Monotone Multispecies Flows

Franca Hoffmann

Computing and Mathematical Sciences
California Institute of Technology

IMAG, Granada, 18 September 2025

Gradient flows face-to-face

Caltech

Collaboration with



Ricardo Baptista
University of Toronto



Lauren Conger
Caltech



Eric Mazumdar
Caltech



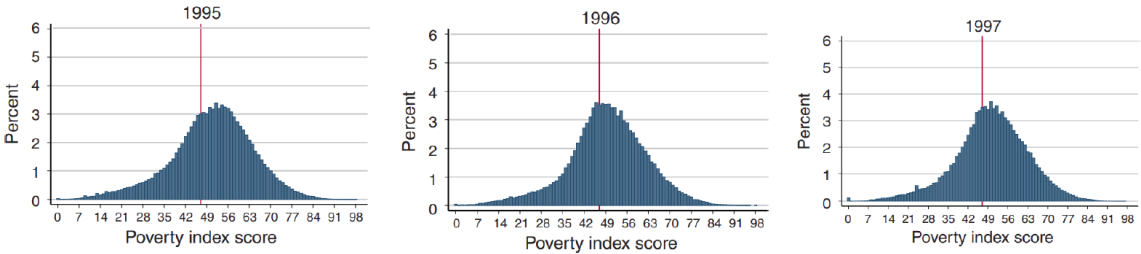
Lillian Ratliff
University of Washington

References:

- Strategic Distribution Shift of Interacting Agents via Coupled Gradient Flows (NeurIPS 2023)
- Coupled Wasserstein Gradient Flows for Min-Max and Cooperative Games (`https://arxiv.org/abs/2411.07403`)
- Computing Optimal Transport Plans via Min-Max Gradient Flows (CDC 2025)
- Monotone Multispecies Flows (`https://arxiv.org/abs/2506.22947`)

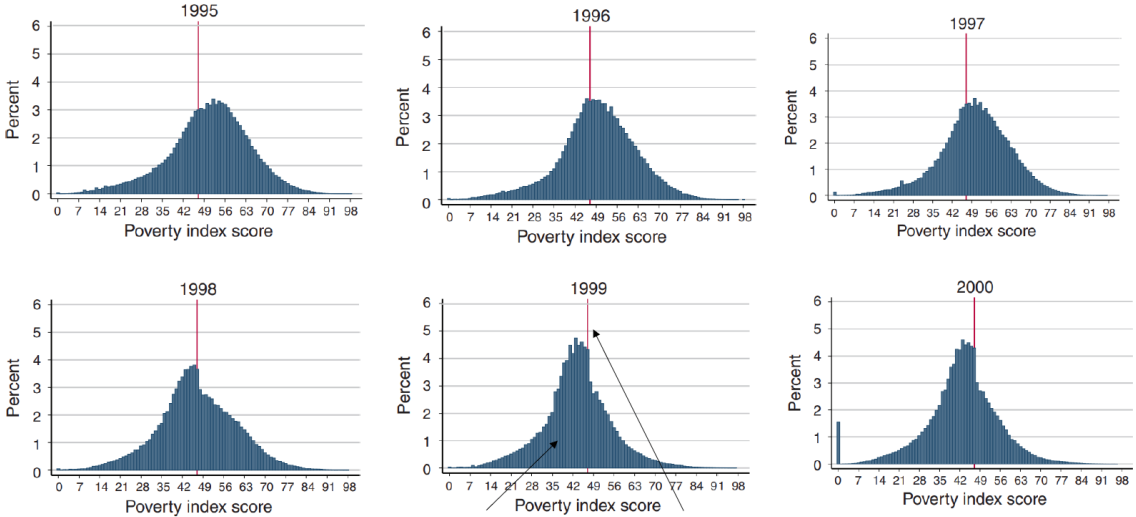
Strategic Behavior

How can we model distribution shift?



Algorithm for score, as a function of questions, made available July 1997

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

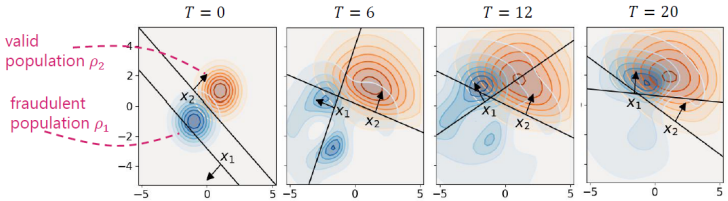


Can we model as a gradient flow? Issue: multiple species.

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

Banking and Loan Applicants

- population ρ_1 puts forward valid new account requests; population ρ_2 submits fraudulent request.
- (ρ_1, ρ_1) are distributions over income, customer age, number of months at a current residence, etc.
- Two banks (x_1, x_2) apply classifiers to separate the fraudulent and valid applicants.
- Applicants choose to affiliate with one of the two banks.
- Bank classifiers and applicants manipulate data over time.



Under what conditions is the system stable?

How does each bank perform?

How many fraudulent accounts are opened?

Need multispecies PDE analysis tools!

[2] Eliot Shekhtman and Sarah Dean. Strategic Usage in a Multi-Learner Setting. en. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

Overview

Goal: characterize long-time behavior for multispecies systems,
e.g. strategic agents interacting with learning algorithm, biological systems.

Issue: joint systems are not gradient flows; convexity arguments not applicable.

Idea: in game theory and optimization, monotonicity extends convexity/dissipativity to a *set* of operators.

Key question: can we develop a notion of monotonicity for general multispecies dynamics?

Three contributions:

1. Defining monotonicity for mass-preserving flow equations.
2. Showing new asymptotic results for multispecies systems.
3. Insights for applications.

Our Setting

We analyze dynamics

general mass-preserving flow: $$\partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0 \qquad \forall\, i \in [1, \dots, n]\,, \tag{1}$$

gradient flow (special case): $$\partial_t \rho_i = -\nabla_{\mathscr{W}_2, \rho_i} F_i[\rho] = \operatorname{div}_{x_i}\left(\rho_i \nabla_{x_i} \delta_{\rho_i} F_i[\rho]\right) \qquad \forall\, i \in [1, \dots, n]\,. \tag{2}$$

Our Setting

We analyze dynamics

$$\text{general mass-preserving flow:} \quad \partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0 \qquad \forall\, i \in [1, \dots, n]\,, \tag{1}$$

$$\text{gradient flow (special case):} \quad \partial_t \rho_i = -\nabla_{\mathscr{W}_2, \rho_i} F_i[\rho] = \operatorname{div}_{x_i}\left(\rho_i \nabla_{x_i} \delta_{\rho_i} F_i[\rho]\right) \qquad \forall\, i \in [1, \dots, n]\,. \tag{2}$$

Notation:

$$v[\rho](x) := \begin{bmatrix} v_1[\rho](x_1) \\ \vdots \\ v_n[\rho](x_n) \end{bmatrix}, \quad \rho := \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_n \end{bmatrix}, \quad \overline{\nabla \delta_\rho} F[\rho](x) := \begin{bmatrix} \nabla_{x_1} \delta_{\rho_1} F_1[\rho](x) \\ \vdots \\ \nabla_{x_n} \delta_{\rho_n} F_n[\rho](x) \end{bmatrix}$$

- Product space: $\overline{\mathcal{P}_2} = \mathcal{P}_2(\mathbb{R}^{d_1}) \times \cdots \times \mathcal{P}_2(\mathbb{R}^{d_n})$
- $\overline{W}(\rho, \mu)^2 = \sum_{i=1}^n \mathscr{W}_2(\rho_i, \mu_i)^2$ where $\mathscr{W}_2$ is Wasserstein-2 metric
- Admissible couplings: $\Gamma(\rho^{(0)}, \rho^{(1)}) \coloneqq \Gamma_1(\rho_1^{(0)}, \rho_1^{(1)}) \times \cdots \times \Gamma_n(\rho_n^{(0)}, \rho_n^{(1)})$
- Optimal couplings: $\Gamma_i^* = \left\{ \gamma_i \,:\, \gamma_i \in \operatorname{argmin}_{\hat{\gamma}_i \in \Gamma_i(\rho_i^{(0)}, \rho_i^{(1)})} \int \|x_i - y_i\|^2 \mathrm{d}\hat{\gamma}_i(x_i, y_i) \right\}$

Example: Energies with Nonlinear Diffusion

Consider the coupled gradient flow system

$$\partial_t \rho_1 = -\nabla_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$

with energies

$$F_1[\rho] = \iint f_1(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_1 \int U_{m_1}(\rho_1)\mathrm{d}x_1 + \frac{1}{2}\int W_1(x_1,\hat{x}_1)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_1(\hat{x}_1)\,,$$

$$F_2[\rho] = \iint f_2(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_2 \int U_{m_2}(\rho_2)\mathrm{d}x_2 + \frac{1}{2}\int W_2(x_2,\hat{x}_2)\mathrm{d}\rho_2(x_2)\mathrm{d}\rho_2(\hat{x}_2)\,.$$

Example: Energies with Nonlinear Diffusion

Consider the coupled gradient flow system

$$\partial_t \rho_1 = -\nabla_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$

with energies

$$F_1[\rho] = \iint f_1(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_1 \int U_{m_1}(\rho_1)\mathrm{d}x_1 + \frac{1}{2}\int W_1(x_1,\hat{x}_1)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_1(\hat{x}_1)\,,$$

$$F_2[\rho] = \iint f_2(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_2 \int U_{m_2}(\rho_2)\mathrm{d}x_2 + \frac{1}{2}\int W_2(x_2,\hat{x}_2)\mathrm{d}\rho_2(x_2)\mathrm{d}\rho_2(\hat{x}_2)\,.$$

What conditions on F_1 and F_2 result in contraction in $\overline{W}$?

