\gamma G_a(z,x)|^2 \mathrm{d}\gamma(z,x)$,

$$G_a(\gamma_0) - G_a(\gamma_1) \leq \overline{W}(\gamma_0,\gamma_1)\sqrt{D_a(\gamma_0)} - \frac{\lambda_a}{2}\,\overline{W}(\gamma_0,\gamma_1)^2\,.$$

- **Generalized Log-Sobolev inequality:** for the unique minimizer γ_* of G_a,

$$D_a(\gamma) \geq 2\lambda_a\, G_a(\gamma|\gamma_*)\,.$$

- **Talagrand inequality:** for the unique minimizer γ_* of G_a,

$$\overline{W}(\gamma,\gamma_*)^2 \leq \frac{2}{\lambda_a} G_a(\gamma\,|\,\gamma_*)\,.$$

- Differentiating G_a along solutions γ_t:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} G_a(\gamma_t) &= -\int \left\|\nabla_z\delta_\rho G_a[\gamma_t](z)\right\|^2 \mathrm{d}\rho_t(z) - \int \left\|\nabla_x\delta_\mu G_a[\gamma_t](x)\right\|^2 \mathrm{d}\mu_t(x)\\ &= -D_a(\gamma_t) \leq -2\lambda_a G_a(\gamma_t\,|\,\gamma_*) \quad + \text{Gronwall} + \text{Talagrand.}\end{aligned}$$

Competing Objectives

Min-Max problem

$$\partial_t \rho = +\nabla_{W_2,\rho} G_c(\rho,\mu)\,, \qquad \partial_t \mu = -\nabla_{W_2,\mu} G_c(\rho,\mu)\,.$$

Energy functional

$$G_c(\rho,\mu) = \iint f_1(z,x)\mathrm{d}\rho(z)\mathrm{d}\mu(x) + \iint f_2(z,x)\mathrm{d}\bar{\rho}(z)\mathrm{d}\mu(x) \\ + \frac{\beta}{2}\int \|x - x_0\|^2 \mathrm{d}\mu(x) - \frac{1}{2}\int \rho W * \rho - \alpha \int \rho \log\left(\frac{\rho}{\tilde{\rho}}\right)\,.$$

Example: classification of 1D features

- $f_1(z,x) =$ cost if agents (label 0) at z and algorithm at x (now maximized by agents);
- $f_2(z,x) =$ cost if agents (label 1) at z and algorithm at x;
- $W(z) =$ interaction potential for agents (label 0).

Nash Equilibrium

Definition

A pair of measures $\gamma_* = (\rho_*, \mu_*) \in \mathcal{P}(\mathbb{R}^d) \times \mathcal{P}(\mathbb{R}^d)$ is a **Nash equilibrium** for the competitive objective case if it satisfies

$$G_c(\rho_*, \mu_*) \geq G_c(\rho, \mu_*) \quad \forall\ \rho \in \mathcal{P}(\mathbb{R}^d)\,,$$
$$G_c(\rho_*, \mu_*) \leq G_c(\rho_*, \mu) \quad \forall\ \mu \in \mathcal{P}(\mathbb{R}^d)\,.$$

Competing Objectives: Additional Assumptions

Upper bounds for f_1 and f_2

- There exists a constant $\Lambda_1 > 0$ such that

$$\nabla_z^2 f_1(z,x) \preceq \Lambda_1 I_d \qquad \text{for all } (z,x) \in \mathbb{R}^d \times \mathbb{R}^d .$$

- For any $R > 0$ there exists a constant $c_2 = c_2(R) \in \mathbb{R}$ such that

$$\sup_{x \in B_R(0)} \int f_2(z,x) \mathrm{d}\bar{\rho}(z) < c_2 .$$

- The function f_1 satisfies $\|\nabla_{xz}^2 f_1(z,x)\|_2 \leq L$ for all $x, z \in \mathbb{R}^d$ for some $L \geq 0$.
- There exists a constant $\Lambda_2 > 0$ such that $\nabla_x^2 f_1(z,x) \preceq \Lambda_2 \operatorname{Id}_d$ for all $(z,x) \in \mathbb{R}^d \times \mathbb{R}^d$.

Competing Objectives: Results (new)

Consider solutions $\gamma_t := (\rho_t, \mu_t)$ to the dynamics

$$\partial_t \rho = +\nabla_{W_2,\rho} G_c(\rho, \mu)\,, \qquad \partial_t \mu = -\nabla_{W_2,\mu} G_c(\rho, \mu)\,. \tag{4}$$

Theorem (Existence of Nash equilibrium)

There exists a unique critical point γ_* for G_c over $\mathcal{P}$ which is also a steady state for equation (4) and a Nash equilibrium. $\gamma_* = (\rho_*, \mu_*)$ satisfies $\rho_* \in \mathcal{P}_2 \cap L^1_+(\mathbb{R}^d)$ with $\|\rho_*\|_1 = 1$, and $\mu_* \in \mathcal{P}_2$ with $\|\mu_*\|_1 = 1$.

