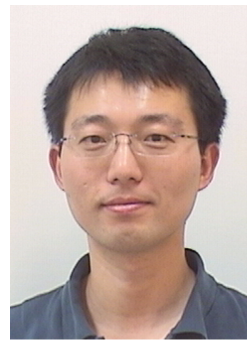
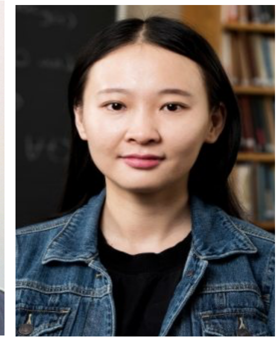
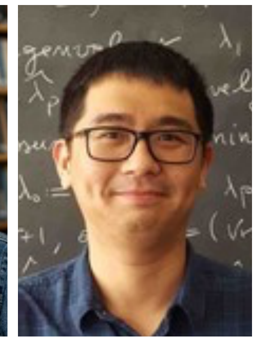
Learning Interaction Laws in Particle- and Agent-based Systems

MPML Series

Mauro Maggioni
mauromaggioni.duckdns.org
Johns Hopkins University

AFOSR NSF Bloomberg Philanthropies
SIMONS FOUNDATION

Joint work with:


F. Lu S. Tang, M. Zhong, J. Miller

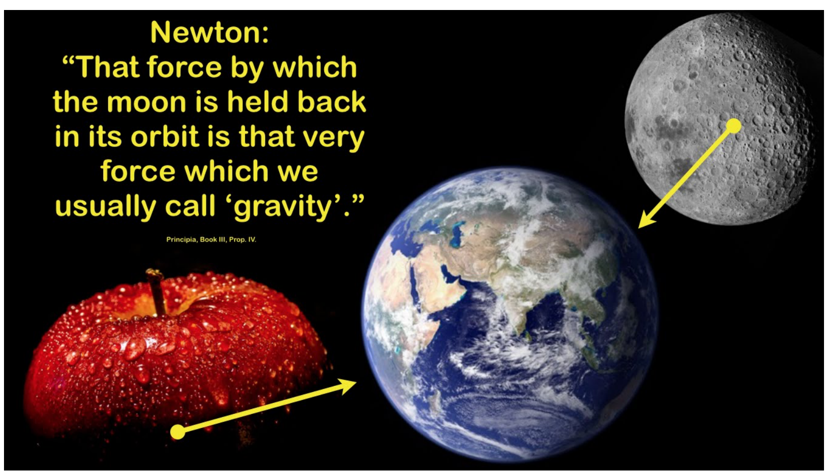
Felix Munoz, https://www.youtube.com/watch?v=OxYn3e_imhA

BBC Blue Planet (clip from YouTube)

Learning Interaction Laws

Problem: Given observations of trajectories of a dynamical system of interacting agents, learn the interaction rules.
Motivation: particle-/agent-based systems ubiquitous in Physics, Biology, social sciences, Economics, ... Beyond model-based interaction rules.
Further goals: hypothesis testing for agent-based systems; transfer learning; agents on networks; collaborative and competitive games.

Renewed interest in learning ODE's and PDE's in Applied Math and Engineering e.g. S. Osher, H. Shaeffer, N. Kutz, Y. Kevrekidis, D. Giannakis, C. Shü tte, R. Ward...; also in the ML community, e.g. Battaglia, Tenenbaum et al., especially in view of applications to control/reinforcement learning.



From https://www.youtube.com/watch?v=gJhn7WmXWVY

Felix Munoz, https://www.youtube.com/watch?v=OxYn3e_imhA

Estimation/Learning for ODE systems

Suppose we have a system driven by of ODEs in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) \qquad , \mathbf{x} \in \mathbb{R}^D, \mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$$

and we are given observations of positions and velocities

$$(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\dots,L; m=1,\dots,M} \ ,$$

where:

· $0 = t_1 < \cdots < t_L = T$;

· m indexes trajectories corresponding to different initial conditions at $t_1 = 0$

Objective: construct an estimator $\hat{\mathbf{f}}$ that is close to $\mathbf{f}$.

No randomness: this is an approximation problem.

Randomness: this is a statistical problem. Sources of randomness:

· initial conditions $\mathbf{x}(0)$ are random, e.g $\sim_{\text{i.i.d}} \mu_0$, a prob. meas. on $\mathbb{R}^D$;

· the observations are corrupted by noise.

Nonparametric regression

Suppose we have a system driven by of ODEs in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) \qquad , \mathbf{x} \in \mathbb{R}^D, \mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$$

and we are given observations of positions and velocities

$$(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\dots,L; m=1,\dots,M} \,,$$

where:

- $0 = t_1 < \cdots < t_L = T$;
- m indexes trajectories corresponding to different initial conditions at $t_1 = 0$

Objective: construct an estimator $\hat{\mathbf{f}}$ that is close to $\mathbf{f}$.

Given $(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\dots,L; m=1,\dots,M}$, with $\mathbf{x}^{(m)}(0) \sim_{\text{i.i.d.}} \mu_0$, we want to approximate the unknown $\mathbf{f}$ in $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$.

Possible approach: regression. In regression one is given pairs

$$\{(\mathbf{z}_i, f(\mathbf{z}_i) + \eta_i)\}_{i=1}^n \,, \text{with } \mathbf{z}_i \sim_{\text{i.i.d.}} \rho \,,$$

ρ a probability measure on $\mathbb{R}^D$ and η independent noise, and outputs an estimator $\hat{\mathbf{f}}_n$.

Nonparametric regression

Suppose we have a system driven by of ODEs in the form

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t)) \qquad , \mathbf{x} \in \mathbb{R}^D, \mathbf{f} : \mathbb{R}^D \to \mathbb{R}^D$$

and we are given observations of positions and velocities

$$(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\dots,L;m=1,\dots,M} \, ,$$

where:

- $0 = t_1 < \cdots < t_L = T$;
- m indexes trajectories corresponding to different initial conditions at $t_1 = 0$

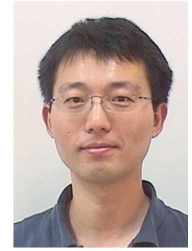
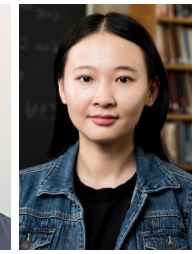
Objective: construct an estimator $\hat{\mathbf{f}}$ that is close to $\mathbf{f}$.

Map our problem to regression? $\mathbf{z}_i = \mathbf{x}^{(m)}(t_l)$ and $\mathbf{f}(\mathbf{z}_i) = \dot{\mathbf{x}}^{(m)}(t_l)$ (so i runs over the product of the index sets for l and m). Warning: no independence in l.

Even if we pretended to have independence, without further assumptions on $\mathbf{f}$, besides s-Hölder regularity, the best attainable rate is $n^{-\frac{s}{2s+D}}$, where $n = LM$ (L observations in each of M trajectories) and $D = Nd$ (N agents in $\mathbb{R}^d$).

For a system of N agents in $\mathbb{R}^d$, $D = Nd$ is typically very large, and the rate $n^{-\frac{s}{2s+D}}$ unsatisfactory. Further assumptions are needed for better rates.

Agent-based systems

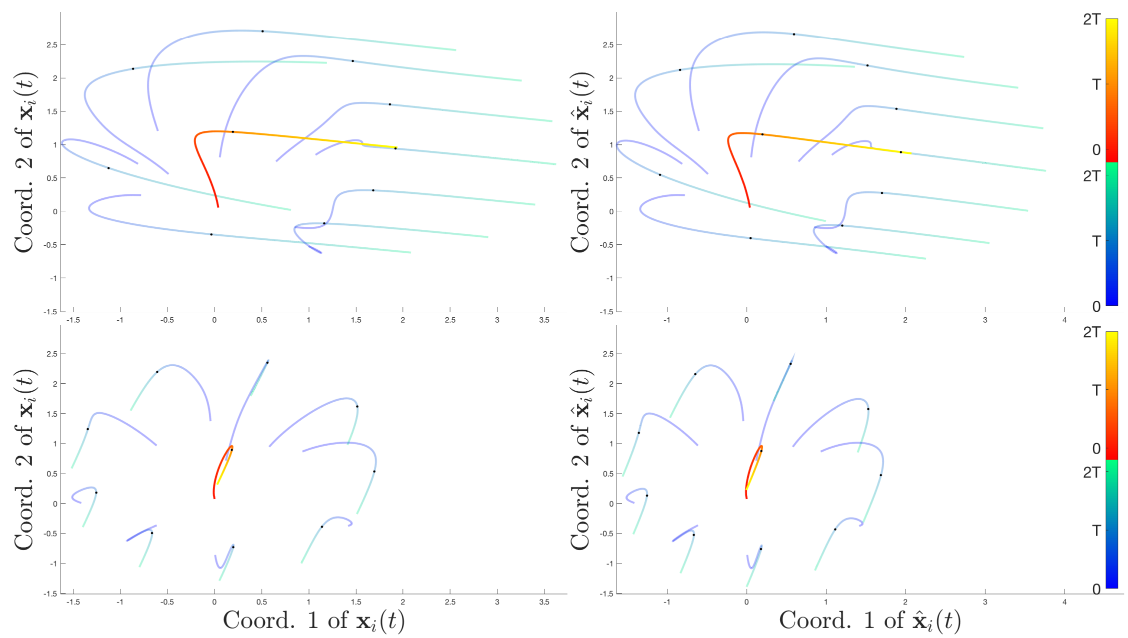

Particle- and agent-based systems are driven by ODEs with special structure. A widely used model:

$$\dot{\mathbf{x}}_i^{(m)} = \frac{1}{N}\sum_{i'=1}^{N} \phi(||\mathbf{x}_i^{(m)} - \mathbf{x}_{i'}^{(m)}||)(\mathbf{x}_{i'}^{(m)} - \mathbf{x}_i^{(m)})$$

Given observations $\{(\mathbf{x}_i, \dot{\mathbf{x}}_i)\}_{i=1}^{N}$ at different times $\{t_l\}_{l=1}^{L}$ and/or for different initial conditions $\{\mathbf{x}^{(m)}(0)\}_{m=1}^{M}$, we want to learn the interaction kernel ϕ. Different limits: $N \to +\infty$ (mean-field limit, joint work with M. Fornasier and M. Bongini), $M \to +\infty$ (joint current work with F. Lu, M. Zhong and S. Tang).

Interesting extensions to:
- higher-order systems,
- stochastic systems,
- agents of different types,
- varying environment.

...



Second-order prey-predator model. Left: true trajectories; Right: trajectories with learned interactions.

The Mean-field limit

Rewriting

$$\dot{\mathbf{x}}_i = \frac{1}{N}\sum_{i'} \phi(||\mathbf{x}_i - \mathbf{x}_{i'}||)(\mathbf{x}_{i'} - \mathbf{x}_i) = \frac{1}{N}\sum_{i'} \frac{\Phi'(||\mathbf{x}_i - \mathbf{x}_{i'}||)}{||\mathbf{x}_i - \mathbf{x}_{i'}||}(\mathbf{x}_i - \mathbf{x}_{i'})$$

we see this is the gradient flow of the energy $\mathcal{J}_N(\mathbf{X}) = \frac{1}{2N}\sum_{i,i'=1}^{N} \Phi(||\mathbf{x}_i - \mathbf{x}_{i'}||)$.

Considering the measure $\mu^N(t) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\mathbf{x}_i(t)}$, we may let $N \to +\infty$ to obtain (under suitable regularity assumptions on Φ) the mean field equations

$$\partial_t \mu(t) = -\nabla \cdot \left(\left(-\frac{\Phi'(||\cdot||)}{||\cdot||} * \mu(t)\right)\mu(t)\right), \quad \mu(0) = \mu_0 .$$

This is also a gradient flow for the energy $\mathcal{J}(\mu) = \int_{\mathbb{R}^{d\times d}} \Phi(||\mathbf{x}-\mathbf{y}||)d\mu(\mathbf{x})d\mu(\mathbf{y})$ on the space of probability measures with Wasserstein distance.

Estimation in the limit as $N \to \infty$: studied in *Inferring Interaction Rules from Observations of Evolutive Systems I: The Variational Approach*, M. Bongini, M. Fornasier, M. Hansen, MM, published in M3S, 2017

Measures on pairwise distances

Observations: $\{(\mathbf{x}_i, \dot{\mathbf{x}}_i)^{(m)}(t_l)\}_{i=1,l=1,m=1}^{N,L,M}$, where $\mathbf{x}^{(m)}(0) \sim \mu_0$ for some μ_0 on $\mathbb{R}^d$. Note that each state of the system is in $\mathbb{R}^{dN}$.
All we want however is the one-dimensional interaction kernel ϕ in the equations

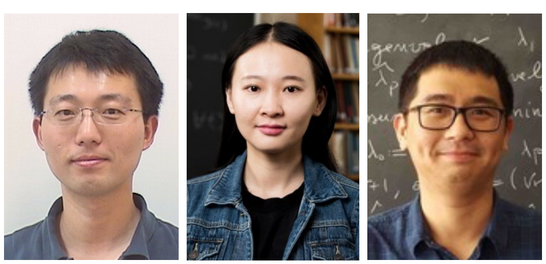
$$\dot{\mathbf{x}}_{i'}^{(m)}(t) = \frac{1}{N} \sum_{i'=1}^{N} \phi(\underbrace{||\mathbf{x}_{i'}^{(m)}(t) - \mathbf{x}_i^{(m)}(t)||}_{r_{ii'}^{(m)}(t)})(\mathbf{x}_{i'}^{(m)}(t) - \mathbf{x}_i^{(m)}(t)) \,.$$

For a fixed $t = t_l$ and m, we cannot solve for $\phi(r_{ii'})$: $N(N-1)/2$ unknowns and only dN knowns (typically $d \ll N$). We have to leverage observations in time.

At time scale $[0, T]$, we define the probability measure on $\mathbb{R}_+$:

$$\rho_T^L(r) := \mathbb{E}_{x(0)\sim\mu_0} \frac{1}{L} \sum_{l=1}^{L} \frac{1}{\binom{N}{2}} \sum_{i,i'=1, i<i'}^{N} \delta_{r_{ii'}^{(m)}(t_l)}(r) \,.$$

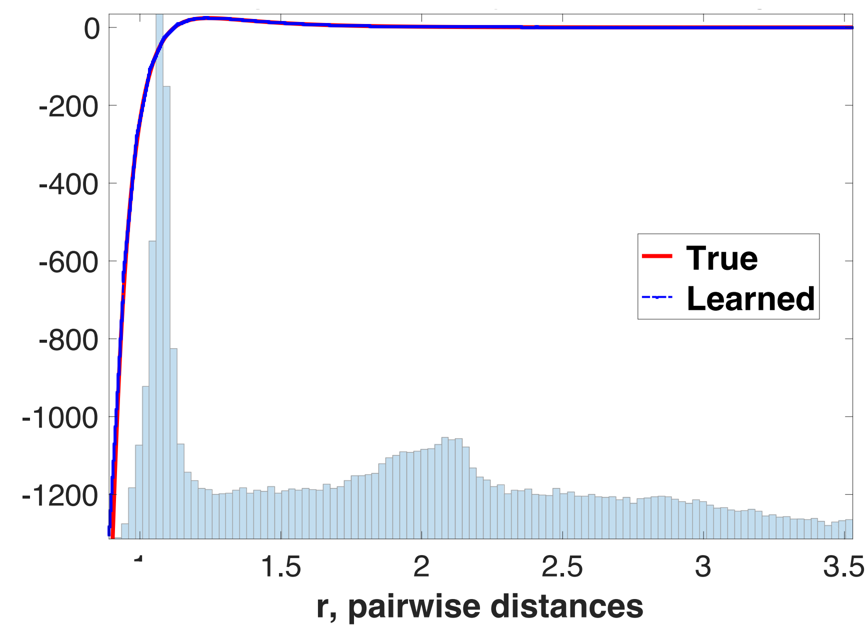
- ↑ average over initial conditions
- ↑ average over observations in time
- ↑ average over pairs of agents
- ↑ δ on $\mathbb{R}_+$ at every observed pairwise distance

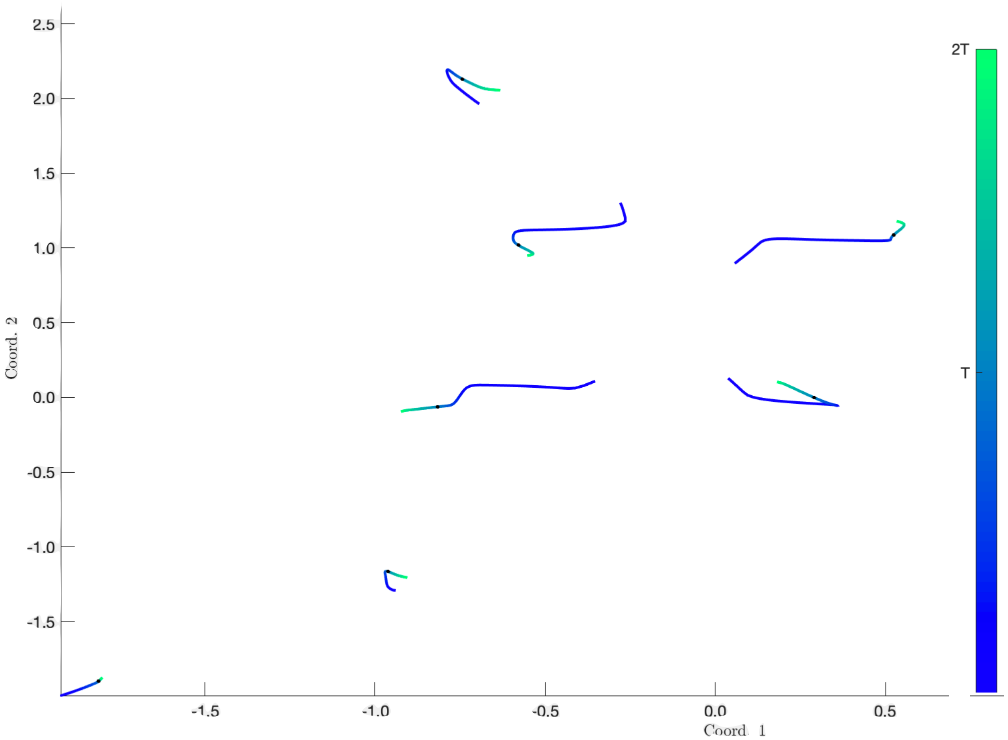


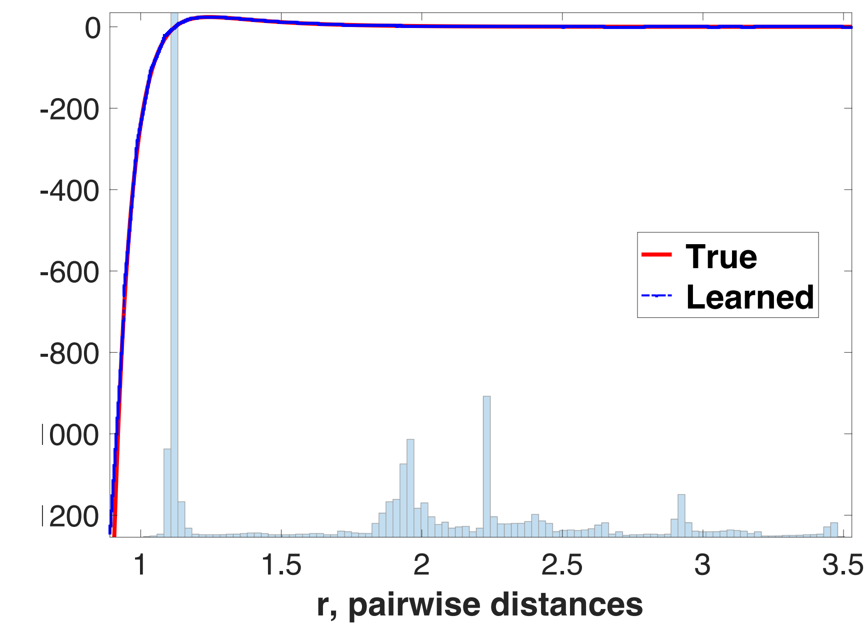
Nonparametric inference of interaction laws in systems of agents from trajectory data, Fei Lu, Ming Zhong, Sui Tang, and MM, P.N.A.S., 2019

Example: L-J kernel and ρ_L^T

Example. The Lennard Jones force is the derivative of the potential
$V_{LJ}(r) = 4\epsilon\left(\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right)$.
Right figure: In blue the LJ ϕ,
in red an empirical estimate of ρ_T^L,
for a system of $N = 7$ agents.
Two cases: L, T small, and L, T large.