Example: Energies with Nonlinear Diffusion

Consider the coupled gradient flow system

$$\partial_t \rho_1 = -\nabla_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$

with energies

$$F_1[\rho] = \iint f_1(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_1 \int U_{m_1}(\rho_1)\mathrm{d}x_1 + \frac{1}{2}\int W_1(x_1,\hat{x}_1)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_1(\hat{x}_1)\,,$$

$$F_2[\rho] = \iint f_2(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_2 \int U_{m_2}(\rho_2)\mathrm{d}x_2 + \frac{1}{2}\int W_2(x_2,\hat{x}_2)\mathrm{d}\rho_2(x_2)\mathrm{d}\rho_2(\hat{x}_2)\,.$$

What conditions on F_1 and F_2 result in contraction in $\overline{W}$?

General setting: n-species system, *not* a gradient flow.

Why monotonicity?

- **Optimization:** Monotonicity is a condition for existence of Nash equilibrium for multi-energy optimization problem:

solve system for ρ^*

$$\begin{cases} \rho_1^* = \operatorname{argmin}_{\rho_1 \in \mathcal{P}_2} F_1[\rho_1, \rho_{-1}^*] \\ \quad \vdots \\ \rho_n^* = \operatorname{argmin}_{\rho_n \in \mathcal{P}_2} F_n[\rho_n, \rho_{-n}^*] \end{cases} .$$

ρ_{-i} denotes $(\rho_j)_j$ for $j \neq i$

Why monotonicity?

- **Optimization:** Monotonicity is a condition for existence of Nash equilibrium for multi-energy optimization problem:

solve system for ρ^*

$$\begin{cases} \rho_1^* = \operatorname{argmin}_{\rho_1 \in \mathcal{P}_2} F_1[\rho_1, \rho_{-1}^*] \\ \quad \vdots \\ \rho_n^* = \operatorname{argmin}_{\rho_n \in \mathcal{P}_2} F_n[\rho_n, \rho_{-n}^*] \end{cases} .$$

ρ_{-i} denotes $(\rho_j)_j$ for $j \neq i$

- **Dynamical System:** Monotonicity is a generalization of convexity (dissipativity) for multi-energy systems, resulting in
  - contraction in joint metric;
  - energies do not necessarily decay.

Monotonicity

Definition (Finite-Dimensional Monotonicity)

Let $u_i \in C^1(\mathbb{R}^d; \mathbb{R}^{d_i})$, $d = \sum_{i=1}^n d_i$, $\lambda \in \mathbb{R}$. Then $(u_i)_i$ is λ-monotone if it satisfies

$$\sum_{i=1}^{n} \langle u_i(x) - u_i(x'),\, x_i - x_i' \rangle \geq \lambda \|x - x'\|^2 \quad \forall\, x, x' \in \mathbb{R}^d\,.$$

Monotonicity

Definition (Finite-Dimensional Monotonicity)

Let $u_i \in C^1(\mathbb{R}^d; \mathbb{R}^{d_i})$, $d = \sum_{i=1}^n d_i$, $\lambda \in \mathbb{R}$. Then $(u_i)_i$ is λ-monotone if it satisfies

$$\sum_{i=1}^{n} \left\langle u_i(x) - u_i(x'),\, x_i - x_i' \right\rangle \geq \lambda \|x - x'\|^2 \quad \forall\, x, x' \in \mathbb{R}^d\,.$$

Definition (Monotonicity [New])

If for any $\rho^{(0)}, \rho^{(1)} \in \mathcal{A} \subseteq \overline{\mathcal{P}_2}$ with $\mathcal{A}$ geodesically convex, $\lambda \in \mathbb{R}$,

$$-\sum_{i=1}^{n} \int \left\langle v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i),\, x_i - y_i \right\rangle \mathrm{d}\gamma(x,y) \geq \lambda \int \|x-y\|^2 \mathrm{d}\gamma(x,y) = \lambda \overline{W}(\rho^{(0)}, \rho^{(1)})^2\,, \qquad (3)$$

for all $\gamma \in \Gamma^*(\rho^{(0)}, \rho^{(1)})$, then the velocity fields $\{v_i : \mathcal{A} \to \mathbb{R}^{d_i}\}_{i=1}^n$ are λ-monotone in $\mathcal{A}$.

Monotonicity

Definition (Finite-Dimensional Monotonicity)

Let $u_i \in C^1(\mathbb{R}^d; \mathbb{R}^{d_i})$, $d = \sum_{i=1}^n d_i$, $\lambda \in \mathbb{R}$. Then $(u_i)_i$ is λ-monotone if it satisfies

$$\sum_{i=1}^n \left\langle u_i(x) - u_i(x'),\, x_i - x_i' \right\rangle \geq \lambda \|x - x'\|^2 \quad \forall\, x, x' \in \mathbb{R}^d\,.$$

Definition (Monotonicity [New])

If for any $\rho^{(0)}, \rho^{(1)} \in \mathcal{A} \subseteq \overline{\mathcal{P}_2}$ with $\mathcal{A}$ geodesically convex, $\lambda \in \mathbb{R}$,

$$-\sum_{i=1}^n \int \left\langle v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i),\, x_i - y_i \right\rangle \mathrm{d}\gamma(x,y) \geq \lambda \int \|x-y\|^2 \mathrm{d}\gamma(x,y) = \lambda \overline{W}(\rho^{(0)}, \rho^{(1)})^2\,, \qquad (3)$$

for all $\gamma \in \Gamma^*(\rho^{(0)}, \rho^{(1)})$, then the velocity fields $\{v_i : \mathcal{A} \to \mathbb{R}^{d_i}\}_{i=1}^n$ are λ-monotone in $\mathcal{A}$.

Corollary

The energy functionals $\{F_i : \mathcal{A} \to \mathbb{R}\}_{i=1}^n$ are λ-monotone in $\mathcal{A}$ if $v_i[\rho] = -\nabla_{x_i} \delta_{\rho_i} F_i[\rho]$ are λ-monotone:

$$\sum_{i=1}^n \int \left\langle \nabla_{x_i} \delta_{\rho_i} F_i[\rho^{(0)}](x_i) - \nabla_{x_i} \delta_{\rho_i} F_i[\rho^{(1)}](y_i),\, x_i - y_i \right\rangle \mathrm{d}\gamma_i(x_i, y_i) \geq \lambda \overline{W}(\rho^{(0)}, \rho^{(1)})^2\,.$$

Related Literature: Game Theory

- monotonicity emerges in *diagonally strictly concave games* in[3]
- results for measure-valued functionals new in game theory literature; previous results for Hilbert spaces[4]
- learning algorithms which minimize functions under the assumption of monotone cost structures[5]
- generalizes results in[6][7] beyond min-max setting.
- key examples of monotone games include bounded perturbations; we extend[8] to measures.

---

[3] J. B. Rosen. Existence and Uniqueness of Equilibrium Points for Concave N-Person Games. *Econometrica*, 1965.

[4] Haim Brezis. The Hille-Yosida Theorem. en. In Haim Brezis, editor, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*. Springer, New York, NY, 2011.

[5] Tatiana Tatarenko and Maryam Kamgarpour. Learning Nash Equilibria in Monotone Games. In *2019 IEEE 58th Conference on Decision and Control (CDC)*, 2019. ISSN: 2576-2370.

[6] 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.

[7] Lauren Conger, Franca Hoffmann, Eric Mazumdar, and Lillian J. Ratliff. Coupled Wasserstein Gradient Flows for Min-Max and Cooperative Games. 2024. arXiv:2411.07403 [math].