Theorem (Convergence)

Assume $\lambda_b = \alpha\tilde{\lambda} - \Lambda_1 > 0$ and $\lambda_2 + 2L < \min\{\hat{\beta}, \lambda_b\}$ for $\hat{\beta} := \frac{\beta}{2(\|x_0\|+1)}$. Then

$$\overline{W}(\gamma_t, \gamma_*) \le c e^{-\lambda_c t} \quad \textit{for all } t \ge 0\,,$$

where

$$\lambda_c := \frac{1}{2} \min\{\lambda_1 + \lambda_2 + \beta, \lambda_b\} > 0\,.$$

Competing Objectives: Proof Sketch

- Define $\widehat{G}_c(\gamma, \hat{\gamma}) \coloneqq G_c(\hat{\rho}, \mu) - G_c(\rho, \hat{\mu})$.
- G_c is uniformly displacement $(\lambda_1 + \lambda_2 + \beta)$-convex in μ for any fixed $\rho \in \mathcal{P}$.
- G_c is uniformly displacement λ_b-concave in ρ for any fixed $\mu \in \mathcal{P}$.
- $\widehat{G}_c(\gamma, \hat{\gamma})$ is uniformly displacement $2\lambda_c$-convex in γ for any fixed $\hat{\gamma} \in \mathcal{P}$.
- Existence of a unique Nash equilibrium for G_c:
  generalization of the Browder-Ky Fan fixed point theorem for the map

$$B(\gamma) \coloneqq \left\{ \zeta \in \widehat{\mathcal{P}} \,\middle|\, \widehat{G}_c(\zeta, \gamma) = \min_{\pi \in \widehat{\mathcal{P}}} \widehat{G}_c(\pi, \gamma) \right\} .$$

- Any critical point γ_* of G_c is a steady state and ρ_* satisfies $\mathrm{supp}\ (\rho_*) = \mathrm{supp}\ (\tilde{\rho})$.
- Exponential convergence: explicitly differentiate $\overline{W}(\gamma_t, \gamma^*)^2$.

Competing Objectives: Timescale Separation[5]

Recall the dynamics

$$\partial_t \rho = +\nabla_{W_2,\rho} G_c(\rho,\mu)\,, \qquad \partial_t \mu = -\nabla_{W_2,\mu} G_c(\rho,\mu)\,.$$

- **Fast algorithm:** $\mu(t,\cdot) = \delta_{x(t)}$ and

$$\partial_t \rho = +\left.\nabla_{W_2,\rho} G_c(\rho,\delta_x)\right|_{x=b(\rho)}\,, \qquad b(\rho) \coloneqq \operatorname*{argmin}_{\bar{x}} G_c(\rho,\delta_{\bar{x}})\,.$$

  $\longrightarrow$ Define $G_b(\rho) \coloneqq G_c(\rho, b(\rho))$ and consider $\partial_t \rho = +\nabla_{W_2,\rho} G_b(\rho)$.
  $\longrightarrow$ **Examples:** Online advertising, Uber/Lyft, ...

- **Fast agents:** $\mu(t,\cdot) = \delta_{x(t)}$ and

$$r(x(t)) \coloneqq \operatorname*{argmax}_{\hat{\rho}\in\mathcal{P}} G_c(\hat{\rho},\delta_{x(t)})\,, \qquad \frac{\mathrm{d}}{\mathrm{d}t} x(t) = -\left.\nabla_x G_c(\rho,\delta_{x(t)})\right|_{\rho=r(x(t))}\,.$$

  $\longrightarrow$ Define $G_d(\rho) \coloneqq G_c(r(x),\delta_x)$ and consider $\dot{x}(t) = -\nabla_x G_d(x(t))$.
  $\longrightarrow$ **Examples:** Loans, government policies, video game rule updates, ...

[5] Lauren Conger, F H, Eric Mazumdar, and Lillian Ratliff. Strategic distribution shift of interacting agents via coupled gradient flows. In *Advances in Neural Information Processing Systems (NeurIPS)*. Curran Associates, Inc., 2023.