The estimator

Observations: $\{(\mathbf{x}_i^{(m)}, \dot{\mathbf{x}}_i^{(m)})(t_l)\}_{I=1,l=1,m=1}^{N,L,M}$, for M different initial conditions i.i.d. $\sim \mu_0$, from

$$\dot{\mathbf{x}}_i^{(m)}(t) = \frac{1}{N}\sum_{i'} \phi(||\mathbf{x}_{i'}^{(m)}(t) - \mathbf{x}_i^{(m)}(t)||)(\mathbf{x}_{i'}^{(m)}(t) - \mathbf{x}_i^{(m)}(t)) =: \mathbf{f}_{\phi}(\mathbf{x}_i^{(m)}(t))\,.$$

linear map applied to unknown ϕ

Consider the empirical error functional

$$\mathcal{E}_{L,M}(\varphi) := \frac{1}{LMN}\sum_{l,m,i=1}^{L,M,N} \left\|\dot{\mathbf{x}}_i^{(m)}(t_l) - \mathbf{f}_{\varphi}(\mathbf{x}_i^{(m)}(t_l))\right\|^2\,.$$

Our estimator is defined as a minimizer of $\mathcal{E}_{L,M}$ over $\varphi \in \mathcal{H}$, a suitable hypothesis space of functions on $\mathbb{R}_+$, with $\dim(\mathcal{H}) = n$ (with $n = n(M)$):

$$\hat{\phi}_{L,M,\mathcal{H}} := \arg\min_{\varphi\in\mathcal{H}} \mathcal{E}_{L,M}(\varphi)\,.$$

For $\mathcal{H}$ linear subspace, this is a least squares problem (Gauss, Legendre); the subspace serves as a regularizer.

Coercivity condition

$$\mathcal{E}_{L,M}(\varphi) := \frac{1}{LMN}\sum_{l,m,i=1}^{L,M,N} \left\|\dot{\mathbf{x}}_i^{(m)}(t_l) - \mathbf{f}_\varphi(\mathbf{x}_i^{(m)}(t_l))\right\|^2,$$

$$\hat{\phi}_{L,M,\mathcal{H}} := \arg\min_{\varphi\in\mathcal{H}} \mathcal{E}_{L,M}(\varphi)\,.$$

We shall assume that the unknown interaction kernel ϕ is in the admissible class $\mathcal{K}_{R,S} := \{\varphi \in C^1(\mathbb{R}_+) :\ \mathrm{supp}.\varphi \subset [0,R], \sup_{r\in[0,R]} |\varphi(r)| + |\varphi'(r)| \leq S\}$.
Coercivity condition: $\forall\varphi\ :\ \varphi(\cdot)\cdot \in \mathcal{H}$, for $c_{L,N,\mathcal{H}}$

$$c_{L,N,\mathcal{H}}\|\varphi(\cdot)\cdot\|^2_{L^2(\rho_T^L)} \leq \frac{1}{NL}\sum_{l,i=1}^{L,N} \mathbb{E}\left\|\frac{1}{N}\sum_{i'=1}^{N} \varphi(r_{ii'}(t_l))\mathbf{r}_{ii'}(t_l)\right\|^2.$$

Lemma. Coercivity $\implies$ unique minimizer of $\lim_{M\to+\infty}\mathcal{E}_{L,M}(\varphi)$ over $\varphi \in \mathcal{H}$

$$\varphi - \phi \in \mathcal{H} \implies c_{L,N,\mathcal{H}}||\varphi(\cdot)\cdot - \phi(\cdot)\cdot||^2_{L^2(\rho_T^L)} \leq \mathcal{E}_{L,\infty}(\varphi - \phi)$$

The coercivity constant $c_{L,N,\mathcal{H}}$ also controls the condition number of the matrix in the least squares problem yielding $\hat{\phi}_{L,M,\mathcal{H}}$.

Bias/variance trade-off

$$\mathcal{E}_{L,M}(\varphi) := \frac{1}{LMN} \sum_{l,m,i=1}^{L,M,N} \left\| \dot{\mathbf{x}}_i^{(m)}(t_l) - \mathbf{f}_\varphi(\mathbf{x}_i^{(m)}(t_l)) \right\|^2,$$

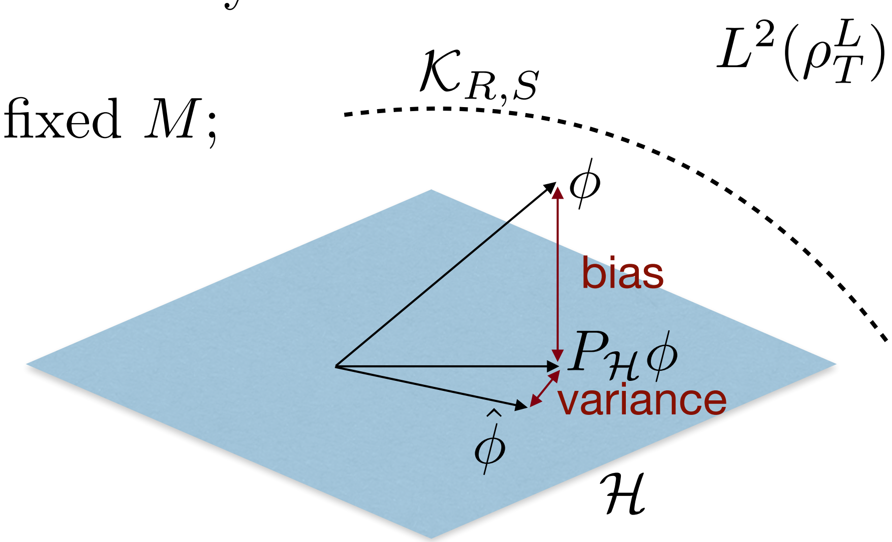
$$\hat{\phi}_{L,M,\mathcal{H}} := \arg\min_{\varphi \in \mathcal{H}} \mathcal{E}_{L,M}(\varphi)\,.$$

$+$ coercivity

bias decreases as $\dim \mathcal{H}$ increases; depends only on approximation properties of $\mathcal{H}$

variance increases as $\dim \mathcal{H}$ increases, for fixed M; measures randomnesso of $\hat{\phi} \in \mathcal{H}$

Pick $\dim \mathcal{H}$ an increasing function of M, to attain the minimum of the sum of bias (squared) and variance.



Unlike regression, we do not have access to values of ϕ, but only observations that are linear functions (via f_ϕ) of ϕ; coercivity implies stable invertibility.

Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \leq c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot) - \phi(\cdot)\|_{L^\infty([0,R])} \leq c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \leq \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}} .$$

- The good: Rate in M is optimal, in fact even optimal in the case of regression, where we would be given $(r_m, \phi(r_m))_{m=1}^M$.
- The bad: no dependency on L. Numerical examples: suggest in some cases that effective sample size is $LM = \#\text{obs}$.

We choose $\mathcal{H}_n$ to be the space of piecewise linear functions on a uniform partition of cardinality n of $[0, R_{\max}]$ (estimated supp.ρ_L^T), for $n = n_*$.

In the end solving the minimization problem is a least-squares problem in $n = n_*$ dimensions. Algorithms for constructing the LS matrix and computing the estimator run in time $O(N^2Ld \cdot M + Mn_*^2)$ (online versions also possible).

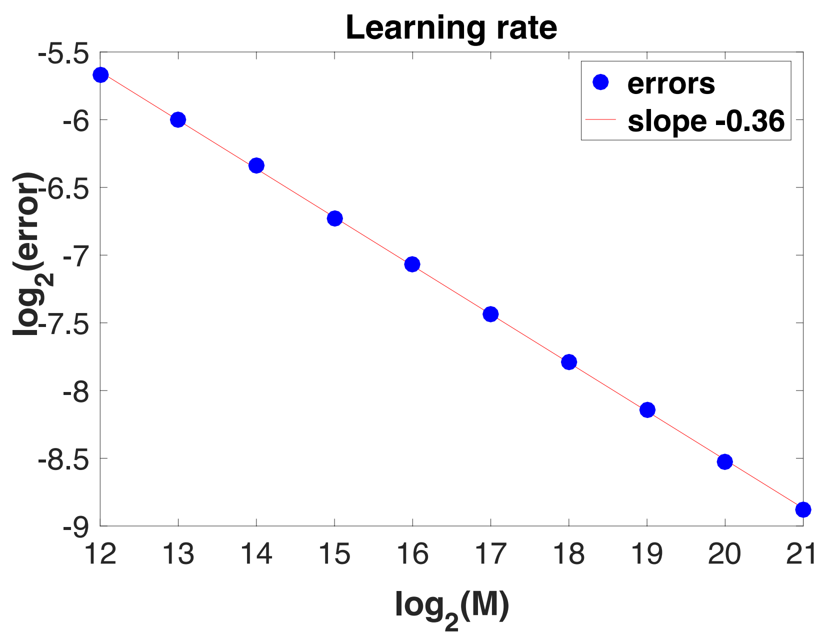
Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \leq c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot) - \phi(\cdot)\|_{L^\infty([0,R])} \leq c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \leq \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}}.$$

- The good: Rate in M is optimal, in fact even optimal in the case of regression, where we would be given $(r_m, \phi(r_m))_{m=1}^M$.
- The bad: no dependency on L.

Example. The Lennard Jones kernel is *not* admissible, yet since particles rarely get very close to each other, we obtain a convergence rate close to optimal.



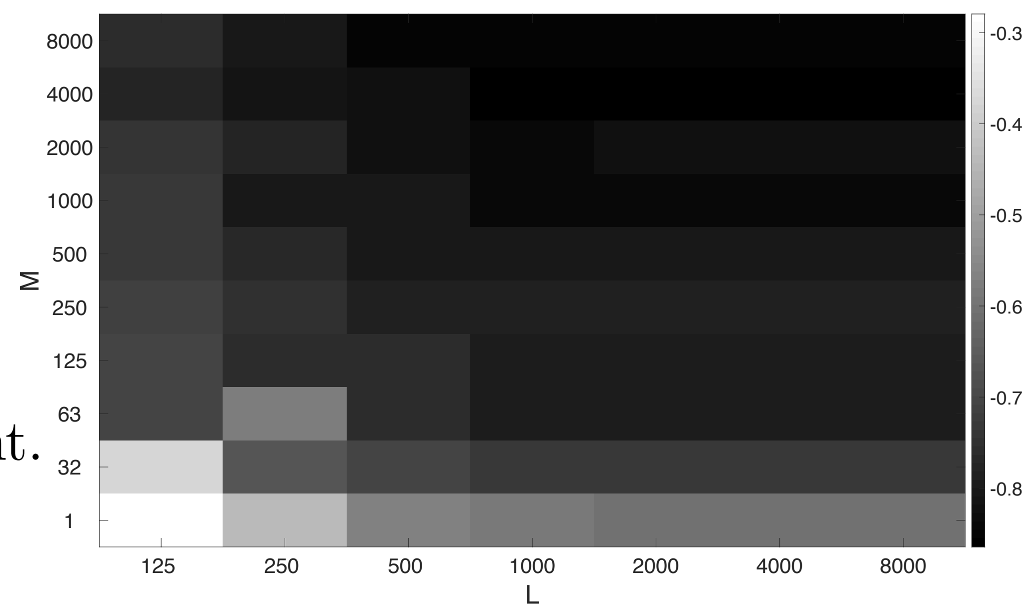
Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \leq c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot) - \phi(\cdot)\|_{L^\infty([0,R])} \leq c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \leq \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}} .$$

- The good: Rate in M is optimal, in fact even optimal in the case of regression, where we would be given $(r_m, \phi(r_m))_{m=1}^M$.
- The bad: no dependency on L.

Numerical results suggest that the effective sample size should scale linearly in L, at least up to a point.

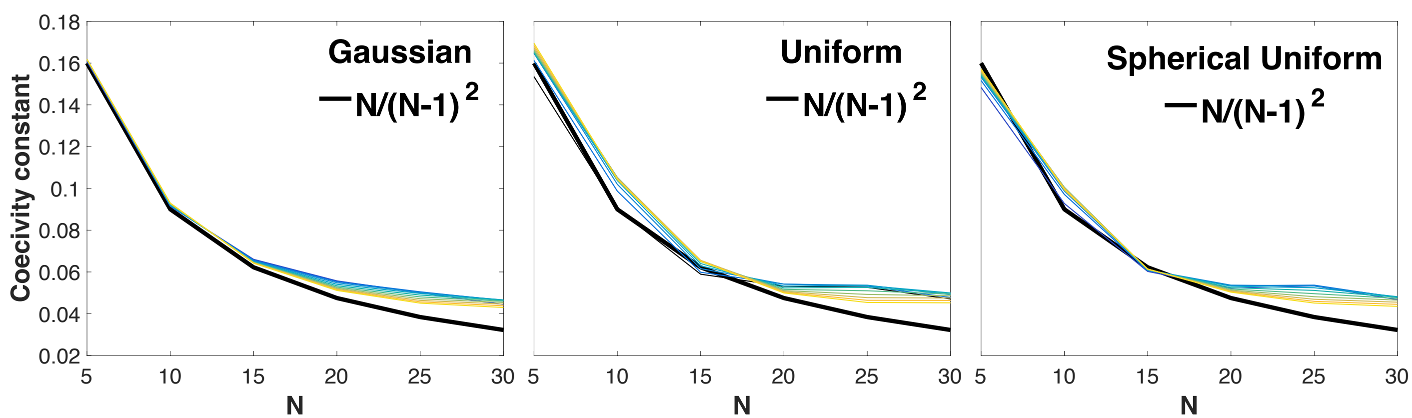


Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \le c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot) - \phi(\cdot)\|_{L^\infty([0,R])} \le c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \le \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}}.$$

$c_{L,N,\mathcal{H}}$ can be as small as $\frac{N-1}{N^2}$, but in fact we conjecture that under some general conditions it is independent of N when evaluated on compact subspaces $\mathcal{H} \subset L^2(\rho_T^L)$. We can prove this in special cases, for $L = 1$ and μ_0 exchangeable Gaussian.



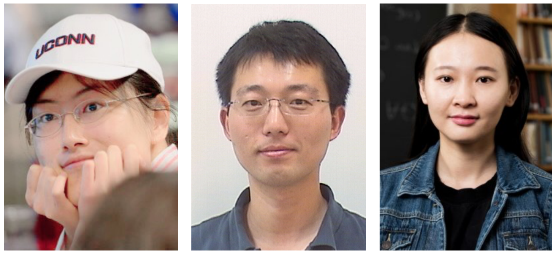
Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \le c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot)-\phi(\cdot)\|_{L^\infty([0,R])} \le c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \le \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}} .$$

Theorem. Suppose $L=1, N>1$ and assume that the distribution of $\mathbf{X}(t_1) = (\mathbf{x}_1(t_1),\dots,\mathbf{x}_N(t_1))$ is exchangeable Gaussian with $\mathrm{cov}(\mathbf{x}_i) - \mathrm{cov}(\mathbf{x}_i,\mathbf{x}_{i'}) = \lambda I_d$ for some constant $\lambda > 0$ (and all i, i'). Then the coercivity condition holds in $L^2(\rho_T^L)$ with constant $c_L = \frac{N-1}{N^2}$, and on any compact hypothesis space $\mathcal{H} \subset L^2(\rho_T^L)$ with a constant $c_{L,N,\mathcal{H}} = c_{\mathcal{H}}$ independent of N.

Many significant generalizations possible guaranteeing coercivity.



On the identifiability of interaction functions in systems of interacting particles, Z. Li, F. Lu, MM, S. Tang, C. Zhang, to appear in Stochastic Processes and their Applications, arxiv.org/pdf/1912.11965.pdf

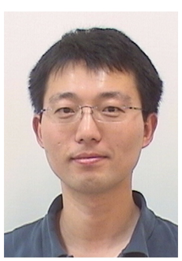
Main Theorem (first order systems)

Theorem. Let $\{\mathcal{H}_n\}_n \subseteq \mathcal{H}$ be a sequence of subspaces of $L^\infty[0,R]$, with $\dim(\mathcal{H}_n) \le c_0 n$ and $\inf_{\varphi\in\mathcal{H}_n} \|\varphi(\cdot) - \phi(\cdot)\|_{L^\infty([0,R])} \le c_1 n^{-s}$, for some constants $c_0, c_1, s > 0$. It exists, for example, if ϕ is s-Hölder regular.
Choose $n_* = (M/\log M)^{\frac{1}{2s+1}}$: then for some $C = C(c_0, c_1, R, S)$

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \le \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+1}} .$$

This result may be extended to heterogeneous agent systems (e.g. prey-predator) [Lu, Tang, MM, on arXiV].