[8] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Related Literature: PDE Analysis

- joint system is not necessarily a gradient flow[9]
- monotonicity is a joint notion of displacement convexity, where a system is driven by multiple energies.[10][11]
- characterization of the equilibria of a two-species gradient flow system with nonlocal interaction[12]
- existence of solutions for multispecies systems with nonlocal interaction[13] or cross-diffusion[14]
- vector-valued optimal transport[15]

---

[9] Qiang Du, Kaizheng Wang, Edith Zhang, and Chenyang Zhong. A Particle Algorithm for Mean-Field Variational Inference. 2024. arXiv:2412.20385.

[10] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in wasserstein spaces. *Probability Theory and Related Fields*, 1, 2023.

[11] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces. 2023. arXiv:2305.05211.

[12] José Antonio Carrillo, Marco Di Francesco, Antonio Esposito, Simone Fagioli, and Markus Schmidtchen. Measure solutions to a system of continuity equations driven by Newtonian nonlocal interactions. en. *Discrete and Continuous Dynamical Systems*, 2020

[13] Marco Di Francesco and Simone Fagioli. A nonlocal swarm model for predator-prey interactions. en. *Mathematical Models and Methods in Applied Sciences*, 2016.

[14] Alpár R. Mészáros and Guy Parker. On a cross-diffusion system with independent drifts and no self-diffusion: the existence of totally mixed solutions. 2025.

[15] Katy Craig, Nicolás García Trillos, and Dorde Nikolić. Vector valued optimal transport: from dynamic to static formulations. 2025. arXiv:2505.03670.

Admissible vs Optimal Plans

Theorem (Admissible Couplings)

If for any $\rho^{(0)}, \rho^{(1)} \in \overline{\mathcal{P}_2}$ *and* $\gamma^* \in \Gamma^*(\rho^{(0)}, \rho^{(1)})$ *it holds*

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i^*(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma^*(x, y)\,,$$

then for any $\gamma \in \Gamma(\rho^{(0)}, \rho^{(1)})$,

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma(x, y)\,.$$

[16] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in wasserstein spaces. *Probability Theory and Related Fields*, 1, 2023.

Admissible vs Optimal Plans

Theorem (Admissible Couplings)

If for any $\rho^{(0)}, \rho^{(1)} \in \overline{\mathcal{P}_2}$ *and* $\gamma^* \in \Gamma^*(\rho^{(0)}, \rho^{(1)})$ *it holds*

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i^*(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma^*(x, y)\,,$$

then for any $\gamma \in \Gamma(\rho^{(0)}, \rho^{(1)})$,

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma(x, y)\,.$$

Proof follows from[16]. Monotonicity inequality with admissible plans (not necessarily optimal) does not hold for general $\mathcal{A} \subset \overline{\mathcal{P}_2}$! For example, fails for $v[\rho] = -\nabla \log \rho$, $\mathcal{A} = L^1 \subset \overline{\mathcal{P}_2}$.

[16] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in wasserstein spaces. *Probability Theory and Related Fields*, 1, 2023.

Admissible vs Optimal Plans

Theorem (Admissible Couplings)

If for any $\rho^{(0)}, \rho^{(1)} \in \overline{\mathcal{P}_2}$ *and* $\gamma^* \in \Gamma^*(\rho^{(0)}, \rho^{(1)})$ *it holds*

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i^*(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma^*(x, y)\,,$$

then for any $\gamma \in \Gamma(\rho^{(0)}, \rho^{(1)})$,

$$-\sum_{i=1}^{n} \int \Big\langle x_i - y_i,\, v_i[\rho^{(0)}](x_i) - v_i[\rho^{(1)}](y_i) \Big\rangle \, \mathrm{d}\gamma_i(x_i, y_i) \geq \lambda \int \|x - y\|^2 \mathrm{d}\gamma(x, y)\,.$$

Proof follows from[16]. Monotonicity inequality with admissible plans (not necessarily optimal) does not hold for general $\mathcal{A} \subset \overline{\mathcal{P}_2}$! For example, fails for $v[\rho] = -\nabla \log \rho$, $\mathcal{A} = L^1 \subset \overline{\mathcal{P}_2}$. Need Diracs in $\mathcal{A}$ for theorem to hold.

[16] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. Dissipative probability vector fields and generation of evolution semigroups in wasserstein spaces. *Probability Theory and Related Fields*, 1, 2023.

Assumptions for main results

$$\partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0\,, \qquad i \in [1, \dots, n] \tag{1}$$

Assume

- **Absolutely continuous curves:** The solution $\rho(t)$ is absolutely continuous.
- **Monotonicity:** The velocity field v is λ-monotone in some set $\mathcal{A} \subseteq \overline{\mathcal{P}_2}$ with $\lambda > 0$.
- **No escape:** The solution satisfies $\rho(t) \in \mathcal{A} \subseteq \overline{\mathcal{P}_2}$ for all $t \geq 0$.
- **Lower semicontinuity:** For any fixed $\rho \in \mathcal{A}$, the functional $F_i[\cdot, \rho_{-i}]$ is l.s.c wrt the weak topology over the domain $\mathcal{A}_i$.

Long-Time Behavior

$$\partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0\,, \qquad i \in [1, \dots, n] \tag{1}$$

Theorem (Main Result)

Under these assumptions,

(a) *there exists a unique steady state* ρ^∞ *of* (1) *and if* $\mathcal{A}$ *is compact in* $\overline{W}$*, then* $\rho^\infty \in \mathcal{A}$*;*

(b) *if* $\rho^\infty \in \mathcal{A}$*, then*

$$\overline{W}(\rho(t), \rho^\infty) \le e^{-\lambda t} \overline{W}(\rho(0), \rho^\infty)\,.$$

(c) *If* $v_i = -\nabla_{x_i} \delta_{\rho_i} F_i[\rho]$*,* F_i *lsc, and* $\rho^\infty \in \mathcal{A}$*, then* ρ^∞ *is the unique Nash equilibrium of* F *in* $\mathcal{A}$*.*

Long-Time Behavior

$$\partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0\,, \qquad i \in [1, \dots, n] \tag{1}$$

Theorem (Main Result)

Under these assumptions,

(a) *there exists a unique steady state ρ^∞ of (1) and if $\mathcal{A}$ is compact in $\overline{W}$, then $\rho^\infty \in \mathcal{A}$;*

(b) *if $\rho^\infty \in \mathcal{A}$, then*

$$\overline{W}(\rho(t), \rho^\infty) \le e^{-\lambda t} \overline{W}(\rho(0), \rho^\infty)\,.$$

(c) *If $v_i = -\nabla_{x_i} \delta_{\rho_i} F_i[\rho]$, F_i lsc, and $\rho^\infty \in \mathcal{A}$, then ρ^∞ is the unique Nash equilibrium of F in $\mathcal{A}$.*

Monotonicity with $\lambda > 0$ implies existence of a unique steady state and exponential convergence to it.

Long-Time Behavior

$$\partial_t \rho_i + \operatorname{div}\left(\rho_i v_i[\rho]\right) = 0\,, \qquad i \in [1, \dots, n] \tag{1}$$

Theorem (Main Result)

Under these assumptions,

(a) *there exists a unique steady state* ρ^∞ *of* (1) *and if* $\mathcal{A}$ *is compact in* $\overline{W}$*, then* $\rho^\infty \in \mathcal{A}$*;*

(b) *if* $\rho^\infty \in \mathcal{A}$*, then*

$$\overline{W}(\rho(t), \rho^\infty) \leq e^{-\lambda t} \overline{W}(\rho(0), \rho^\infty)\,.$$

(c) *If* $v_i = -\nabla_{x_i} \delta_{\rho_i} F_i[\rho]$*,* F_i *lsc, and* $\rho^\infty \in \mathcal{A}$*, then* ρ^∞ *is the unique Nash equilibrium of* F *in* $\mathcal{A}$*.*

Monotonicity with $\lambda > 0$ implies existence of a unique steady state and exponential convergence to it.

$\Rightarrow$ **Other results:** uniform bounds on second moments, exponential convergence of $\sum_{i=1}^n \int \|v_i[\rho](x_i)\|^2 \mathrm{d}\rho_i(x_i)$, second order monotonicity conditions.

Properties in Gradient Flow Setting

(C1) λ-monotonicity implies λ-convexity: if F is λ-monotone in $\mathcal{A}$, then each $F_i[\cdot, \rho_{-i}]$ is λ-displacement convex in $\mathcal{A}_i$ for all $\rho_{-i} \in \mathcal{A}_{-i}$.

[17] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Properties in Gradient Flow Setting

(C1) λ-monotonicity implies λ-convexity: if F is λ-monotone in $\mathcal{A}$, then each $F_i[\cdot, \rho_{-i}]$ is λ-displacement convex in $\mathcal{A}_i$ for all $\rho_{-i} \in \mathcal{A}_{-i}$. Converse false in general. Equivalent if $F_i[\rho_i, \rho_{-i}] = F_i[\rho_i]$.