Competing Objectives: Results

Theorem (Fast Algorithm)

(a) *There exists a unique maximizer ρ_∞ of $G_b(\rho)$, which is also a steady state. Moreover, $\rho_\infty \in L^1(\mathbb{R}^d)$, has the same support as $\tilde{\rho}$, and its density is continuous.*

(b) *The solution ρ_t converges exponentially fast to ρ_∞ with rate λ_b in $G_b(\cdot \,|\, \rho_\infty)$ and W_2,*

$$G_b(\rho_t \,|\, \rho_\infty) \le e^{-2\lambda_b t} G_a(\rho_0 \,|\, \rho_\infty) \quad \text{and} \quad W_2(\rho_t, \rho_\infty) \le c e^{-\lambda_b t} \quad \text{for all } t \ge 0\,,$$

where $c > 0$ is a constant only depending on ρ_0, ρ_∞ and the parameter $\lambda_b := \alpha\tilde{\lambda} - \Lambda_1 > 0$.

Theorem (Fast Agents)

(a) *There exists a unique minimizer x_∞ of $G_d(x)$ which is also a steady state.*

(b) *The vector $x(t)$ solving the dynamics with initial condition $x(0) \in \mathbb{R}^d$ converges exponentially fast to x_∞,*

$$\|x(t) - x_\infty\| \le e^{-\lambda_d t} \|x(0) - x_\infty\|\,,$$

$$G_d(x(t)) - G_d(x_\infty) \le e^{-2\lambda_d t} \left(G_d(x(0)) - G_d(x_\infty)\right)$$

for all $t \ge 0$ for parameter $\lambda_d := \lambda_1 + \lambda_2 + \beta > 0$.

Comments on proof

Fast algorithm:

- The best response $b(\rho)$ is uniformly bounded.
- G_b is upper semi-continuous with respect to the weak-$*$ topology.
- G_b is λ_b uniformly displacement concave.
  Remark: concavity of G_b is unknown, use Danskin's Theorem (game theory).
- Existence of maximizer for G_b: direct method in the calculus of variations.
- Any maximizer G_b is a steady state and satisfies $\text{supp }(\rho_*) = \text{supp }(\tilde{\rho})$.
- Convergence follows from functional inequalities as before.

Fast agents:

- For each $x \in \mathbb{R}^d$ there exists a unique maximizer $\rho_* := r(x)$ solving $\text{argmax}_{\hat{\rho}\in\mathcal{P}} G_c(\hat{\rho}, x)$. Further, $r(x) \in L^1(\mathbb{R}^d) \cap \mathcal{P}_2(\mathbb{R}^d)$, $\text{supp }(r(x)) = \text{supp }(\tilde{\rho})$.
- Consider $x_n \to \bar{x}$ in $\mathbb{R}^d$. Define $F_n(\rho) := -G_c(\rho, x_n)$ and $\bar{F}(\rho) := -G_c(\rho, \bar{x})$. Then $F_n \xrightarrow{\Gamma} \bar{F}$ in the narrow topology.
- The best response $r(x) \in \mathcal{P}_2(\mathbb{R}^d)$ is continuous in $x \in \mathbb{R}^d$ in the narrow topology.
- Danskin's Result: We have $\nabla_x G_d(x) = (\nabla_x G_c(\rho, x))|_{\rho=r(x)}$, and G_d is strongly λ_d convex.

Modeling Societal Systems

How can we model population dynamics under algorithmic influence?

Real-world questions can we answer:

- Does the model describe real-world data?
- What dynamical features do we observe?
- How fast should the algorithm learn?
- Is gradient descent an optimal learning strategy for the algorithm?
- When do other state-of-the-art techniques for performative prediction fail?
- What if the algorithm only has access to samples from the population?
- Learning cost and interactions from data?

Modeling Societal Systems

How can we model population dynamics under algorithmic influence?

Real-world questions can we answer:

- Does the model describe real-world data? Yes
- What dynamical features do we observe? e.g. polarization
- How fast should the algorithm learn? timescale selection is critical for design
- Is gradient descent an optimal learning strategy for the algorithm? Not necessarily!
- When do other state-of-the-art techniques for performative prediction fail? Moment models are not sufficiently detailed.
- What if the algorithm only has access to samples from the population? interesting problem for future research!
- Learning cost and interactions from data? kernel methods

$\longrightarrow$ ask Lauren Conger! (Tuesday poster session)

Takeaway: PDEs provide a powerful and necessary tool for understanding strategic distribution shift.

Current Work & Future Directions

- Generalizing results for timescale separated case.
- Adversarial setting without timescale separation.
- N-player game with arbitrary dynamics: competition among multiple algorithms and multiple populations.
- Analysis when the algorithm only has access to population samples.
- Learning PDE dynamics from data.


Thank You