It may also be extended to multi-variable interaction kernels, depending on more observables than pairwise distances, as well as second-order systems (work in progress).




Learning interaction kernels in heterogeneous systems of agents from multiple trajectories, F. Lu, MM, S. Tang, arxiv 1910.04832

Learning theory for inferring interaction kernels in second-order interacting agent systems, J.Miller, M. Zhong, S. Tang, MM, in preparation.

Errors on trajectories

Standard arguments yield bounds on trajectories between trajectories of the true system and those of the system driven by the estimated interaction kernel.

Proposition. Assume $\widehat{\phi}(\|\cdot\|)\cdot \in \mathrm{Lip}(\mathbb{R}^d)$, with Lipschitz constant C_{Lip}. Let $\widehat{\mathbf{X}}(t)$ and $\mathbf{X}(t)$ be the solutions of systems with kernels $\widehat{\phi}$ and ϕ respectively, started from the same initial condition. Then for each trajectory

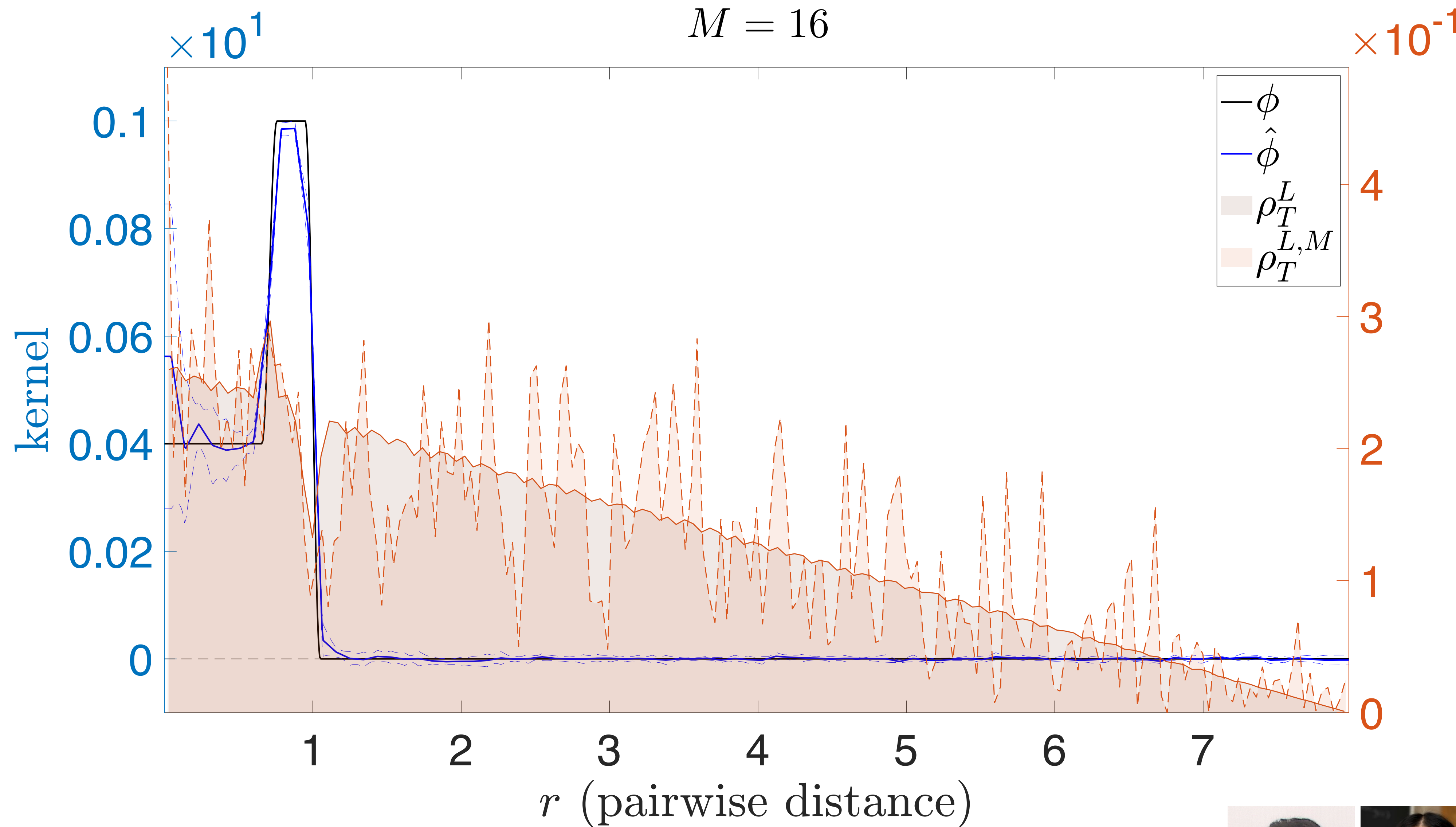
$$\sup_{t\in[0,T]} \|\widehat{\mathbf{X}}(t) - \mathbf{X}(t)\|^2 \leq 2Te^{8T^2C_{\mathrm{Lip}}^2}\int_0^T \left\|\dot{\mathbf{X}}(t) - \mathbf{f}_{\hat{\phi}}(\mathbf{X}(t))\right\|^2 dt\,,$$

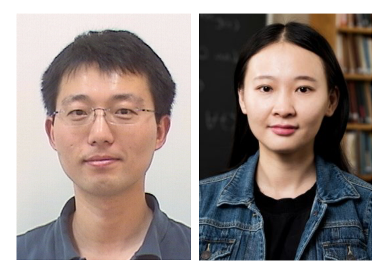
and on average w.r.t. the distribution μ_0 of initial conditions:

$$\mathbb{E}_{\mu_0}[\sup_{t\in[0,T]} \|\widehat{\mathbf{X}}(t) - \mathbf{X}(t)\|] \leq C(T, C_{\mathrm{Lip}})\sqrt{N}\|\hat{\phi}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_T)}\,,$$

where $C(T, C_{\mathrm{Lip}})$ is a constant depending on T and C_{Lip}.

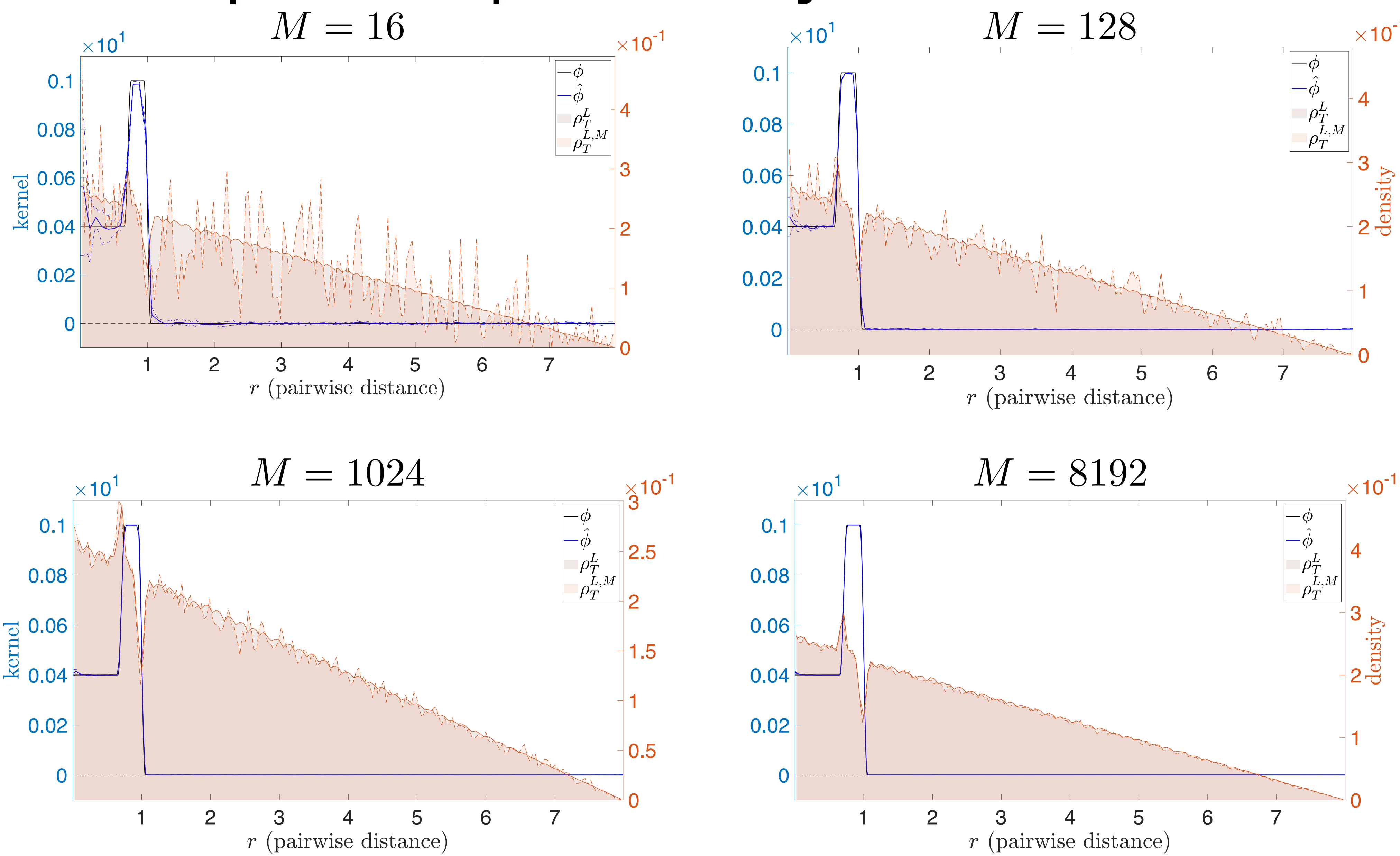
Examples: Opinion Dynamics





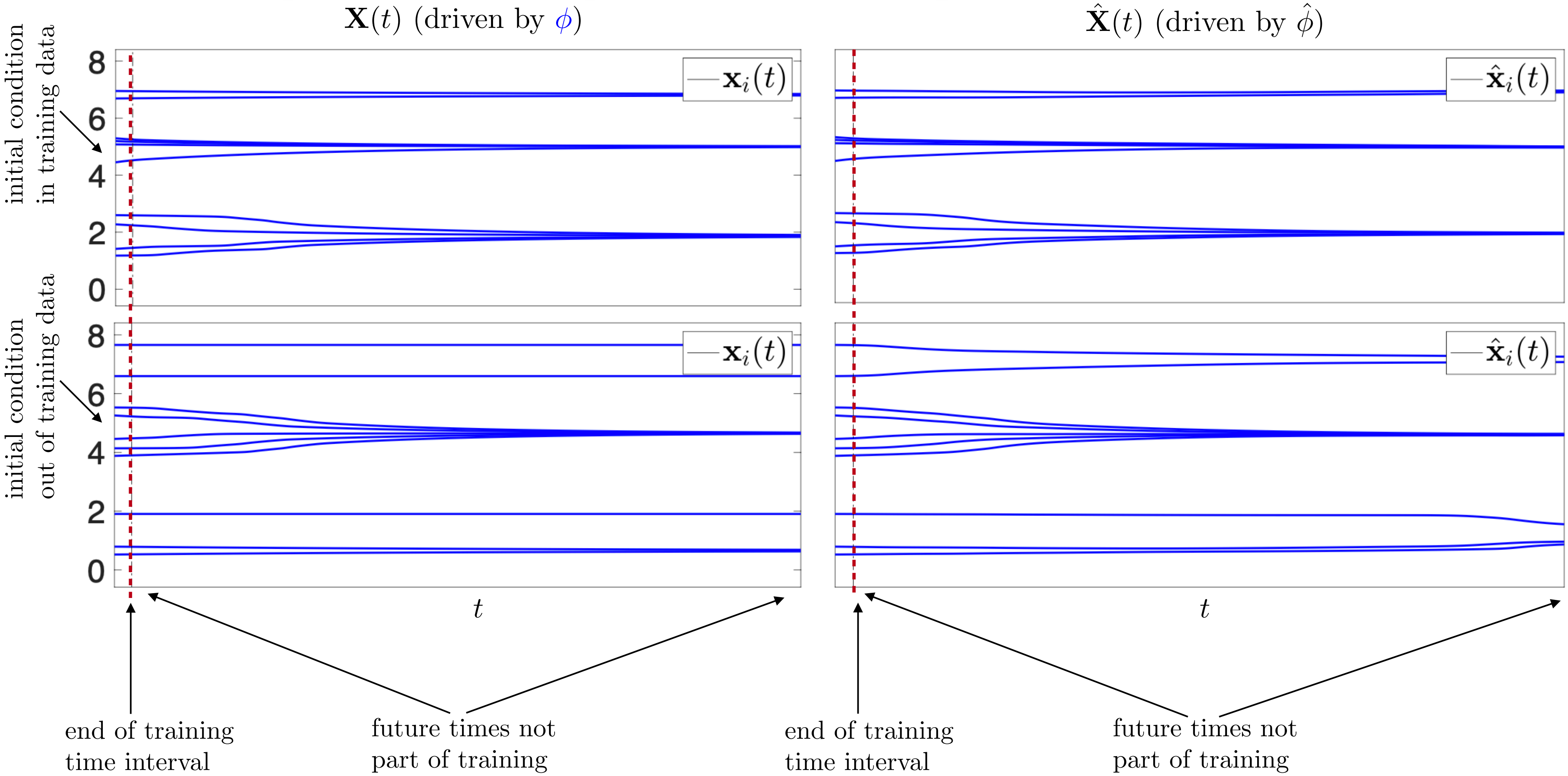
Learning interaction kernels in heterogeneous systems of agents from multiple trajectories,
F. Lu, MM, S. Tang, ArXiv 1910.04832, submitted.

Examples: Opinion Dynamics

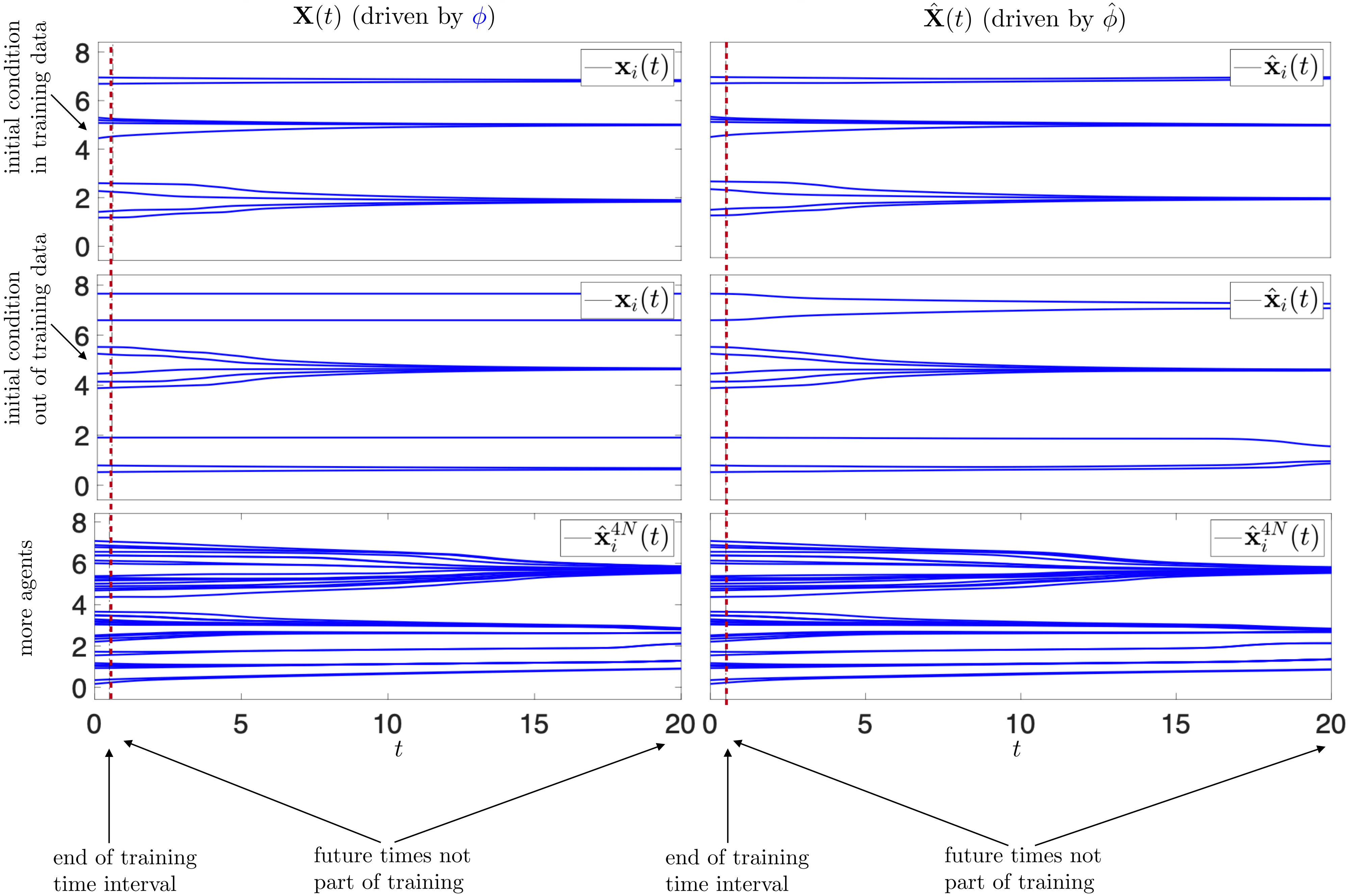


d	N	$M_{\rho_T^L}$	L	$[t_1; t_L; t_f]$	μ_0	$\deg(\psi)$	n
1	10	10^5	51	$[0; 0.5; 20]$	$\mathcal{U}([0, 8])$	0	$60(\frac{M}{\log M})^{\frac{1}{3}}$