[17] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Properties in Gradient Flow Setting

(C1) λ-monotonicity implies λ-convexity: if F is λ-monotone in $\mathcal{A}$, then each $F_i[\cdot, \rho_{-i}]$ is λ-displacement convex in $\mathcal{A}_i$ for all $\rho_{-i} \in \mathcal{A}_{-i}$. Converse false in general. Equivalent if $F_i[\rho_i, \rho_{-i}] = F_i[\rho_i]$.

(C2) Additivity of monotonicity: if F is λ-monotone in $\mathcal{A}$ and E is λ-monotone in $\mathcal{B}$, then $F + E$ is λ-monotone in $\mathcal{A} \cap \mathcal{B}$.

[17] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Properties in Gradient Flow Setting

(C1) λ-monotonicity implies λ-convexity: if F is λ-monotone in $\mathcal{A}$, then each $F_i[\cdot, \rho_{-i}]$ is λ-displacement convex in $\mathcal{A}_i$ for all $\rho_{-i} \in \mathcal{A}_{-i}$. Converse false in general. Equivalent if $F_i[\rho_i, \rho_{-i}] = F_i[\rho_i]$.

(C2) Additivity of monotonicity: if F is λ-monotone in $\mathcal{A}$ and E is λ-monotone in $\mathcal{B}$, then $F + E$ is λ-monotone in $\mathcal{A} \cap \mathcal{B}$.

(C3) Monotonicity of cross-interaction kernels: $\mathcal{W}_{ij} = \int W_{ij}(x_i, x_j) \mathrm{d}\rho_i(x_i) \mathrm{d}\rho_j(x_j)$, $F_i = \sum_{j \neq i} \mathcal{W}_{ij}$. Let α_{ij} be the Lipschitz coefficient of $\nabla_{x_i} W_{ij}(x_i, \cdot)$ and $\sum_{j,j \neq i} \nabla^2_{x_i} W_{ij}(x_i, x_j) \succeq c_i \, \mathrm{Id}_{d_i}$. Define

$$\Lambda := \begin{bmatrix} c_1 & -\alpha_{12} & \dots & -\alpha_{1n} \\ -\alpha_{21} & c_2 & & -\alpha_{2n} \\ \vdots & & \ddots & \vdots \\ -\alpha_{n1} & -\alpha_{n2} & \dots & c_n \end{bmatrix}.$$

Then F is λ-monotone in $\mathcal{A}$, where $\frac{1}{2}(\Lambda + \Lambda^\top) \succeq \lambda \, \mathrm{Id}_d$.

---

[17] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Properties in Gradient Flow Setting

(C1) λ-monotonicity implies λ-convexity: if F is λ-monotone in $\mathcal{A}$, then each $F_i[\cdot, \rho_{-i}]$ is λ-displacement convex in $\mathcal{A}_i$ for all $\rho_{-i} \in \mathcal{A}_{-i}$. Converse false in general. Equivalent if $F_i[\rho_i, \rho_{-i}] = F_i[\rho_i]$.

(C2) Additivity of monotonicity: if F is λ-monotone in $\mathcal{A}$ and E is λ-monotone in $\mathcal{B}$, then $F + E$ is λ-monotone in $\mathcal{A} \cap \mathcal{B}$.

(C3) Monotonicity of cross-interaction kernels: $\mathcal{W}_{ij} = \int W_{ij}(x_i, x_j) \mathrm{d}\rho_i(x_i) \mathrm{d}\rho_j(x_j)$, $F_i = \sum_{j \neq i} \mathcal{W}_{ij}$. Let α_{ij} be the Lipschitz coefficient of $\nabla_{x_i} W_{ij}(x_i, \cdot)$ and $\sum_{j,j \neq i} \nabla^2_{x_i} W_{ij}(x_i, x_j) \succeq c_i \, \mathrm{Id}_{d_i}$. Define

$$\Lambda := \begin{bmatrix} c_1 & -\alpha_{12} & \dots & -\alpha_{1n} \\ -\alpha_{21} & c_2 & & -\alpha_{2n} \\ \vdots & & \ddots & \vdots \\ -\alpha_{n1} & -\alpha_{n2} & \dots & c_n \end{bmatrix}.$$

Then F is λ-monotone in $\mathcal{A}$, where $\frac{1}{2}(\Lambda + \Lambda^\top) \succeq \lambda \, \mathrm{Id}_d$.

If W_{ij} are convex enough in i, and effect of other species is bounded, $\Rightarrow$ monotonicity.[17]

[17] Alexander Davydov, Saber Jafarpour, and Francesco Bullo. Non-euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 2022.

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot \mathscr{W}_{2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot \mathscr{W}_{2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1, x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

For $\eta \in \mathcal{P}(\mathbb{R}^k)$, the nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta \log \eta & \text{for } m = 1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m > 1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases}$$

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

For $\eta \in \mathcal{P}(\mathbb{R}^k)$, the nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta\log\eta & \text{for } m=1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m>1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases}$$

- Use finite-infinite dimensional monotonicity equivalence to show that if $[f_1, f_2]$ is λ-monotone, then

$$\left[\iint f_1(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2), \iint f_2(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2)\right]$$

  is λ-monotone in $\overline{\mathcal{P}_2}$.

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

For $\eta \in \mathcal{P}(\mathbb{R}^k)$, the nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta\log\eta & \text{for } m=1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m>1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases}$$

- Use finite-infinite dimensional monotonicity equivalence to show that if $[f_1, f_2]$ is λ-monotone, then

$$\left[\iint f_1(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2), \iint f_2(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2)\right]$$

  is λ-monotone in $\overline{\mathcal{P}_2}$.
- Use that $\int U_{m_i}(\rho_i)\mathrm{d}x_i$ is 0-convex in L^1 + (C1,C2) (convexity, addivity) to conclude that $F = [F_1, F_2]$ is λ-monotone in $\mathcal{A} = \overline{\mathcal{P}_2} \cap \overline{L^1}$.

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot \mathscr{W}_{2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot \mathscr{W}_{2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1, x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

The nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta \log \eta & \text{for } m = 1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m > 1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{7}$$

- Assuming solution $\rho(t)$ is absolutely continuous and $\rho(t) \in \mathcal{A}$, obtain uniform bound on second moments.

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot_{\mathscr{W}_2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot_{\mathscr{W}_2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1,x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

The nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta\log\eta & \text{for } m=1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m>1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{7}$$

- Assuming solution $\rho(t)$ is absolutely continuous and $\rho(t) \in \mathcal{A}$, obtain uniform bound on second moments.
- Apply main theorem for existence of and convergence to ρ^∞,

$$\overline{W}(\rho(t),\rho^\infty) \le e^{-\lambda t}\overline{W}(\rho(0),\rho^\infty)\,.$$

Example: Monotone Energy with Nonlinear Diffusion

Consider the coupled gradient flow system parameterized by F,

$$\partial_t \rho_1 = -\nabla \cdot \mathscr{W}_{2,\rho_1} F_1[\rho]\,, \qquad \partial_t \rho_2 = -\nabla \cdot \mathscr{W}_{2,\rho_2} F_2[\rho]\,,$$
$$F_i[\rho] = \iint f_i(x_1, x_2)\mathrm{d}\rho_1(x_1)\mathrm{d}\rho_2(x_2) + \alpha_i \int U_{m_i}(\rho_i)\mathrm{d}x_i\,.$$

The nonlinear diffusion terms are given by

$$U_m(\eta) = \begin{cases} \eta \log \eta & \text{for } m = 1\,,\ \eta \ll \mathcal{L}^k\,, \\ \frac{\eta^m}{m-1} & \text{for } m > 1\,,\ \eta \ll \mathcal{L}^k\,, \\ +\infty & \text{otherwise}\,. \end{cases} \tag{7}$$

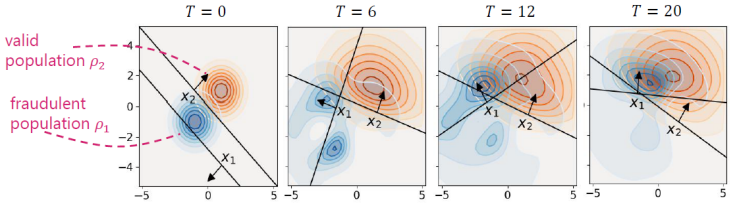
- Assuming solution $\rho(t)$ is absolutely continuous and $\rho(t) \in \mathcal{A}$, obtain uniform bound on second moments.
- Apply main theorem for existence of and convergence to ρ^∞,

$$\overline{W}(\rho(t), \rho^\infty) \leq e^{-\lambda t}\overline{W}(\rho(0), \rho^\infty)\,.$$

- Show lower semincontinuity holds due to monotonicity properties of $[f_1, f_2] \in C^2 \times C^2$, then obtain ρ^∞ is Nash equilibrium of $F = [F_1, F_2]$.