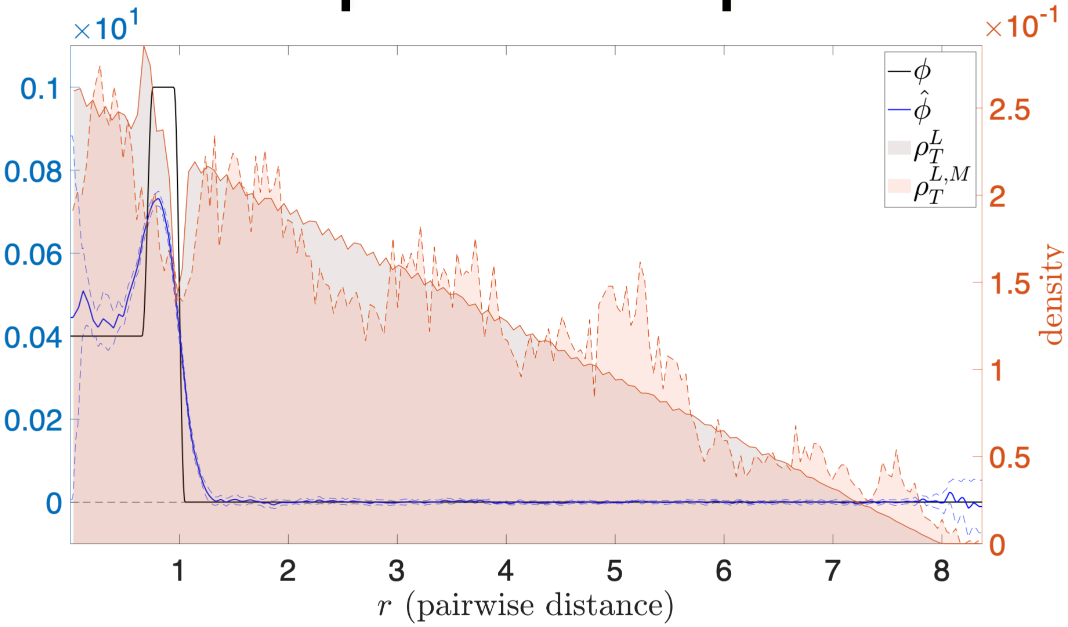
Examples: Opinion Dynamics

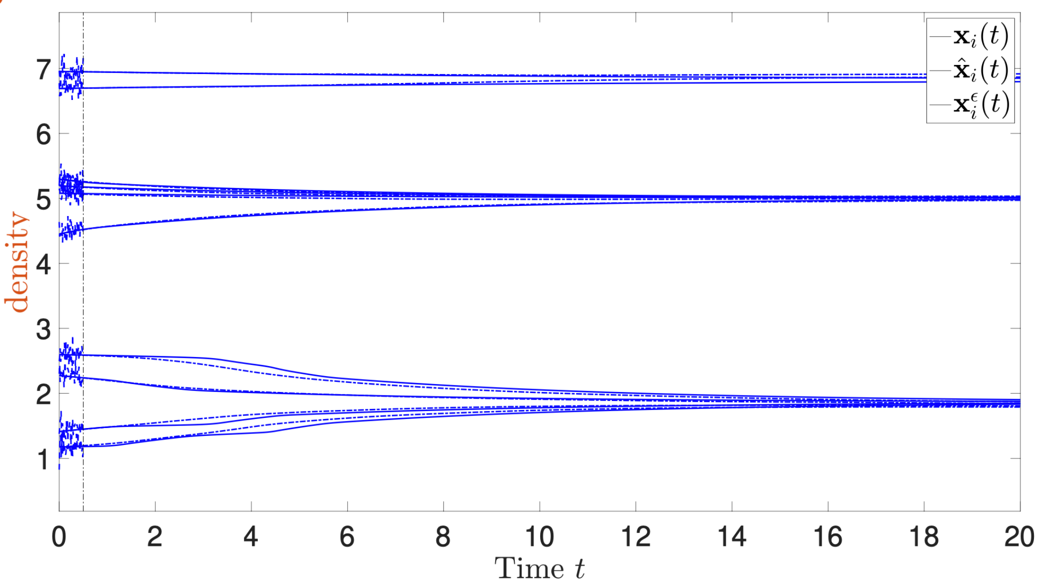


Examples: Opinion Dynamics



Examples: Opinion Dynamics

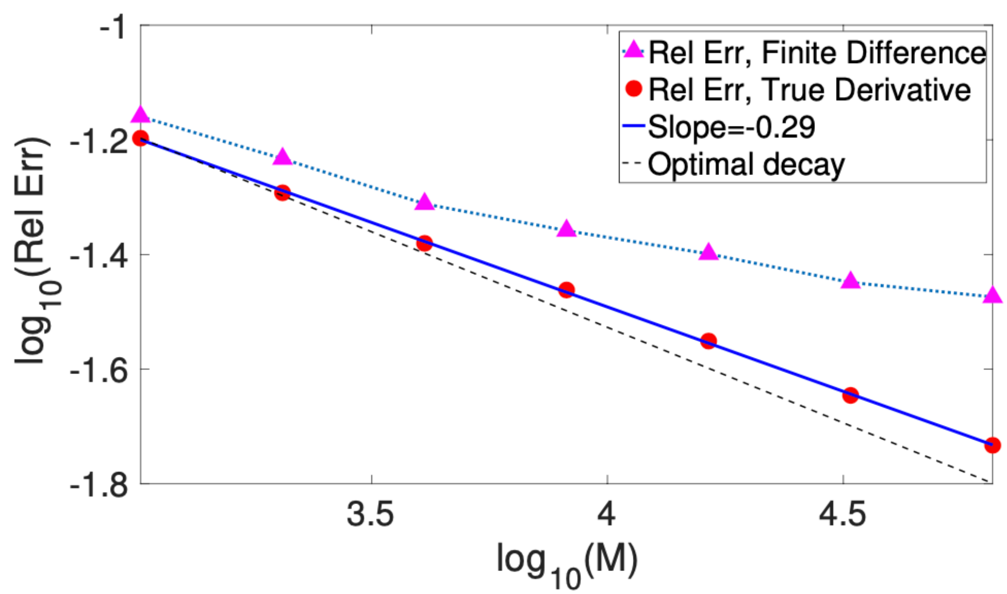




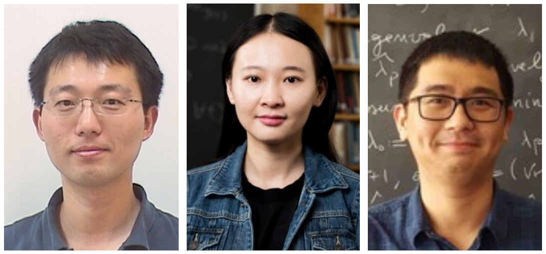
$Unif([-0.15, 0.15])$ noise in the observed positions and velocities.

Observing only positions but not velocities: the numerical approximation error (which depends on L) constrains the ability of estimating the interaction kernel.

From now on, all experiments will **not** have observed velocities.



Example: 2nd order systems

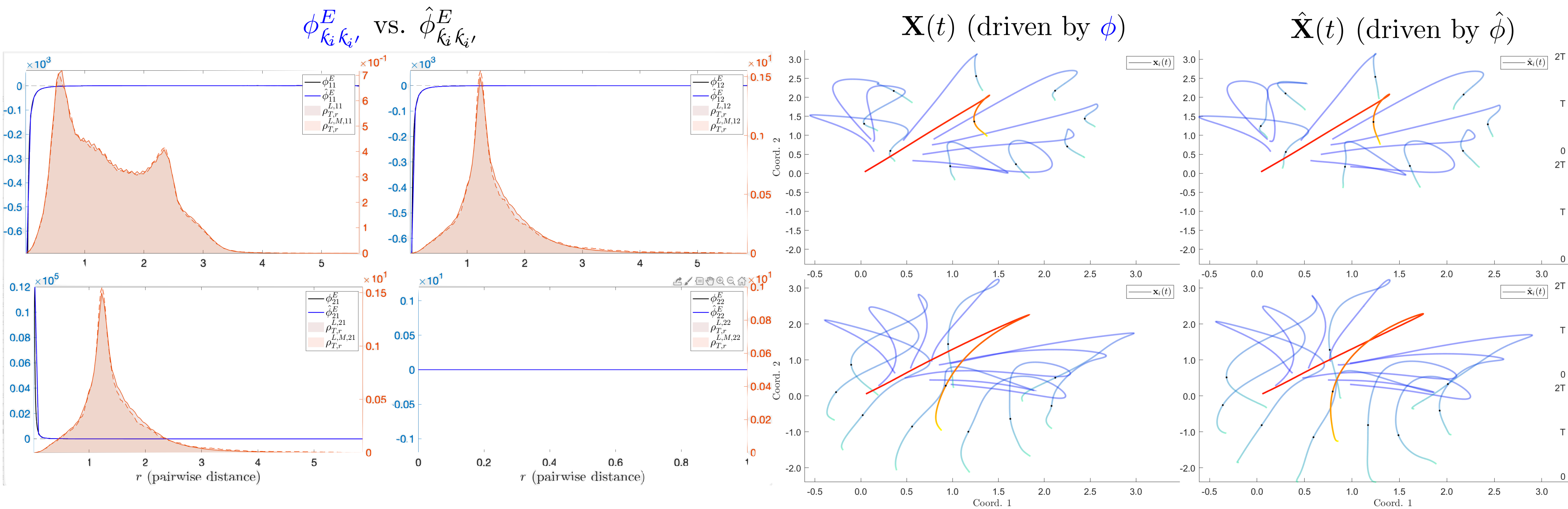


$$\begin{cases} m_i \ddot{\boldsymbol{x}}_i = F_i^{\boldsymbol{v}}(\dot{\boldsymbol{x}}_i, \xi_i) + \sum_{i'=1}^{N} \frac{\kappa^{\boldsymbol{v}}_{k_{i'}}}{N_{k_{i'}}} \left(\phi^E_{k_i k_{i'}}(r_{ii'}) \boldsymbol{r}_{ii'} + \phi^A_{k_i k_{i'}}(r_{ii'}) \dot{\boldsymbol{r}}_{ii'} \right) \\ \dot{\xi}_i = F_i^{\xi}(\xi_i) + \sum_{i'=1}^{N} \frac{\kappa^{\xi}_{k_{i'}}}{N_{k_{i'}}} \phi^{\xi}_{k_i k_{i'}}(r_{ii'}) \xi_{ii'} \end{cases}$$

simple environment (food, light, ...)

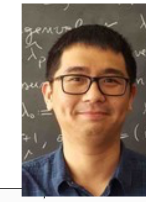
one kernel for each pair of interacting agent **types**

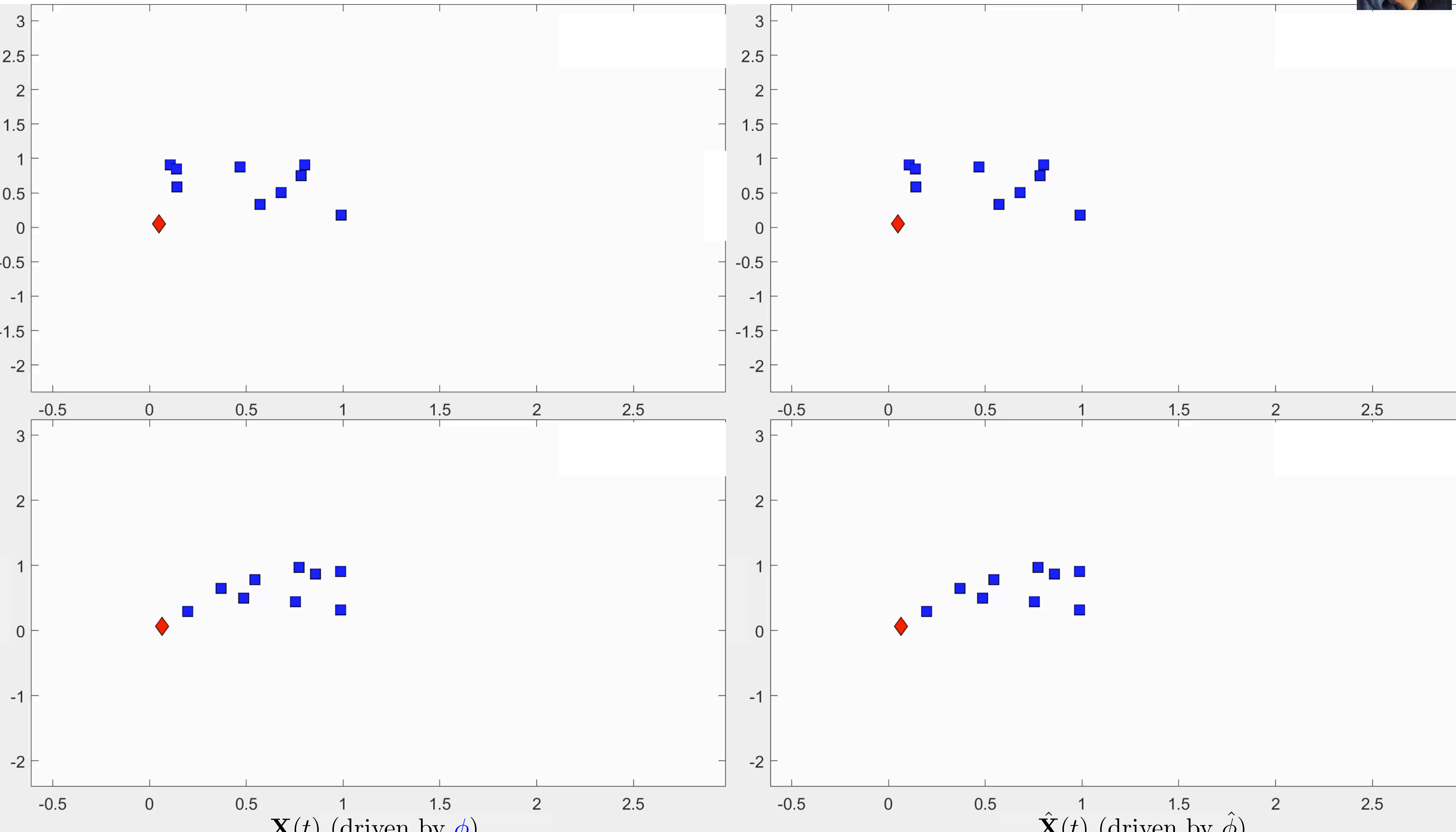
energy and alignment interactions



Example 2nd order Prey-Predator system. Left: the interaction kernels and ρ_L^T's. Right: trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

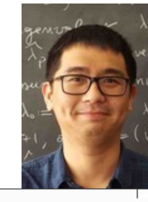
Examples: prey-predator systems

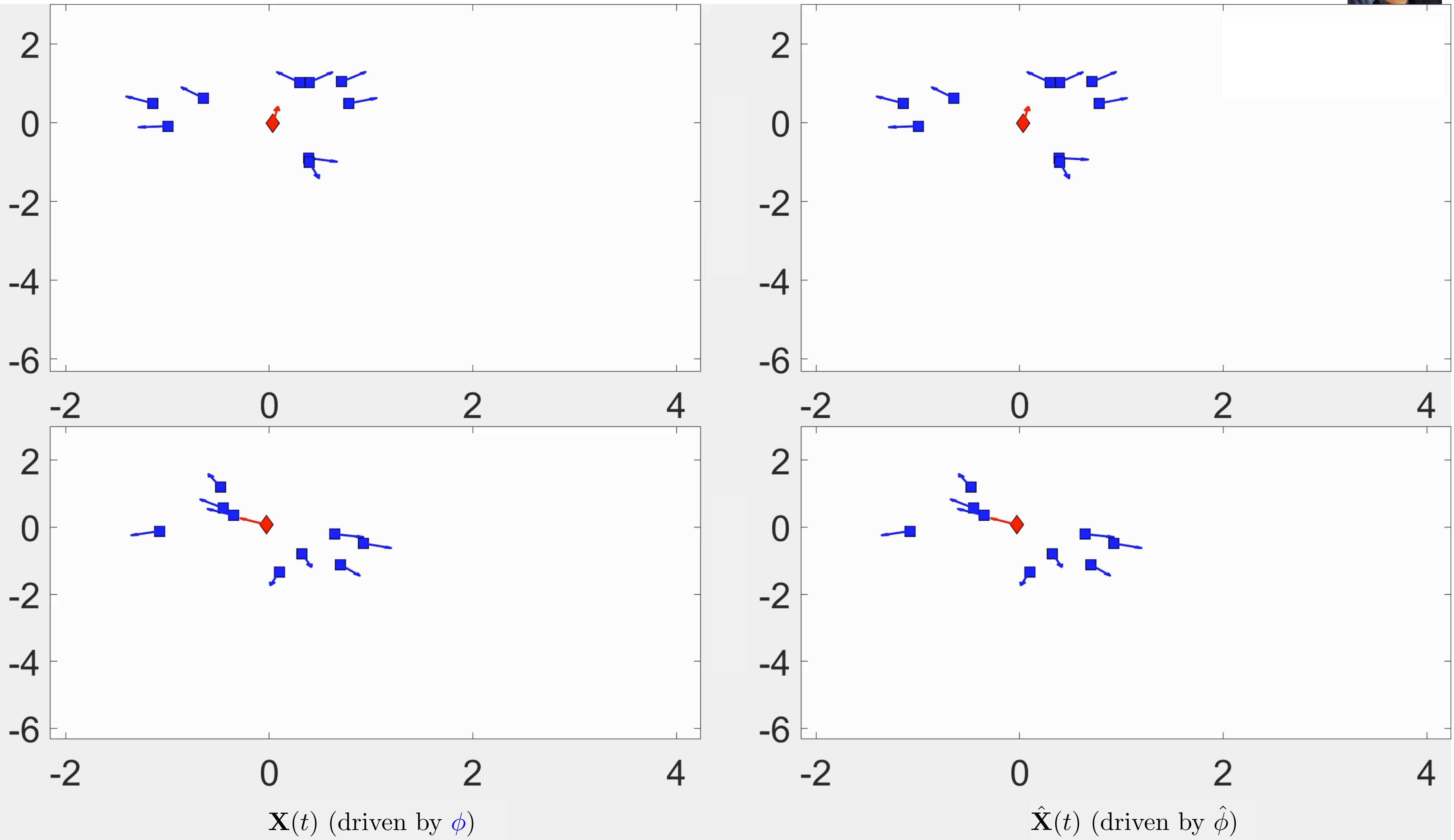




Trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

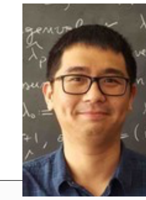
Examples: prey-predator systems

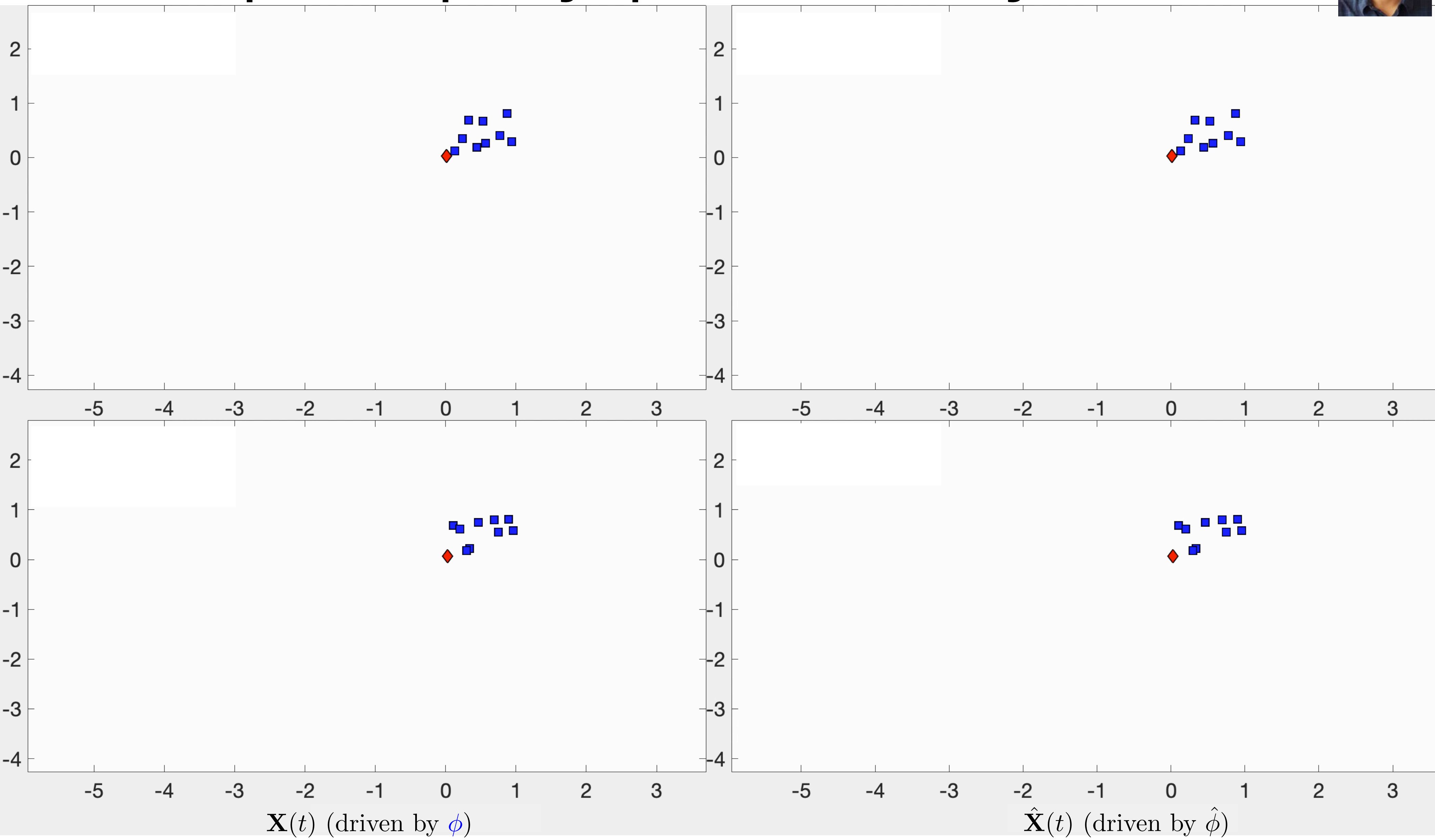




Trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

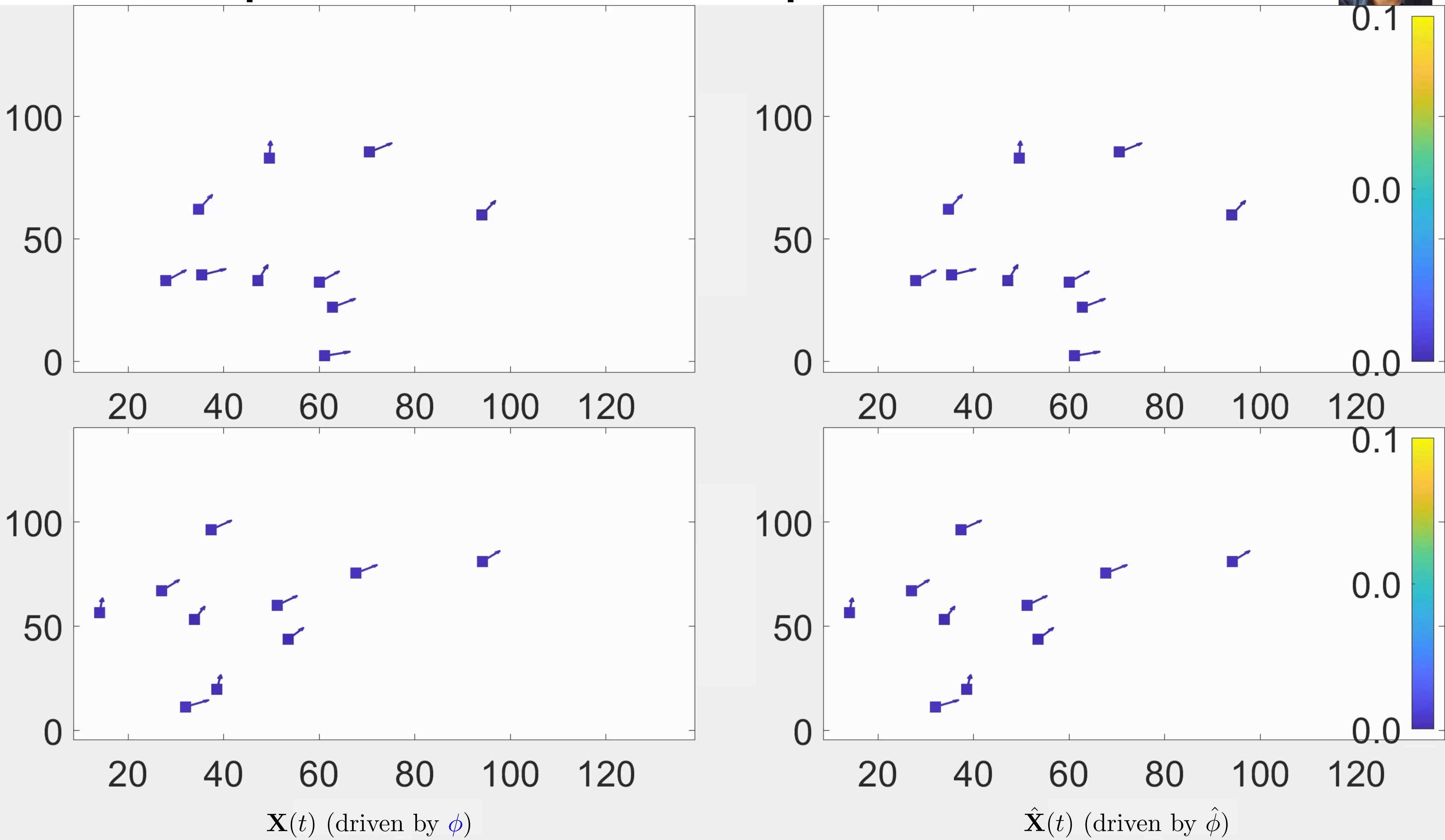
Examples: prey-predator systems





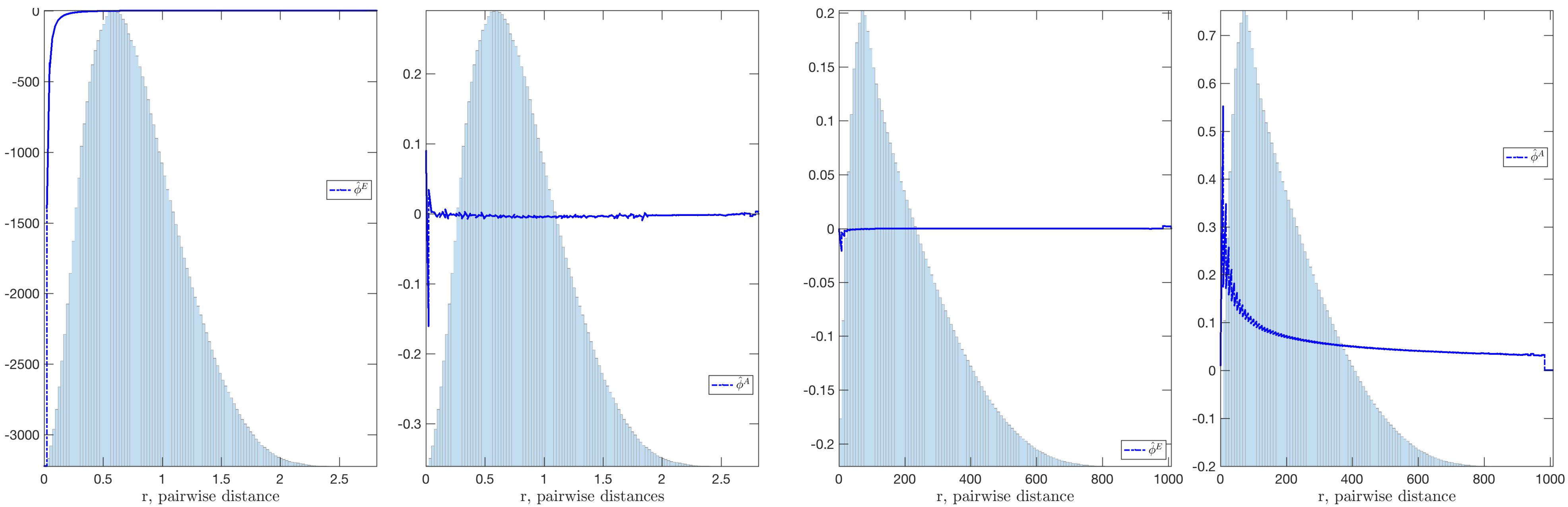
Trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

Examples: model of phototaxis



Trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

Testing hypotheses for agent systems



Example We want to test if a 2nd order system is driven by energy or alignment interactions. Left: we learn a general model (with both types of interaction) on a system with only energy interaction terms: we obtain $\hat{\phi}^A$ is $\approxeq 0$. Right: learning on a system with only alignment term yields $\hat{\phi}^E \approxeq 0$.

Example We want to test if a system is governed by 1st or 2nd order interactions. We are able to tell the difference reliably, by testing the predictions of the learned models on trajectories.

True	Learned as 1^{st} order	Learned as 2^{nd} order
1^{st} order	**0.039 $\pm$ 0.16**	28 ± 21
2^{nd} order	3.1 ± 0.99	**0.58 $\pm$ 0.89**

Emerging behaviors

Ming Zhong,
Jason Miller

Organized collecting stable patterns at large spatial/temporal scale.
Simple, local interaction kernels can learn to complex, organized behavior.
Most of the above is ill-defined, and quotes needed a.e.
Examples include flocking of bird, milling of fish, synchronization in systems of oscillators (neurons, frogs, ...), etc...

In general difficult to characterize and predict; however if robust, we may hope to recover them with systems driven by estimated interaction kernels.

Not only we are often able to recover them in general, but even predict them correctly for each initial condition.

Felix Munoz, https://www.youtube.com/watch?v=OxYn3e_imhA

BBC Blue Planet

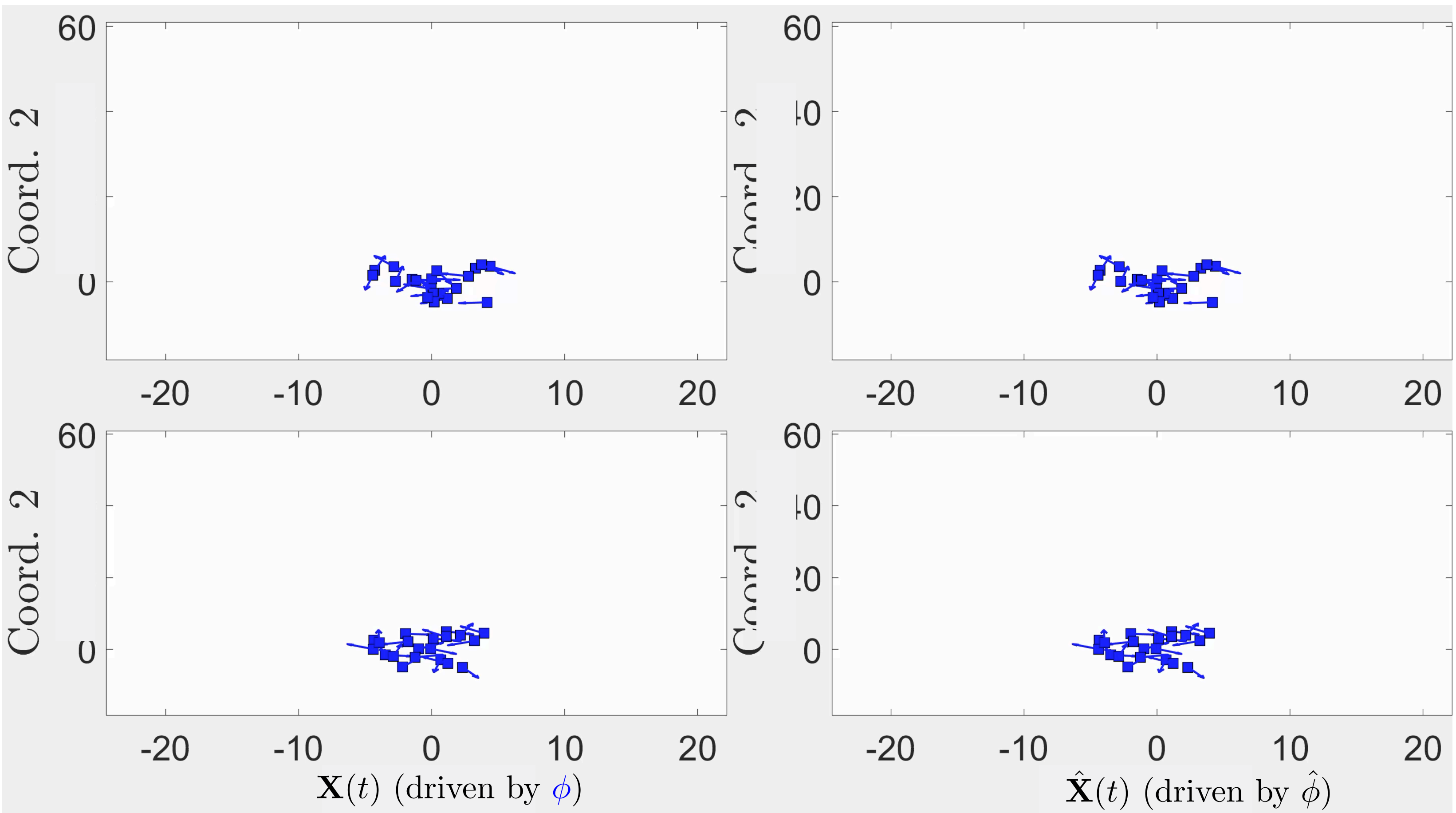
Emerging behaviors: flocking

(*) F. Cucker, J. G. Dong, Avoiding collisions in flocks, IEEE Transactions on Automatic Control, 2010.

The governing equations of Cucker-Smale-Dong (*) dynamics,

$$\ddot{\mathbf{x}}_i = -b_i(t)\dot{\mathbf{x}}_i + \sum_{i'=1}^{N} \left[a_{i,i'}(\mathbf{x})(\dot{\mathbf{x}}_{i'} - \dot{\mathbf{x}}_i) + f(||\mathbf{x}_i - \mathbf{x}_{i'}||^2)(\mathbf{x}_{i'} - \mathbf{x}_i) \right].$$

Here $a_{i,i'}(\mathbf{x}) = H(1 + ||\mathbf{x}_{i'} - \mathbf{x}_i||^2)^{-\beta}$; $b_i : [0,\infty) \to [0,\infty)$ is a bounded and uniformly continuous damping function, and $f : (\delta,\infty) \to [0,\infty)$ is a non-increasing $\mathcal{C}^1$ repulsion function integrable at $+\infty$.



Emerging behaviors: anticipation & flocking

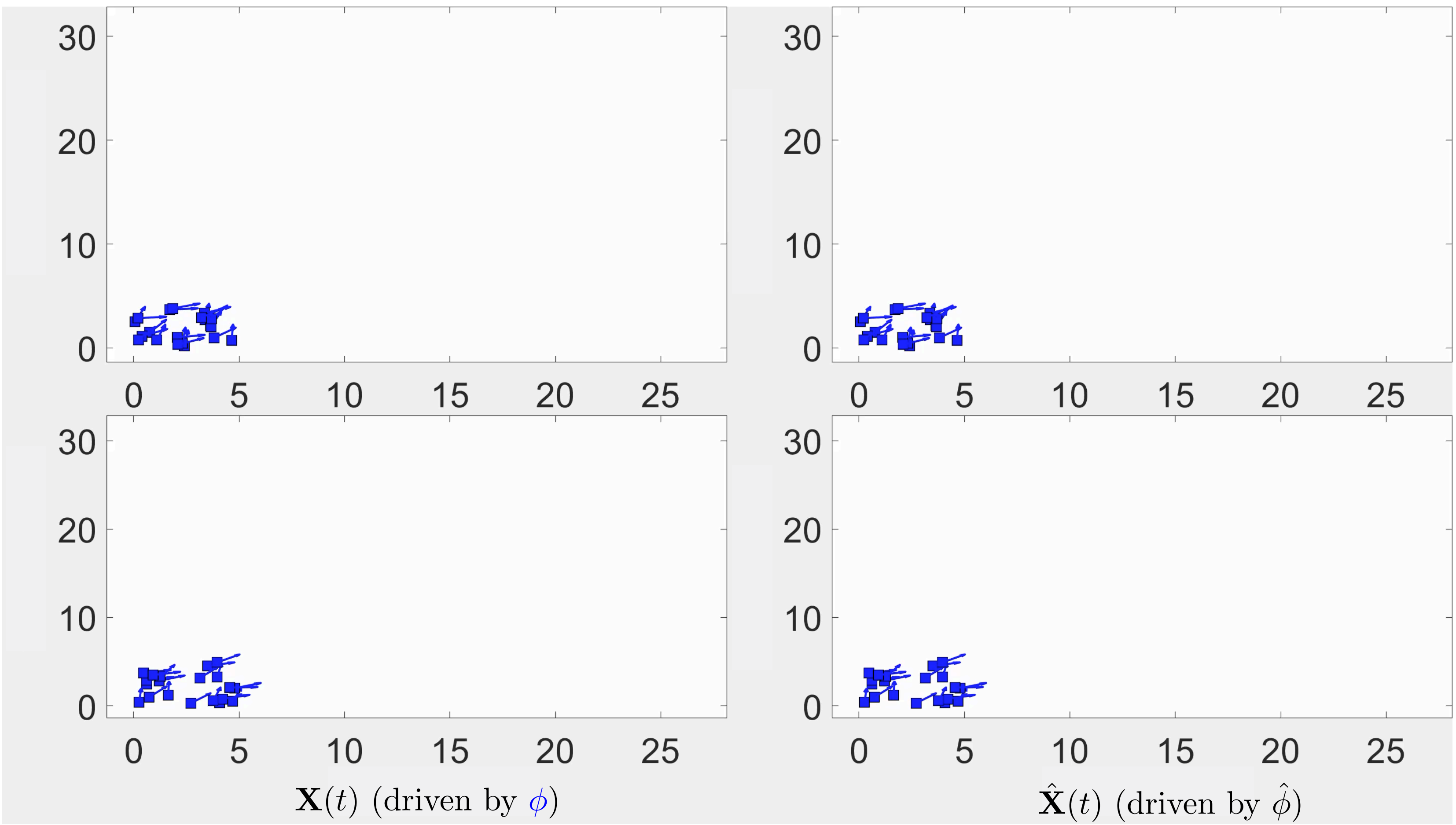
(*) R. Shu and E. Tadmor, Anticipation breeds alignment. arXiv:1905.00633