Application 1: Four-player bank game model

- population ρ_1 puts forward valid new account requests; population ρ_2 submits fraudulent request.
- (ρ_1, ρ_1) are distributions over income, customer age, number of months at a current residence, etc.
- Two banks (x_1, x_2) apply classifiers to separate the fraudulent and valid applicants.
- Applicants choose to affiliate with one of the two banks.
- Bank classifiers and applicants manipulate data over time.



Under what conditions is the system stable?

How does each bank perform?

How many fraudulent accounts are opened?

$\Longrightarrow$ Bank account fraud problem investigated in[18].

[18] Eliot Shekhtman and Sarah Dean. Strategic Usage in a Multi-Learner Setting. en. In *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*. PMLR, 2024.

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

applicants: $$F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

banks: $$F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

$$\text{applicants:} \quad F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

$$\text{banks:} \quad F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

- $a_{ij} \in [0, 1]$ is proportion of population i which purchases from service provider j

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

$$\text{applicants:} \quad F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

$$\text{banks:} \quad F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

- $a_{ij} \in [0, 1]$ is proportion of population i which purchases from service provider j
- applicant utility for population i with bank j is $u_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

$$\text{applicants:} \quad F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

$$\text{banks:} \quad F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

- $a_{ij} \in [0, 1]$ is proportion of population i which purchases from service provider j
- applicant utility for population i with bank j is $u_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$
- bank j classifier loss for interacting with population i is $l_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

$$\text{applicants:} \quad F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

$$\text{banks:} \quad F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

- $a_{ij} \in [0, 1]$ is proportion of population i which purchases from service provider j
- applicant utility for population i with bank j is $u_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$
- bank j classifier loss for interacting with population i is $l_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$
- log concave reference measures $\tilde{\rho}_i$, penalty terms $\alpha_i \mathrm{KL} \rho_i \tilde{\rho}_i$ and $\beta_j \|h_j\|^2$

Application 1: Four-player bank game model

Gradient flow dynamics: $\partial_t \rho_i = -\nabla_{\mathscr{W}_2,\rho_i} F_i[\rho, h]$, $\frac{\mathrm{d}}{\mathrm{d}t} h_j = -\nabla_{h_j} F_{2+j}[\rho, h]$ where

$$\text{applicants:} \quad F_i[\rho, h] = -\sum_{j=1}^{2} \int a_{ij} u_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \sum_{k=1}^{2} \int \rho_i W_{ik} * \rho_k + \alpha_i KL(\rho_i | \tilde{\rho}_i)\,,$$

$$\text{banks:} \quad F_{2+j}[\rho, h] = \sum_{i=1}^{2} \int a_{ij} l_{ij}(x_i, h_j) \mathrm{d}\rho_i(x_i) + \beta_j \|h_j\|^2\,.$$

- $a_{ij} \in [0, 1]$ is proportion of population i which purchases from service provider j
- applicant utility for population i with bank j is $u_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$
- bank j classifier loss for interacting with population i is $l_{ij} : \mathbb{R}^{d_\rho} \times \mathbb{R}^{d_h} \to \mathbb{R}$
- log concave reference measures $\tilde{\rho}_i$, penalty terms $\alpha_i \mathrm{KL} \rho_i \tilde{\rho}_i$ and $\beta_j \|h_j\|^2$

Affiliation dynamics: applicants choose the bank with more desirable classification

$$a_{ij}(t) \coloneqq \frac{\exp[\eta \int u_{ij}(x_i, h_j(t)) \mathrm{d}\rho_i(t, x_i)]}{\sum_{l=1}^{M} \exp[\eta \int u_{il}(x_i, h_l(t)) \mathrm{d}\rho_i(t, x_i)]}\,.$$

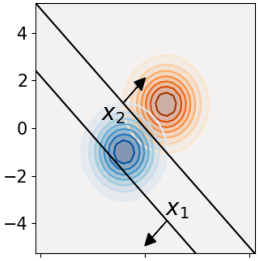
Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

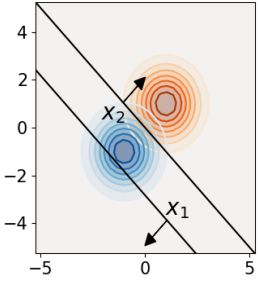
attractive-repulsive kernel W_{22}, fraudulent ρ_2

Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$



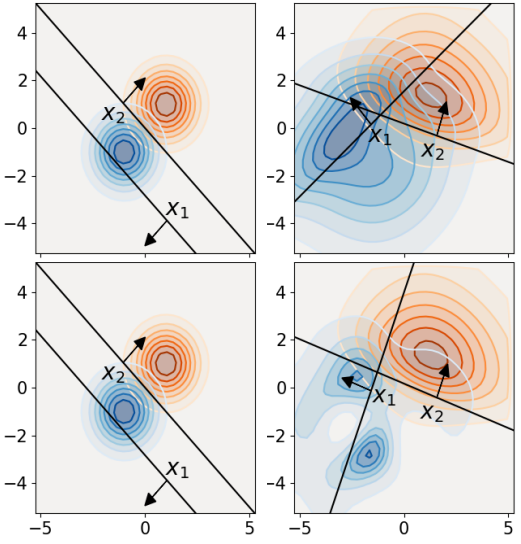
attractive-repulsive kernel W_{22}, fraudulent ρ_2



Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

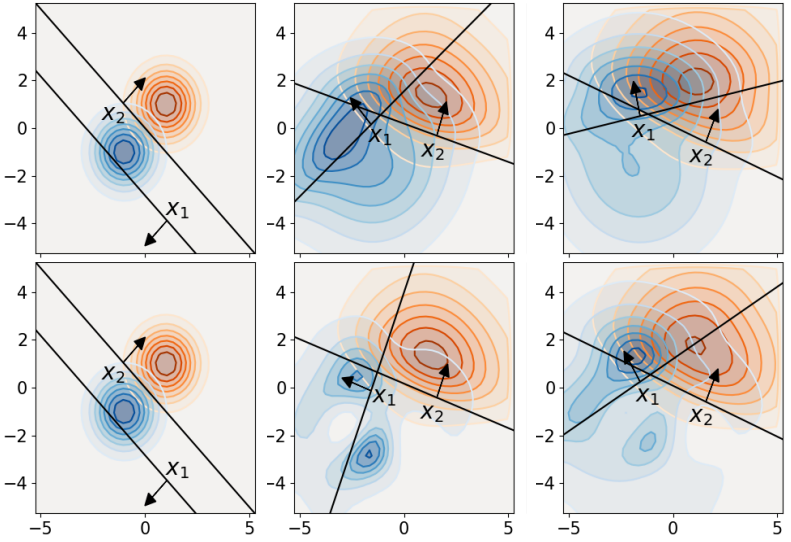
attractive-repulsive kernel W_{22}, fraudulent ρ_2



Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

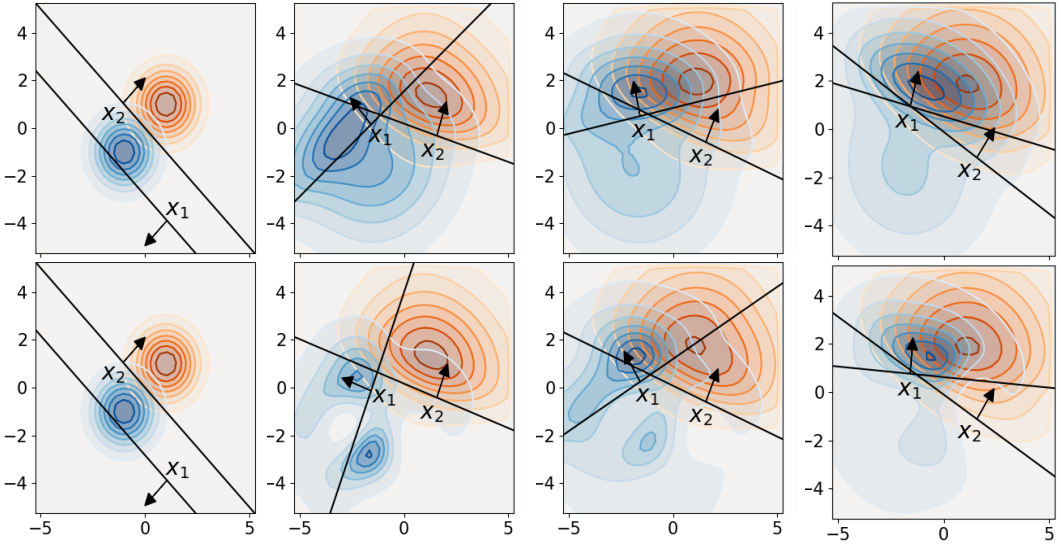
attractive-repulsive kernel W_{22}, fraudulent ρ_2



Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

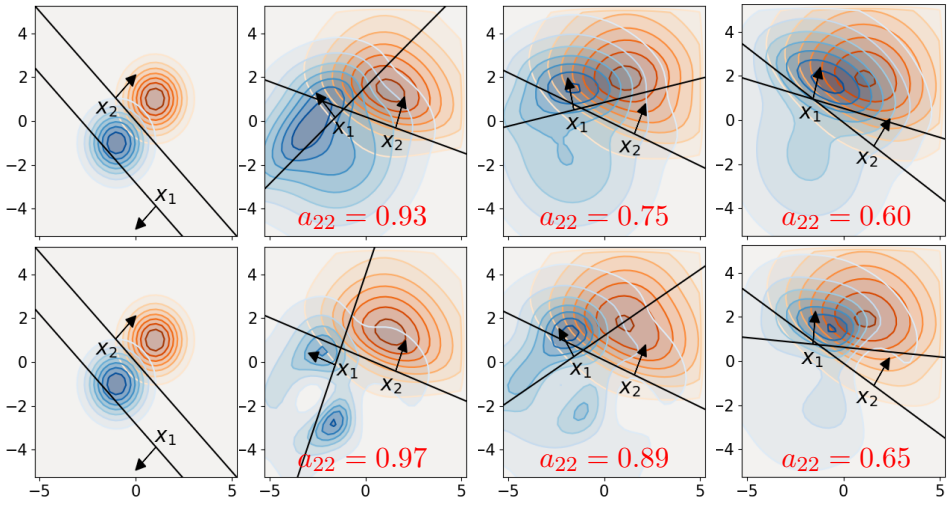
attractive-repulsive kernel W_{22}, fraudulent ρ_2



Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

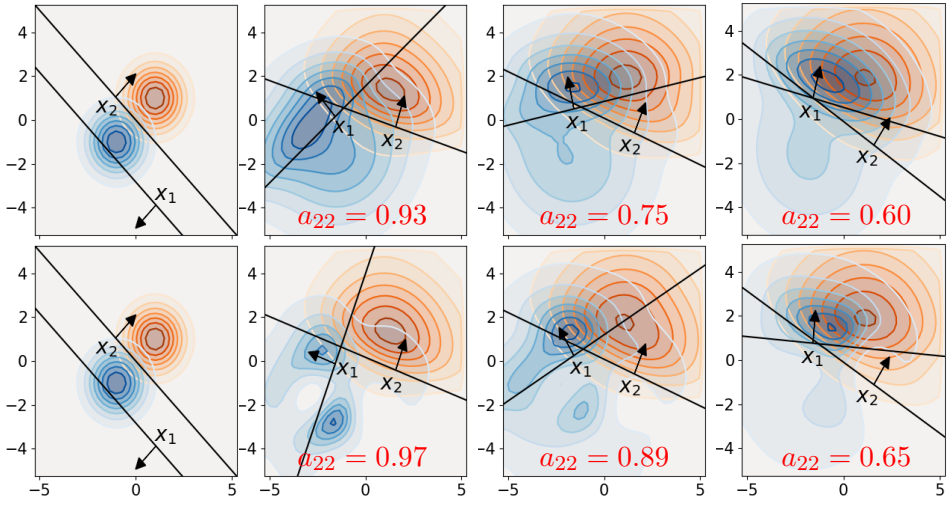
attractive-repulsive kernel W_{22}, fraudulent ρ_2



Application 1: Four-player bank game simulation

no interaction kernels $W_{ik} = 0$

attractive-repulsive kernel W_{22}, fraudulent ρ_2



Nonlocal interaction hurts bank 2; more fraudulent customers.

Application 2: Optimal Transport[19]

The Kantorovich formulation of the optimal transport problem, with cost $c : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}_+$,

$$\inf_{\rho \in \Gamma(\mu,\nu)} \iint c(x,y) \mathrm{d}\rho(x,y)\,. \tag{OT}$$

[19] Lauren Conger, Franca Hoffmann, Ricardo Baptista, and Eric Mazumdar. Computing Optimal Transport Plans via Min-Max Gradient Flows. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, 2025.

Application 2: Optimal Transport[19]

The Kantorovich formulation of the optimal transport problem, with cost $c : \mathbb{R}^{d_x} \times \mathbb{R}^{d_y} \to \mathbb{R}_+$,

$$\inf_{\rho \in \Gamma(\mu,\nu)} \iint c(x,y) \mathrm{d}\rho(x,y)\,. \tag{OT}$$

Equivalent to solving (monotone)

$$\sup_{\Lambda \in \mathbb{R}} \inf_{\rho \in \mathcal{P}_2} E(\rho, \Lambda; \mu, \nu)\,, \tag{ZS}$$

where E is the energy functional

$$E(\rho, \Lambda; \mu, \nu) = \iint c(x,y)\, \mathrm{d}\rho(x,y) + \Lambda \mathrm{KL}\rho_1\mu + \Lambda \mathrm{KL}\rho_2\nu\,,$$

with marginals $\rho_1 := \int \rho(x,y)\mathrm{d}y$ and $\rho_2 := \int \rho(x,y)\mathrm{d}x$ of ρ and $\mathrm{KL}\cdot\cdot$ is the KL divergence.

---

[19] Lauren Conger, Franca Hoffmann, Ricardo Baptista, and Eric Mazumdar. Computing Optimal Transport Plans via Min-Max Gradient Flows. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, 2025.

Application 2: Optimal Transport

To approximate (OT) numerically, we will solve (ZS) using a min-max gradient flow with respect to the metric $\mathcal{W}_2$,

$$\partial_t \rho(t) = -\frac{1}{\beta} \nabla_{\mathcal{W}_2,\rho} E(\rho(t), \Lambda(t); \mu, \nu) \quad \text{(minimize } E\text{)}\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t) = \partial_\Lambda E(\rho(t), \Lambda(t); \mu, \nu) \qquad \text{(maximize } E\text{)}\,.$$

Application 2: Optimal Transport

To approximate (OT) numerically, we will solve (ZS) using a min-max gradient flow with respect to the metric $\mathcal{W}_2$,

$$\partial_t \rho(t) = -\frac{1}{\beta} \nabla_{\mathscr{W}_2,\rho} E(\rho(t), \Lambda(t); \mu, \nu) \quad \text{(minimize } E\text{)}\,,$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t) = \partial_\Lambda E(\rho(t), \Lambda(t); \mu, \nu) \qquad \text{(maximize } E\text{)}\,.$$

Since E is not strongly convex in Λ, we analyze $\beta \to \infty$ (rate will depend on convexity with respect to ρ):

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t) = \partial_\Lambda E(b[\Lambda], \Lambda; \mu, \nu)$$
$$b[\Lambda] \coloneqq \operatorname*{argmin}_{\rho \in \mathcal{P}_2} E(\rho, \Lambda; \mu, \nu)\,.$$

Application 2: Optimal Transport

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t) = \partial_\Lambda E(b[\Lambda], \Lambda; \mu, \nu)$$
$$b[\Lambda] \coloneqq \operatorname*{argmin}_{\rho \in \mathcal{P}_2} E(\rho, \Lambda; \mu, \nu)\,.$$

Theorem (Regularization growth condition)

Under appropriate assumptions, the regularization parameter Λ satisfies

$$\Lambda(t) \leq \sqrt{2(c^* t + c_0)} \quad \forall\, t \geq 0\,,$$

where $c^* < \infty$ is the optimal value of (OT) *and $c_0 > 0$ is a constant dependent on $\Lambda(0) > 0$.*

Application 2: Optimal Transport

$$\frac{\mathrm{d}}{\mathrm{d}t}\Lambda(t) = \partial_\Lambda E(b[\Lambda], \Lambda; \mu, \nu)$$
$$b[\Lambda] \coloneqq \operatorname*{argmin}_{\rho \in \mathcal{P}_2} E(\rho, \Lambda; \mu, \nu)\,.$$

Theorem (Regularization growth condition)

Under appropriate assumptions, the regularization parameter Λ satisfies

$$\Lambda(t) \leq \sqrt{2(c^* t + c_0)} \quad \forall\, t \geq 0\,,$$

where $c^* < \infty$ is the optimal value of (OT) *and $c_0 > 0$ is a constant dependent on $\Lambda(0) > 0$.*

For gradient descent approaches to OT, algorithms with $\Lambda \to \infty$ should not update Λ faster than $O(\sqrt{t})$.

Application 2: Optimal Transport

Particle method for solving min-max dynamics proposed in[20] adapted from[21]. Let $(X_i, Y_i) \sim \rho$; divide the points $(X_j, Y_j)_{j=1}^{2N}$ into two sets and initialize:

$$X_i^{(1)}(0) \sim \mu\,, \qquad Y_i^{(1)}(0) = X_i^{(1)}(0) + \xi_i\,, \qquad \xi_i \sim \mathcal{N}(0, \eta)\,, \qquad \eta \text{ small}$$
$$Y_i^{(2)}(0) \sim \nu\,, \qquad X_i^{(2)}(0) = Y_i^{(2)}(0) + \xi_i$$

[20] Lauren Conger, Franca Hoffmann, Ricardo Baptista, and Eric Mazumdar. Computing Optimal Transport Plans via Min-Max Gradient Flows. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, 2025.

[21] Shu Liu, Haodong Sun, and Hongyuan Zha. Approximating the Optimal Transport Plan via Particle-Evolving Method. 2021. arXiv:2105.06088.

Application 2: Optimal Transport

Particle method for solving min-max dynamics proposed in[20] adapted from[21]. Let $(X_i, Y_i) \sim \rho$; divide the points $(X_j, Y_j)_{j=1}^{2N}$ into two sets and initialize:

$$X_i^{(1)}(0) \sim \mu\,, \qquad Y_i^{(1)}(0) = X_i^{(1)}(0) + \xi_i\,, \qquad \xi_i \sim \mathcal{N}(0, \eta)\,, \qquad \eta \text{ small}$$
$$Y_i^{(2)}(0) \sim \nu\,, \qquad X_i^{(2)}(0) = Y_i^{(2)}(0) + \xi_i$$

Cost starts near zero since $X_i \approx Y_i$, but marginal constraints not satisfied. Flow half of the particles to satisfy constraints, increase constraint penalty Λ

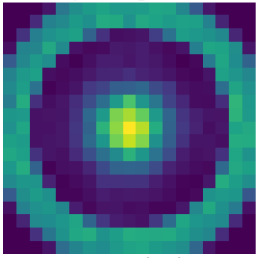
$$\dot{X}_i^{(1)}(t) = 0\,, \qquad \dot{Y}_i^{(1)}(t) = -\frac{1}{\beta}\nabla_y c(X_i^{(1)}, Y_i^{(1)}(t)) + \frac{1}{\beta}\Lambda(t)\nabla_y \log \frac{\rho_2(t, Y_i^{(1)}(t))}{\nu(Y_i^{(1)}(t))}\,,$$
$$\dot{Y}_i^{(2)}(t) = 0, \qquad \dot{X}_i^{(2)}(t) = -\frac{1}{\beta}\nabla_x c(X_i^{(2)}(t), Y_i^{(2)}) + \frac{1}{\beta}\Lambda(t)\nabla_x \log \frac{\rho_1(t, X_i^{(2)}(t))}{\mu(X_i^{(2)}(t))}\,,$$
$$\dot{\Lambda}(t) = \mathrm{KL}\rho_1(t)\mu + \mathrm{KL}\rho_2(t)\nu\,.$$

[20] Lauren Conger, Franca Hoffmann, Ricardo Baptista, and Eric Mazumdar. Computing Optimal Transport Plans via Min-Max Gradient Flows. In *2025 IEEE 64th Conference on Decision and Control (CDC)*, 2025.

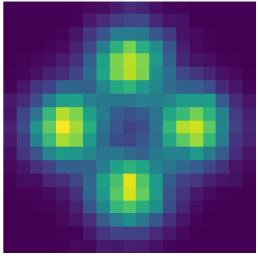
[21] Shu Liu, Haodong Sun, and Hongyuan Zha. Approximating the Optimal Transport Plan via Particle-Evolving Method. 2021. arXiv:2105.06088.

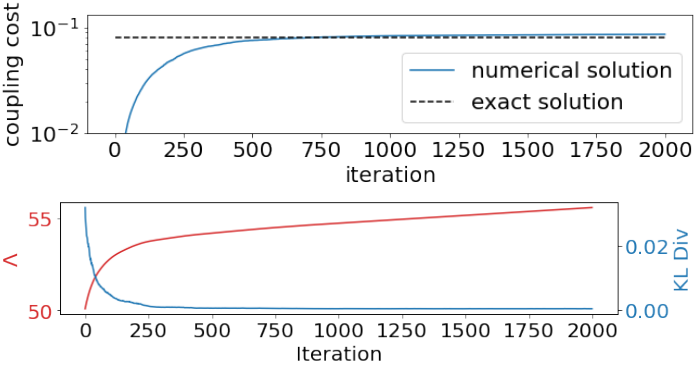
Application 2: Optimal Transport

μ marginal



ν marginal





Coupling cost increases as constraint penalty decreases; Λ increases.

Relation to team mean field games

For a single species, e.g. model for crowd density with intruder passing through, parameterized by $F : \mathcal{P}_2 \to \mathbb{R}$:

$$\inf_{v,\rho} \left\{ \int_0^T \left(\int \frac{e^{-t/\varepsilon}}{2} \|v(t,x)\|^2 \mathrm{d}\rho(t,x) + \frac{e^{-t/\varepsilon}}{\varepsilon} F[\rho(t)] \right) \mathrm{d}t + F[\rho(T)] e^{-T/\varepsilon} \right\},$$
$$\text{such that } \partial_t \rho(t) + \nabla \cdot (v(t)\rho(t)) = 0$$

The optimal (ρ, v) are perturbation of a gradient flow[22]

$$\partial_t \rho = \operatorname{div}\left(\rho \nabla_x \delta_\rho F[\rho]\right) + \varepsilon \operatorname{div}\left(\rho \nabla_x (\partial_t U - \frac{1}{2}|\nabla U|^2)\right)$$
$$-\varepsilon \partial_t U + U + \frac{\varepsilon}{2}|\nabla_x U|^2 = \delta_\rho F[\rho]\,.$$

In the team setting (multispecies), each team is optimizing cost parameterized by $F_i : \overline{\mathcal{P}_2} \to \mathbb{R}$:

$$\partial_t \rho_i = \operatorname{div}_{x_i}\left(\rho_i \nabla_{x_i} \delta_{\rho_i} F_i[\rho]\right) + \varepsilon \operatorname{div}_{x_i}\left(\rho_i \nabla_{x_i} (\partial_t U_i - \frac{1}{2}|\nabla U_i|^2)\right)$$
$$-\varepsilon \partial_t U_i + U_i + \frac{\varepsilon}{2}|\nabla_{x_i} U_i|^2 = \delta_{\rho_i} F_i[\rho]\,.$$

[22] Benjamin J. Zhang and Markos A. Katsoulakis. A mean-field games laboratory for generative modeling. 2023. arXiv:2304.13534 [cs, stat].

Degenerately monotone game

For $\mathcal{H}(\rho_i) = \int \log \rho_i(x_i) \mathrm{d}\rho_i(x_i)$, and $a, b, c > 0$, let

$$F[\rho_1, \rho_2] = \frac{a}{2} \int \|x_1\|^4 \mathrm{d}\rho_1(x_1) - \frac{a}{2} \int \|x_2\|^4 \mathrm{d}\rho_2(x_2) + \frac{b}{2} \iint x_1^\top x_2 \mathrm{d}\rho_1(x_1) \mathrm{d}\rho_2(x_2) + c(\mathcal{H}(\rho_1) - \mathcal{H}(\rho_2)) \,,$$

and consider dynamics

$$\partial_t \rho_1 = -\nabla \cdot \mathscr{W}_{2,\rho_1} F[\rho_1, \rho_2] \qquad \partial_t \rho_2 = \nabla \cdot \mathscr{W}_{2,\rho_2} F[\rho_1, \rho_2] \,.$$

[23] 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.

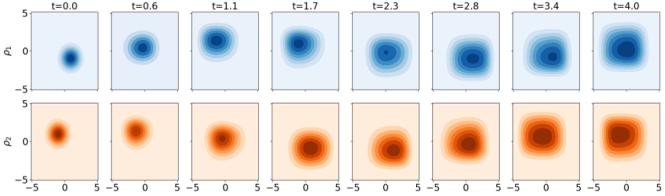
Degenerately monotone game

For $\mathcal{H}(\rho_i) = \int \log \rho_i(x_i) \mathrm{d}\rho_i(x_i)$, and $a, b, c > 0$, let

$$F[\rho_1, \rho_2] = \frac{a}{2} \int \|x_1\|^4 \mathrm{d}\rho_1(x_1) - \frac{a}{2} \int \|x_2\|^4 \mathrm{d}\rho_2(x_2) + \frac{b}{2} \iint x_1^\top x_2 \mathrm{d}\rho_1(x_1) \mathrm{d}\rho_2(x_2) + c(\mathcal{H}(\rho_1) - \mathcal{H}(\rho_2)) \,,$$

and consider dynamics

$$\partial_t \rho_1 = -\nabla \cdot \mathscr{W}_{2,\rho_1} F[\rho_1, \rho_2] \qquad \partial_t \rho_2 = \nabla \cdot \mathscr{W}_{2,\rho_2} F[\rho_1, \rho_2] \,.$$



*Explains why adding diffusion to a nearly-monotone game results in exponential convergence.