$$\ddot{\mathbf{x}}_i = \frac{1}{N}\sum_{i'=1,i'\neq i}^{N} \frac{\tau U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)}{||\mathbf{x}_{i'}-\mathbf{x}_i||}(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)$$

$$+\frac{1}{N}\sum_{i'=1,i'\neq i}^{N}\Big\{\frac{-\tau U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)(\mathbf{x}_{i'}-\mathbf{x}_i)\cdot(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)}{||\mathbf{x}_{i'}-\mathbf{x}_i||^3}$$

$$+\frac{\tau U''(||\mathbf{x}_{i'}-\mathbf{x}_i||)(\mathbf{x}_{i'}-\mathbf{x}_i)\cdot(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)}{||\mathbf{x}_{i'}-\mathbf{x}_i||^2}+\frac{U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)}{||\mathbf{x}_{i'}-\mathbf{x}_i||}\Big\}(\mathbf{x}_{i'}-\mathbf{x}_i).$$

$$U(r) = r^{1.5}/1.5$$



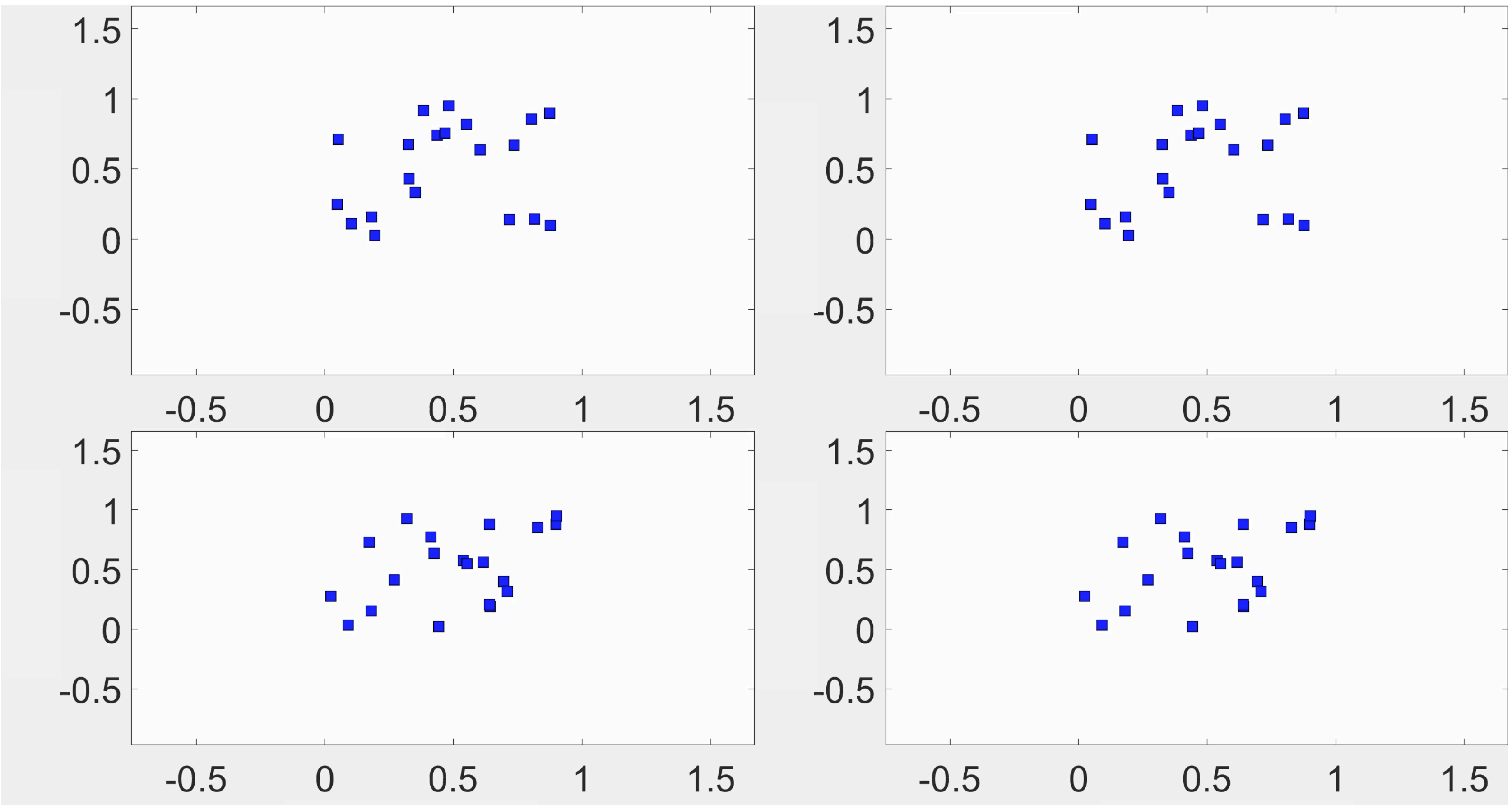
Emerging behaviors: Fish mill patterns

The governing equations of fish milling dynamics in $\mathbb{R}^2$ of (*) are

$$m_i\ddot{\mathbf{x}}_i = \alpha\dot{\mathbf{x}}_i - \beta||\dot{\mathbf{x}}_i||^2\dot{\mathbf{x}}_i - \vec{\nabla}U_i\,,$$

(*) Y. Li Chuang, M. R. D'Orsogna, D. Marthaler, A. L. Bertozzi, L. S. Chayes, Physica D: Nonlinear Phenomena 232 (2007)

with U_i a potential for the interaction of the i^{th} agent with the other agents: $U_i = \sum_{i'=1}^{N}(-C_a \mathrm{e}^{-||\mathbf{x}_i - \mathbf{x}_{i'}||/\ell_a} + C_r \mathrm{e}^{-||\mathbf{x}_i - \mathbf{x}_{i'}||/\ell_r})$.



$\mathbf{X}(t)$ (driven by ϕ) $\hat{\mathbf{X}}(t)$ (driven by $\hat{\phi}$)

Parametrized Families of Interaction Kernels

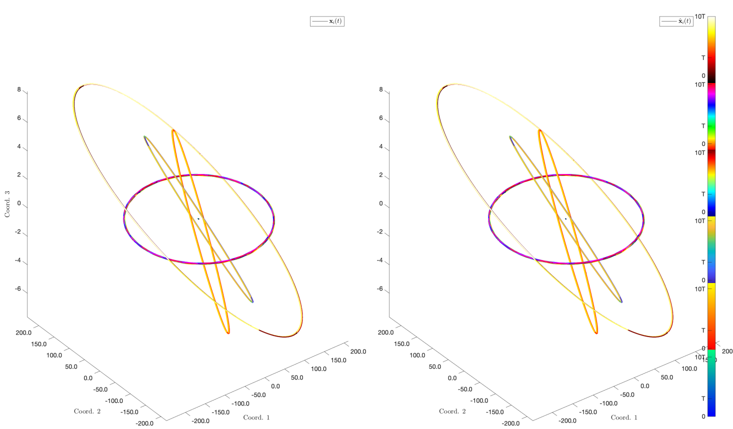
We consider the 5 innermost planets and the Sun in our solar system, moving according to Newton's law:

$$\ddot{\mathbf{x}}_i(t) = \sum_{i'=1}^{N} \frac{G\tilde{m}_{i'}}{||\mathbf{x}_{i'} - \mathbf{x}_i||^3} (\mathbf{x}_{i'} - \mathbf{x}_i) \,,$$

where $\tilde{m}_i$ are the unknown masses, and G is the known gravitational constant.

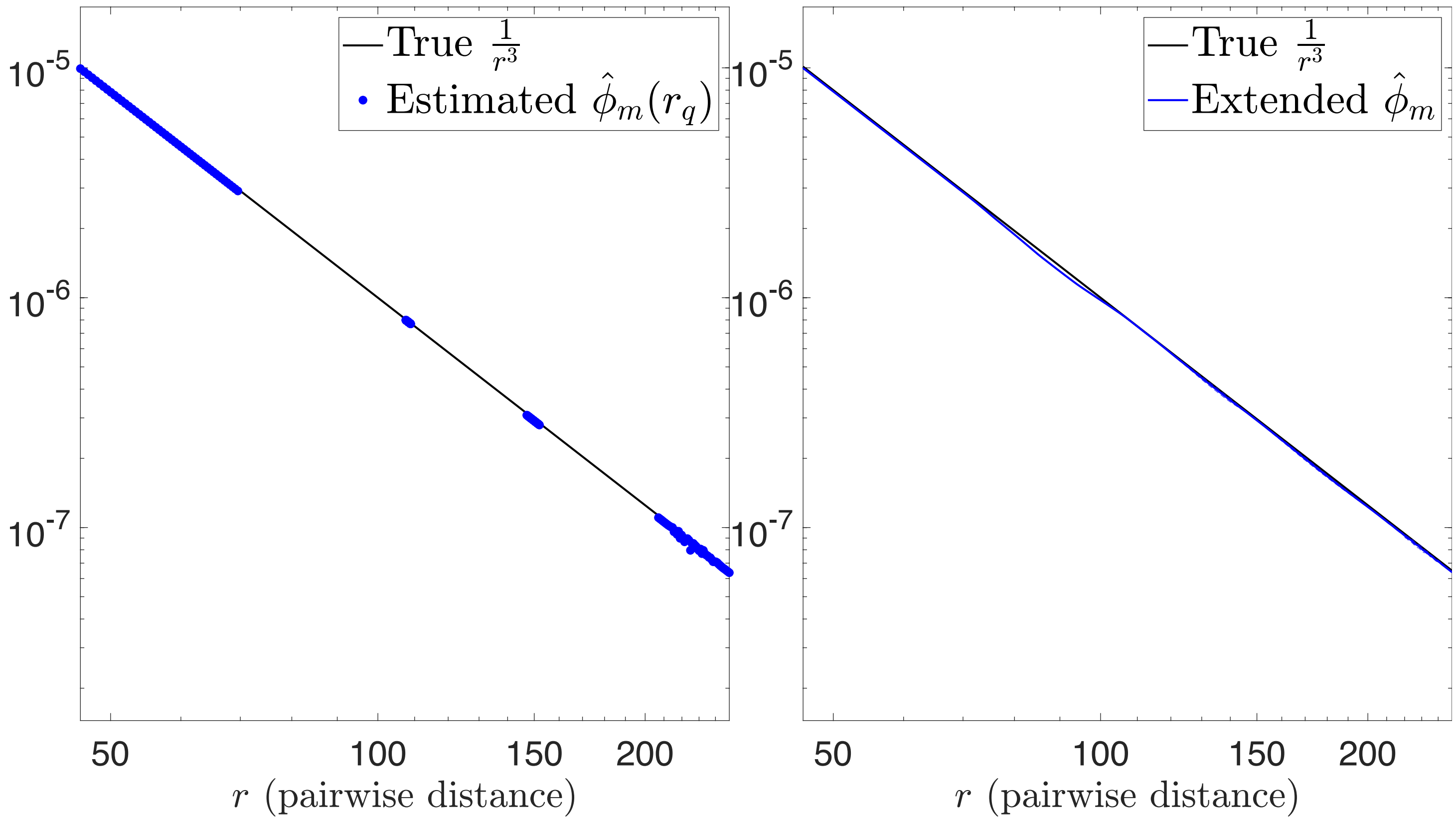
We do not know that the agents follow a common law, modulo parameters (mass), so we treat all of them as being of different type, with interactions $\phi_{ii'}$.

We will then discover the relationships between these interaction kernels, namely that they are all multiple of $1/r^3$, and the multiple is mass.



Data-driven Discovery of Emergent Behaviors in Collective Dynamics,
Physica D: Nonlinear Phenomena, 2019

Discovering the parameters



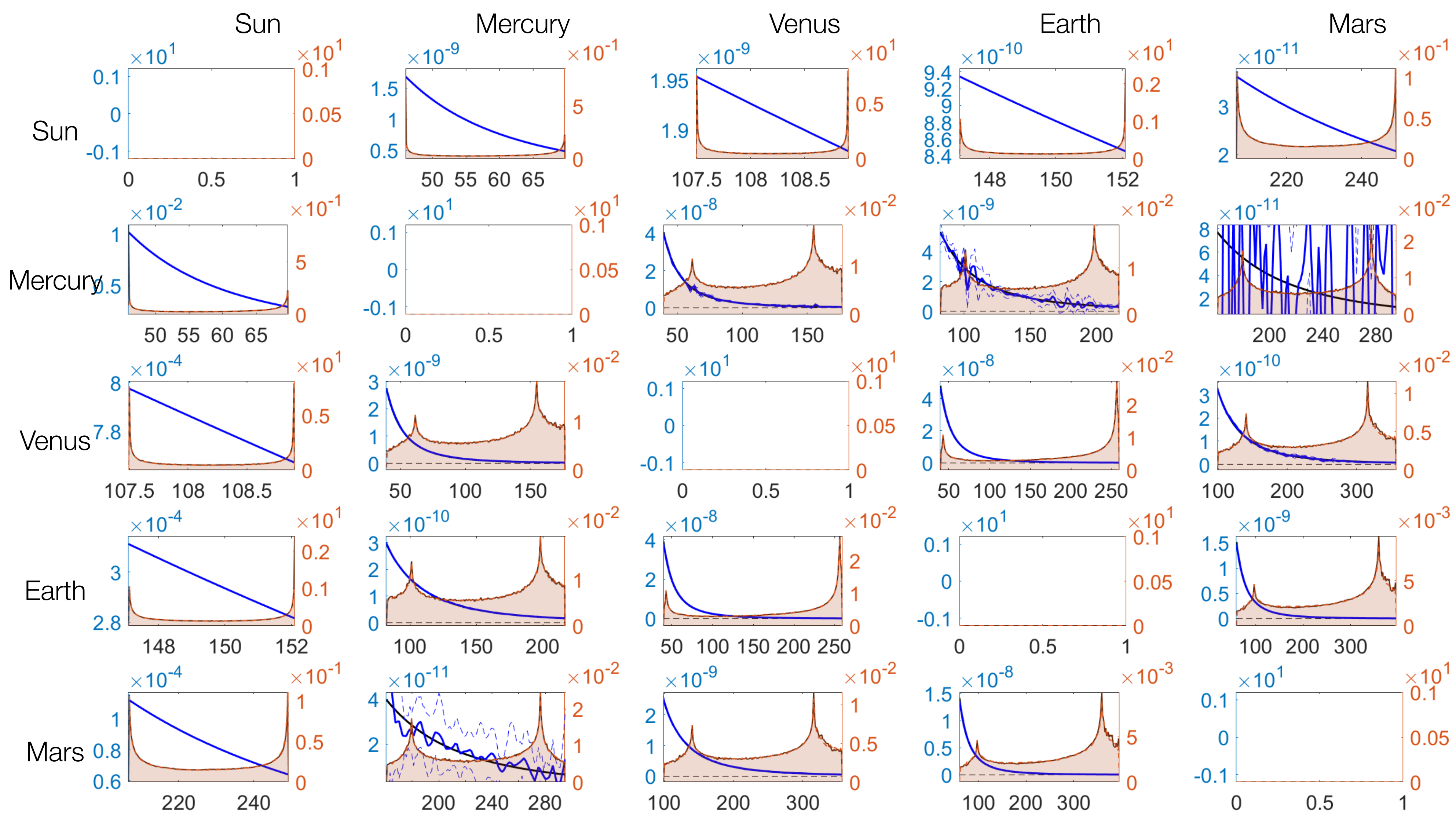
We are able to estimate to identify, in a fully nonparametric fashion, $1/r^3$.

We also estimate masses quite accurately.