*Rule #5 in[23].

[23] 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.

Takeaways & Current Work

1. Defined monotonicity for mass-preserving flow dynamics. ✓
2. Properties of monotonicity and relation to convexity. ✓
3. Showed new asymptotic results for multispecies systems: contraction in $\overline{W}$ etc. ✓

Ongoing work:

- existence of solutions for multispecies gradient flow dynamics
- allowing mass to switch between species; applications in economic markets
- algorithms for solving vector-valued optimal transport


THANK YOU

Proof of main theorem

(a) Time derivative of the joint Wasserstein-2 metric between solutions ρ and η to (1) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{W}(\rho(t),\eta(t))^2 = 2\int \langle x-y,\, v[\eta(t)](x) - v[\rho(t)](y)\rangle \,\mathrm{d}\gamma(t,x,y)\,,$$

via[24] where $\gamma_i(t) \in \Gamma^*(\rho_i(t),\eta_i(t))$. Apply monotonicity inequality, for contraction in $\overline{W}$. Apply[25] to obtain the existence of a unique stationary point $\rho^\infty \in \overline{\mathcal{P}_2}$.

(b) Compute time derivative of $\overline{W}(\rho(t),\rho^\infty)$ to obtain convergence to steady state.

[24] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces. 2023. arXiv:2305.05211, Theorem 3.14.

[25] José A. Carrillo and Guiseppe Toscani. Contractive Probability Metrics and Asymptotic Behavior of Dissipative Kinetic Equations. *Rivista di Matematica della Universita di Parma*, 2007, Lemma 7.3.

Proof of main theorem

(a) Time derivative of the joint Wasserstein-2 metric between solutions ρ and η to (1) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{W}(\rho(t),\eta(t))^2 = 2\int \langle x-y,\, v[\eta(t)](x) - v[\rho(t)](y)\rangle \,\mathrm{d}\gamma(t,x,y)\,,$$

via[24] where $\gamma_i(t) \in \Gamma^*(\rho_i(t),\eta_i(t))$. Apply monotonicity inequality, for contraction in $\overline{W}$. Apply[25] to obtain the existence of a unique stationary point $\rho^\infty \in \overline{\mathcal{P}_2}$.

(b) Compute time derivative of $\overline{W}(\rho(t),\rho^\infty)$ to obtain convergence to steady state.

(c) Use assumptions to show that ρ^∞ is a Nash equilibrium,

$$F_i[\rho_i^\infty,\rho_{-i}^\infty] \le F_i[\rho_i,\rho_{-i}^\infty] \quad \forall\ \rho_i \in \mathcal{A}\,, \quad \forall\ i \in [1,\ldots,n]\,.$$

---

[24] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces. 2023. arXiv:2305.05211, Theorem 3.14.

[25] José A. Carrillo and Guiseppe Toscani. Contractive Probability Metrics and Asymptotic Behavior of Dissipative Kinetic Equations. *Rivista di Matematica della Universita di Parma*, 2007, Lemma 7.3.

Proof of main theorem

(a) Time derivative of the joint Wasserstein-2 metric between solutions ρ and η to (1) is

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{W}(\rho(t),\eta(t))^2 = 2\int \langle x-y,\, v[\eta(t)](x) - v[\rho(t)](y)\rangle \,\mathrm{d}\gamma(t,x,y)\,,$$

via[24] where $\gamma_i(t) \in \Gamma^*(\rho_i(t),\eta_i(t))$. Apply monotonicity inequality, for contraction in $\overline{W}$. Apply[25] to obtain the existence of a unique stationary point $\rho^\infty \in \overline{\mathcal{P}_2}$.

(b) Compute time derivative of $\overline{W}(\rho(t),\rho^\infty)$ to obtain convergence to steady state.

(c) Use assumptions to show that ρ^∞ is a Nash equilibrium,

$$F_i[\rho_i^\infty,\rho_{-i}^\infty] \le F_i[\rho_i,\rho_{-i}^\infty] \quad \forall\ \rho_i \in \mathcal{A}\,, \quad \forall\ i \in [1,\dots,n]\,.$$

(*) Uniform bounds on second moment, Lemma 3.6. Discussion for general metric spaces, Appendix A.2.

---

[24] Giulia Cavagnari, Giuseppe Savaré, and Giacomo Enrico Sodini. A Lagrangian approach to totally dissipative evolutions in Wasserstein spaces. 2023. arXiv:2305.05211, Theorem 3.14.

[25] José A. Carrillo and Guiseppe Toscani. Contractive Probability Metrics and Asymptotic Behavior of Dissipative Kinetic Equations. *Rivista di Matematica della Universita di Parma*, 2007, Lemma 7.3.

Convergence of Velocity Fields

We define a "dissipation"-like function

$$D(\rho) = \sum_{i=1}^{n} D_i(\rho)\,, \quad D_i(\rho) = \frac{1}{2}\int \|v_i[\rho](x_i)\|^2 \mathrm{d}\rho_i(x_i)\,. \tag{4}$$

Convergence of Velocity Fields

We define a "dissipation"-like function

$$D(\rho) = \sum_{i=1}^{n} D_i(\rho)\,, \quad D_i(\rho) = \frac{1}{2}\int \|v_i[\rho](x_i)\|^2 \mathrm{d}\rho_i(x_i)\,. \tag{4}$$

Proposition (Convergence of D)

Let v be λ-monotone in $\mathcal{A}$ with $\lambda \geq 0$, $D(\rho(0)) < \infty$. Then D along solutions to (1) *satisfies**

$$D(\rho(t)) \leq e^{-2\lambda t} D(\rho(0)) \quad \forall\, t \geq 0\,. \tag{5}$$

Convergence of Velocity Fields

We define a "dissipation"-like function

$$D(\rho) = \sum_{i=1}^{n} D_i(\rho)\,, \quad D_i(\rho) = \frac{1}{2}\int \|v_i[\rho](x_i)\|^2 \mathrm{d}\rho_i(x_i)\,. \tag{4}$$

Proposition (Convergence of D)

Let v be λ-monotone in $\mathcal{A}$ with $\lambda \geq 0$, $D(\rho(0)) < \infty$. Then D along solutions to (1) *satisfies**

$$D(\rho(t)) \leq e^{-2\lambda t} D(\rho(0)) \quad \forall\, t \geq 0\,. \tag{5}$$

*Need v_i sufficiently regular in ρ and x, and v_i in subset of tangent space at ρ:

$$v_i[\rho] \in \overline{\{c(r_i - \mathrm{id}) \in L_2(\rho_i) \;:\; (\mathrm{id} \times r_i)_\sharp \rho_i \in \Gamma^*(\rho_i, (r_i)_\sharp \rho_i)\,,\ (r_i)_\sharp \rho_i \in \mathcal{A}_i\,,\ c > 0\}}^{L_2(\rho_i)} \subset \mathcal{T}_\rho \overline{\mathcal{P}_2}.$$

Convergence of Velocity Fields

We define a "dissipation"-like function

$$D(\rho) = \sum_{i=1}^{n} D_i(\rho)\,, \quad D_i(\rho) = \frac{1}{2}\int \|v_i[\rho](x_i)\|^2 \mathrm{d}\rho_i(x_i)\,. \tag{4}$$

Proposition (Convergence of D)

Let v be λ-monotone in $\mathcal{A}$ with $\lambda \geq 0$, $D(\rho(0)) < \infty$. Then D along solutions to (1) *satisfies**

$$D(\rho(t)) \leq e^{-2\lambda t} D(\rho(0)) \quad \forall\, t \geq 0\,. \tag{5}$$

*Need v_i sufficiently regular in ρ and x, and v_i in subset of tangent space at ρ:

$$v_i[\rho] \in \overline{\{c(r_i - \mathrm{id}) \in L_2(\rho_i) \;:\; (\mathrm{id} \times r_i)_\sharp \rho_i \in \Gamma^*(\rho_i, (r_i)_\sharp \rho_i)\,,\; (r_i)_\sharp \rho_i \in \mathcal{A}_i\,,\; c > 0\}}^{L_2(\rho_i)} \subset \mathcal{T}_\rho \overline{\mathcal{P}_2}.$$

Monotonicity with $\lambda > 0$ implies exponential convergence of the velocity fields.