	Sun	Mercury	Venus	Earth	Mars
True Mass	$1.9885 \cdot 10^6$	$3.3 \cdot 10^{-1}$	4.87	5.97	$6.42 \cdot 10^{-1}$
Estimated Mass	$2.01 \cdot 10^6 \pm 1 \cdot 10^4$	$3.35 \cdot 10^{-1} \pm 2 \cdot 10^{-3}$	$4.88 \pm 3 \cdot 10^{-2}$	$6.05 \pm 4 \cdot 10^{-2}$	$6.52 \cdot 10^{-1} \pm 5 \cdot 10^{-3}$
Rel. Err.	$1.1 \cdot 10^{-2} \pm 7 \cdot 10^{-3}$	$1.4 \cdot 10^{-2} \pm 6 \cdot 10^{-3}$	$4 \cdot 10^{-3} \pm 4 \cdot 10^{-3}$	$1.3 \cdot 10^{-2} \pm 6 \cdot 10^{-3}$	$1.6 \cdot 10^{-2} \pm 8 \cdot 10^{-3}$

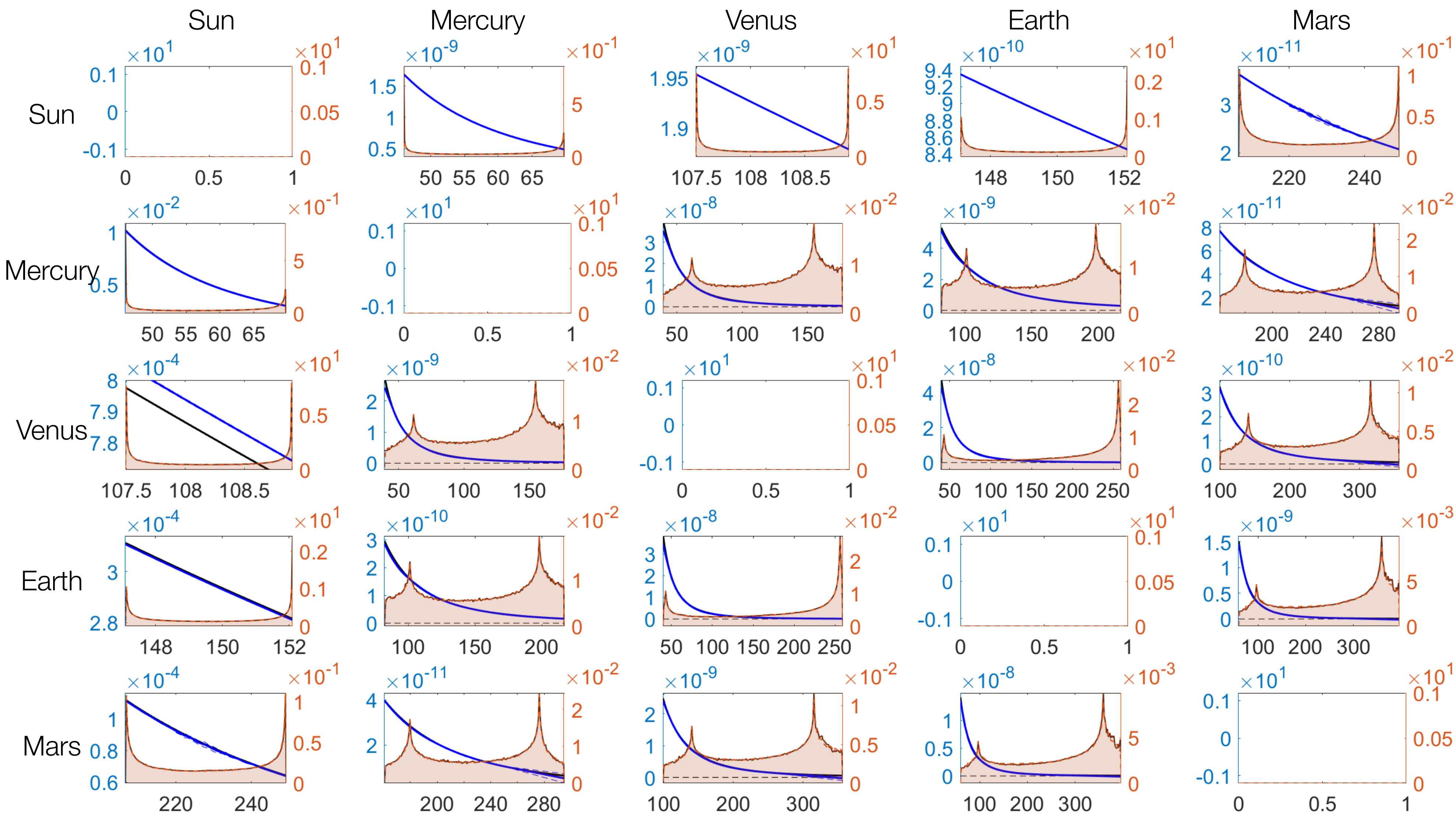
Gravity Kernels - learned independently

Estimated interaction kernels between Sun, Mercury, Venus, Earth and Mars, $\hat{\phi}_{ii'}$, as a function of pairwise distance.



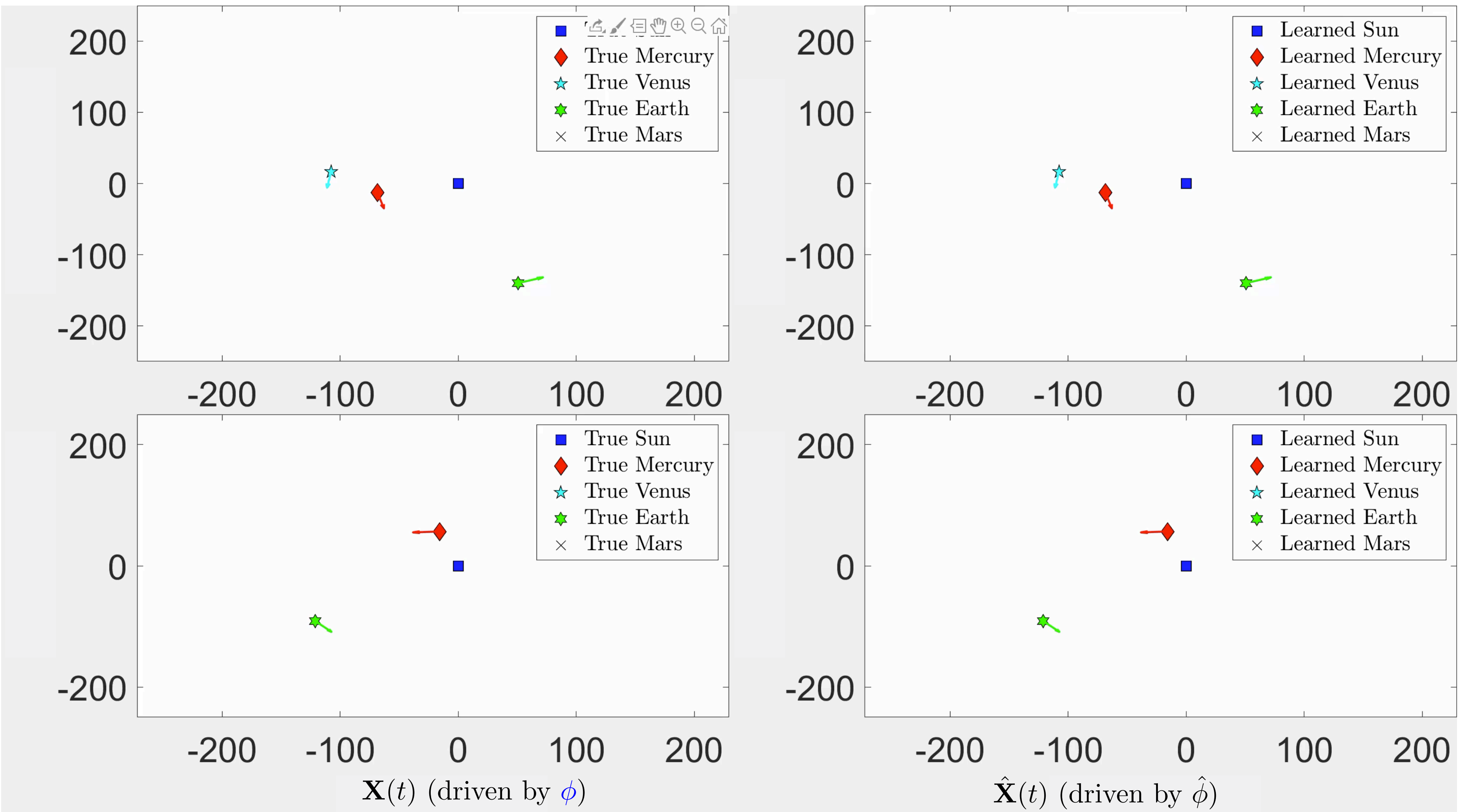
Gravity Kernels, re-estimated

Estimated interaction kernels between Sun, Mercury, Venus, Earth and Mars, $\hat{\phi}_{ii'}$, as a function of pairwise distance.



Trajectories from gravity kernels

Trajectory predictions are accurate (relative errors $O(10^{-4})$ with $M = 500$ and daily observations for about 6 months), exhibit stability, and approximate energy conservation.

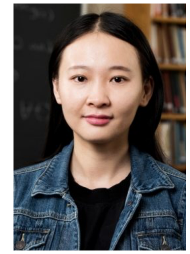


Trajectories from gravity kernels

Trajectory predictions are accurate (relative errors $O(10^{-4})$ with $M = 500$ and daily observations for about $T = 6$ months), exhibit stability, and approximate energy conservation, even for longer times (below, $T_f \sim 2.5$yrs).

	$[0, T]$	$[T, T_f]$
mean_{IC}: Training ICs on $\boldsymbol{x}$	$6.6 \cdot 10^{-4} \pm 2 \cdot 10^{-5}$	$3.9 \cdot 10^{-3} \pm 2 \cdot 10^{-4}$
mean_{IC}: Training ICs on $\boldsymbol{v}$	$3.9 \cdot 10^{-3} \pm 1 \cdot 10^{-4}$	$2.13 \cdot 10^{-2} \pm 8 \cdot 10^{-4}$
std_{IC}: Training ICs on $\boldsymbol{x}$	$5 \cdot 10^{-4} \pm 1 \cdot 10^{-4}$	$2.7 \cdot 10^{-3} \pm 4 \cdot 10^{-4}$
std_{IC}: Training ICs on $\boldsymbol{v}$	$2.5 \cdot 10^{-3} \pm 3 \cdot 10^{-4}$	$1.30 \cdot 10^{-2} \pm 2 \cdot 10^{-4}$
mean_{IC}: Random ICs on $\boldsymbol{x}$	$6.8 \cdot 10^{-4} \pm 2 \cdot 10^{-5}$	$3.9 \cdot 10^{-3} \pm 1 \cdot 10^{-4}$
mean_{IC}: Random ICs on $\boldsymbol{v}$	$3.9 \cdot 10^{-3} \pm 1 \cdot 10^{-4}$	$2.13 \cdot 10^{-2} \pm 6 \cdot 10^{-4}$
mean_{IC}: Random ICs on $\boldsymbol{x}$	$5.3 \cdot 10^{-4} \pm 1 \cdot 10^{-4}$	$2.5 \cdot 10^{-3} \pm 3 \cdot 10^{-4}$
std_{IC}: Random ICs on $\boldsymbol{v}$	$2.6 \cdot 10^{-3} \pm 4 \cdot 10^{-4}$	$1.2 \cdot 10^{-2} \pm 1 \cdot 10^{-3}$

Generalizations



The theory has been extended to:

· multi-type agent systems;

· second-order systems;

· possible interactions with simple environments;

· parametrized families of interaction kernels;

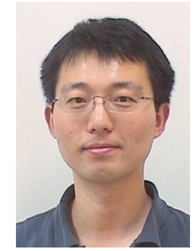
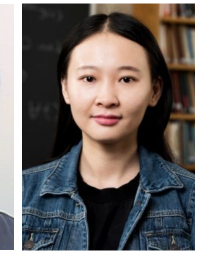
· interaction kernels depending on multiple variables (beyond pairwise distances).

With suitable definitions and assumptions, the results we obtain appear optimal, with a rate depending on the total number v of distinct variables inside the interaction kernels, over all pairs of agent types, e.g.:

$$\mathbb{E}[\|\widehat{\phi}_{L,M,\mathcal{H}_{n_*}}(\cdot)\cdot - \phi(\cdot)\cdot\|_{L^2(\rho_L^T)}] \leq \frac{C}{c_{L,N,\mathcal{H}}}\left(\frac{\log M}{M}\right)^{\frac{s}{2s+v}} .$$

This is joint work with J. Miller, M. Zhong, S. Tang (preprint appearing shortly).

The Stochastic case

We have also generalized these results to the **stochastic** case

$$d\mathbf{x}_{i,t} = \frac{1}{N}\sum_{i'=1}^{N} \phi(\|\mathbf{x}_{j,t} - \mathbf{x}_{i,t}\|)(\mathbf{x}_{j,t} - \mathbf{x}_{i,t})dt + \sigma d\mathbf{B}_{i,t}\,.$$

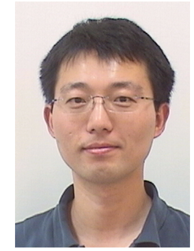
This is joint work with F. Lu and S. Tang, *Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories*, arXiv, 2020. We are currently studying convergence in the stochastic ergodic case as $T \to \infty$.

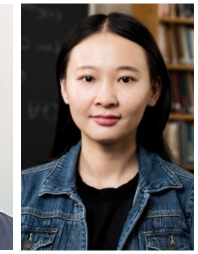
Note that in the stochastic case we do not (cannot!) observe velocities, but only positions. We have studied carefully the dependence on the observation time gap $\Delta t := t_{l+1} - t_l = T/L$:

$$||\hat{\phi}_{L,T,M,\mathcal{H}} - \phi||_{L^2(\rho_T)} \leq ||\hat{\phi}_{T,\infty,\mathcal{H}} - \phi||_{L^2(\rho_T)} + C\left(\sqrt{\frac{n}{M}}\epsilon + \sqrt{\frac{T}{L}}\right),$$

where $\hat{\phi}_{T,\infty,\mathcal{H}}$ is the projection of the true kernel ϕ onto $\mathcal{H}$.

The Stochastic case



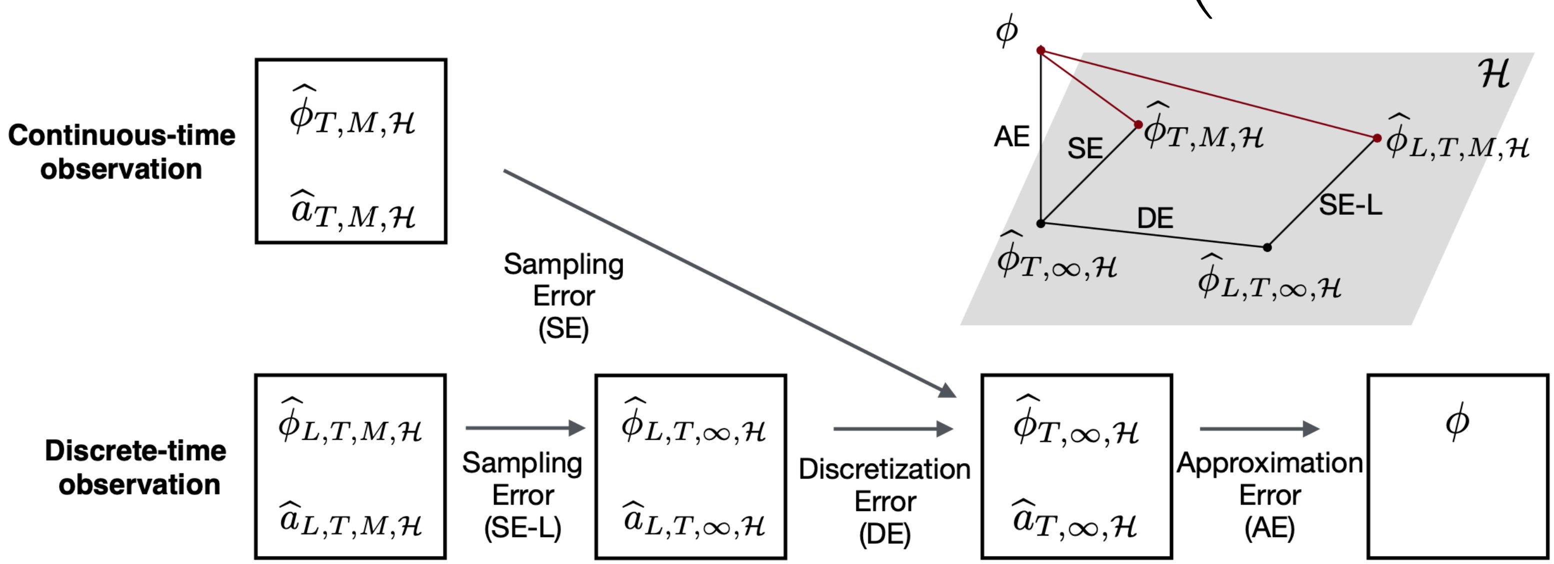


We have also generalized these results to the **stochastic** case

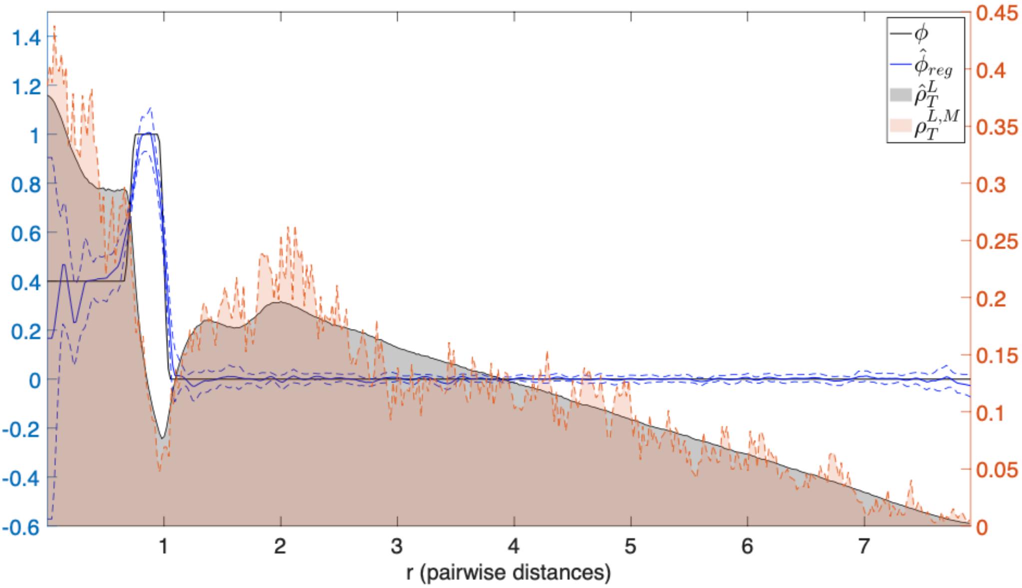
$$d\mathbf{x}_{i,t} = \frac{1}{N}\sum_{i'=1}^{N} \phi(\|\mathbf{x}_{j,t} - \mathbf{x}_{i,t}\|)(\mathbf{x}_{j,t} - \mathbf{x}_{i,t})dt + \sigma d\mathbf{B}_{i,t}\,.$$

Note that in the stochastic case we do not (cannot!) observe velocities, but only positions. We have studied carefully the dependence on the observation time gap $\Delta t := t_{l+1} - t_l = T/L$:

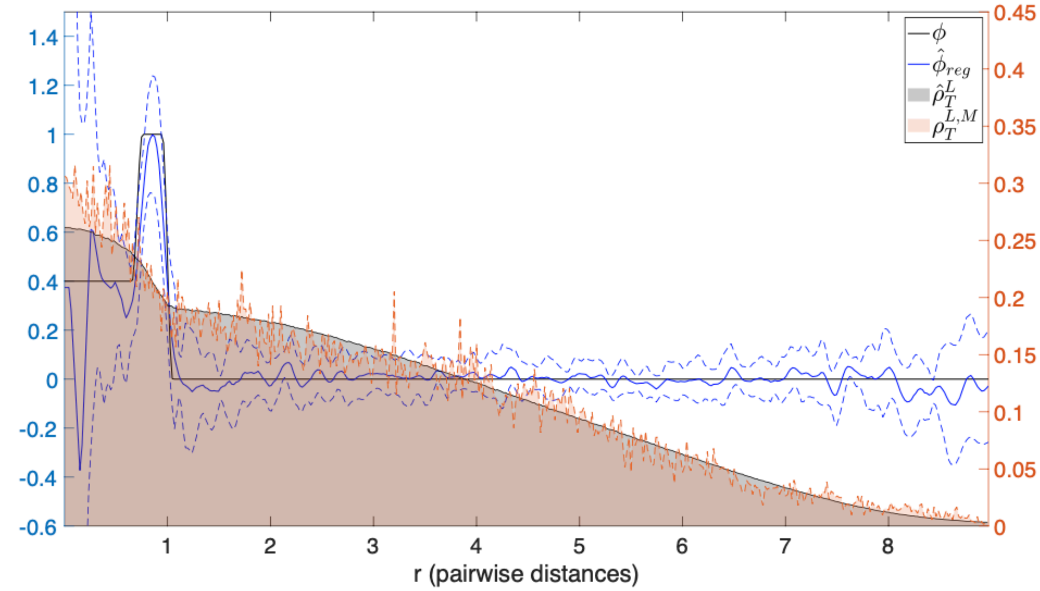
$$||\hat{\phi}_{L,T,M,\mathcal{H}} - \phi||_{L^2(\rho_T)} \leq ||\hat{\phi}_{T,\infty,\mathcal{H}} - \phi||_{L^2(\rho_T)} + C\left(\sqrt{\frac{n}{M}}\epsilon + \sqrt{\frac{T}{L}}\right)\,.$$



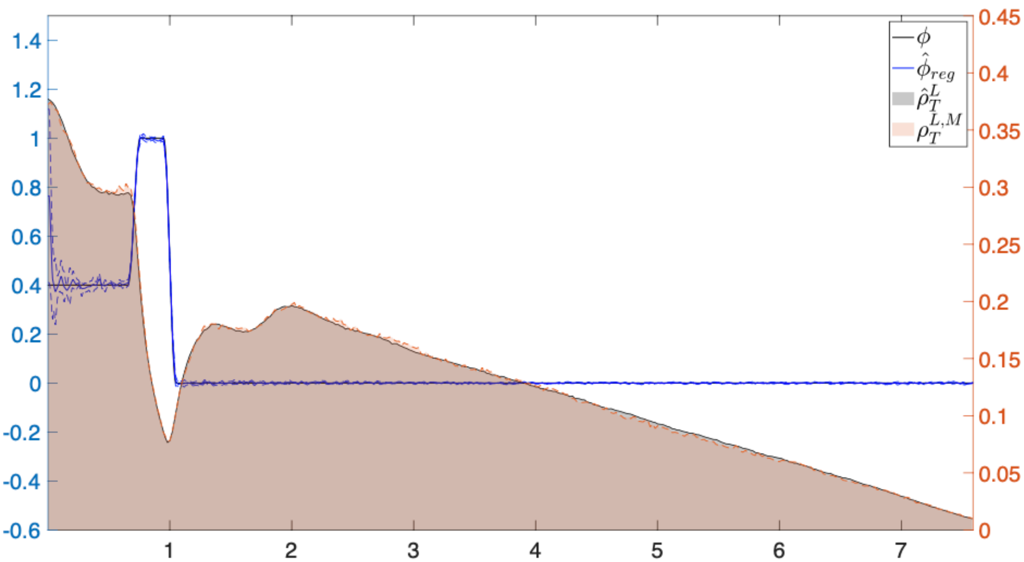
Stochastic Opinion Dynamics



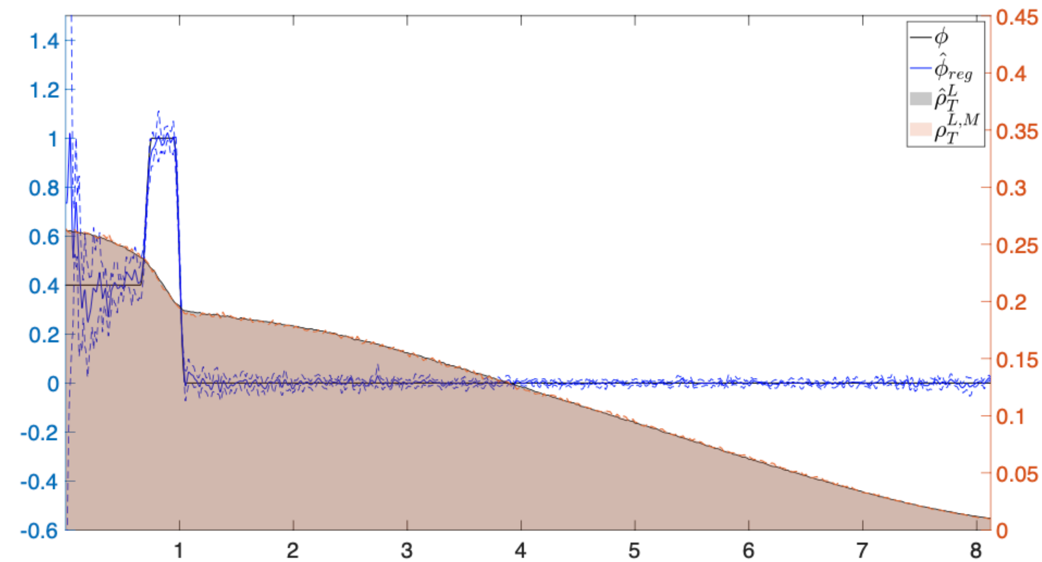
(a) $\sigma = 0.1, M = 32$



(b) $\sigma = 0.5, M = 32$

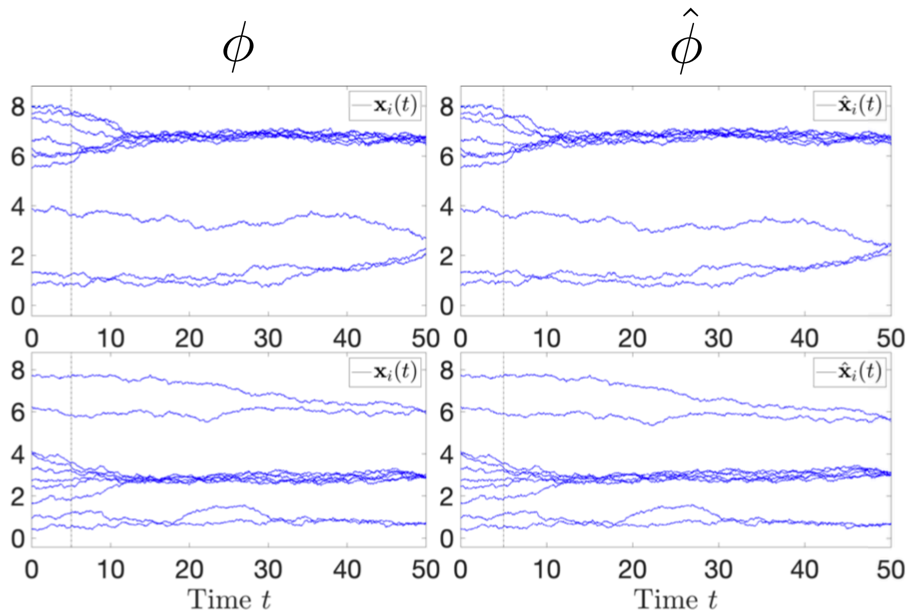


(c) $\sigma = 0.1, M = 4096$

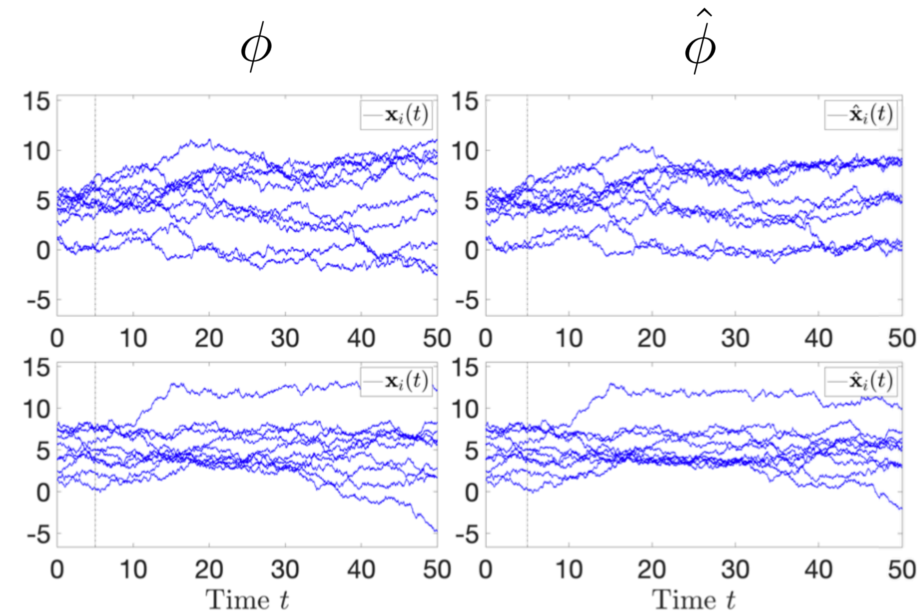


(d) $\sigma = 0.5, M = 4096$

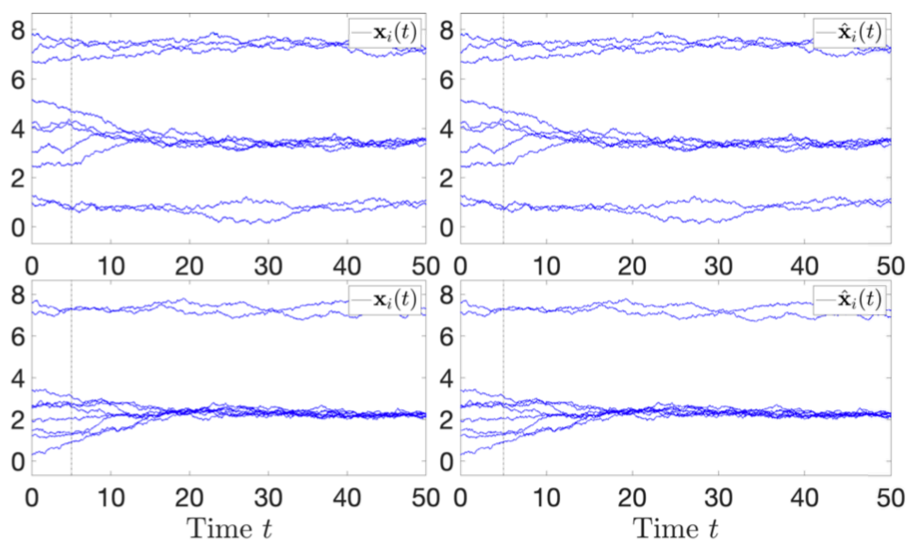
Stochastic Opinion Dynamics



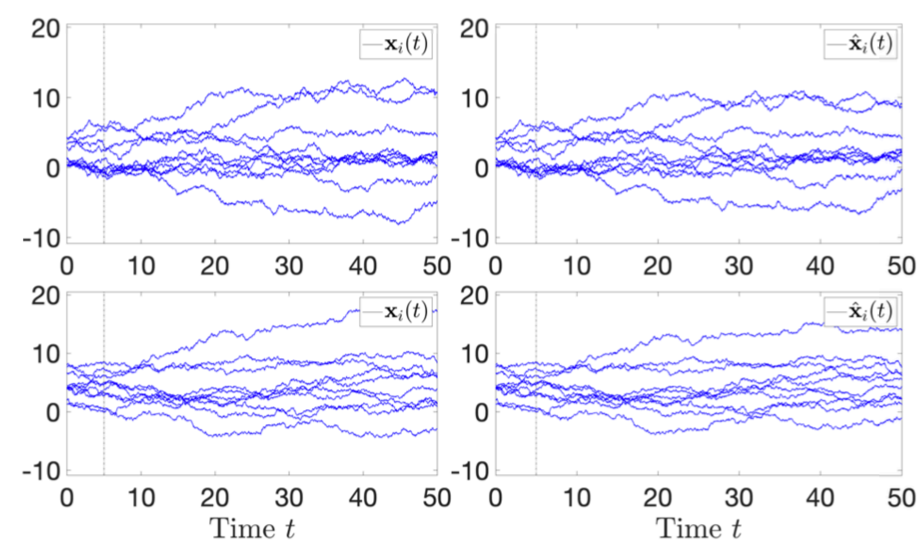
(a) $M = 32, \sigma = 0.1$



(b) $M = 32, \sigma = 0.5$

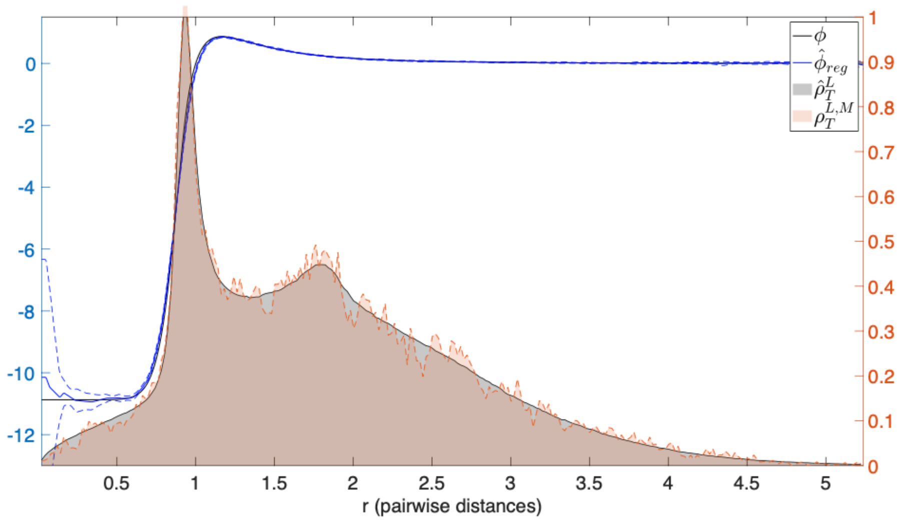


(c) $M = 4096, \sigma = 0.1$

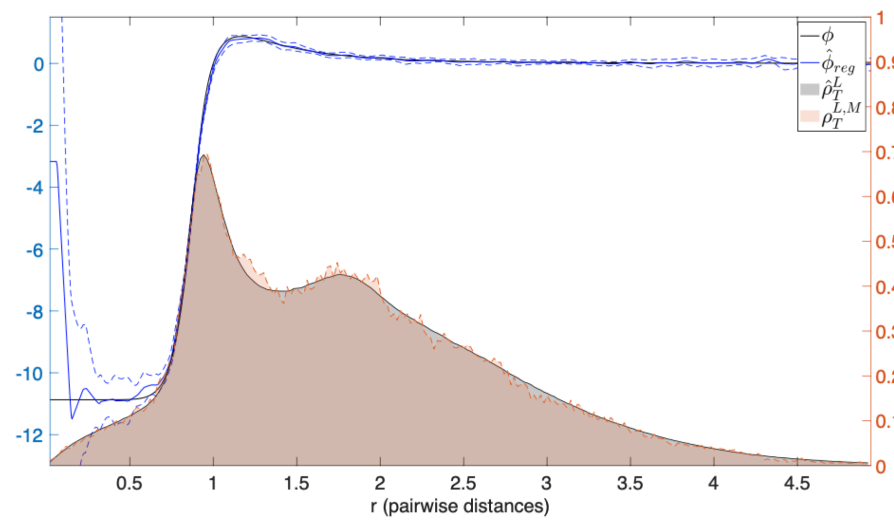


(d) $M = 4096, \sigma = 0.5$

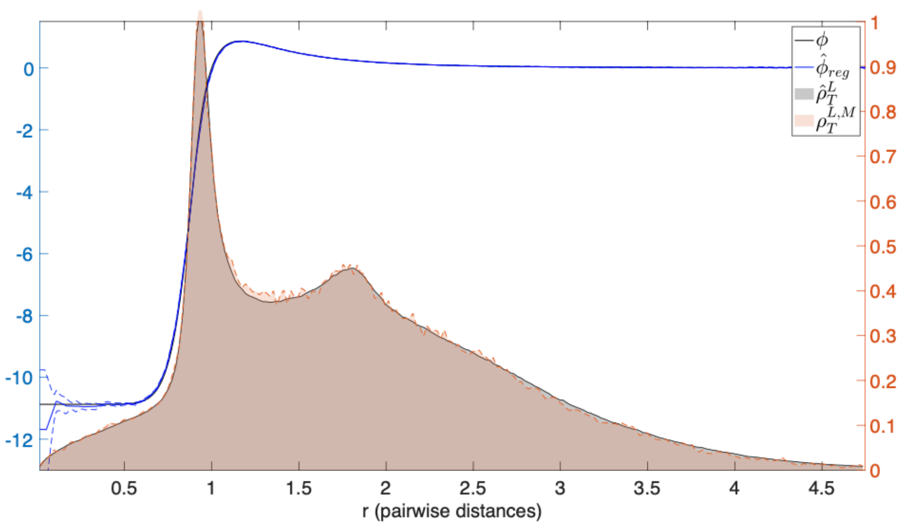
Stochastic Lennard-Jones



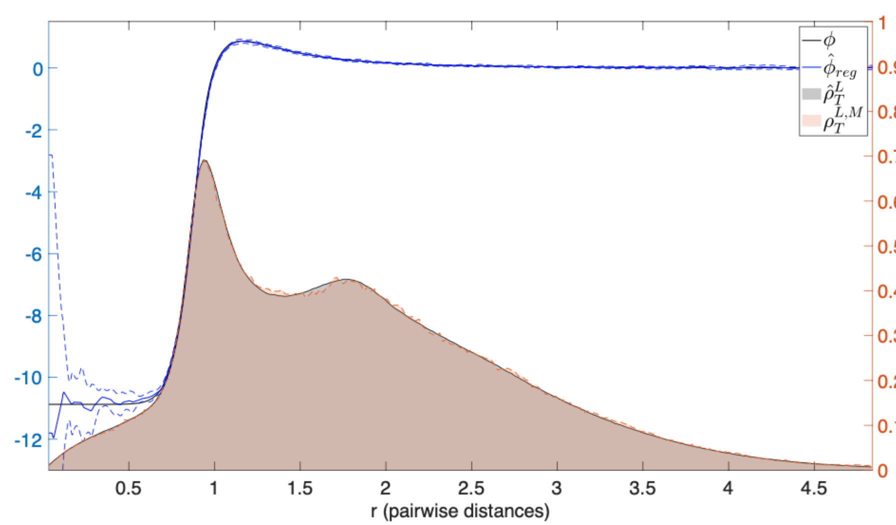
(a) $\sigma = 0.05, M = 128$



(b) $\sigma = 0.25, M = 128$

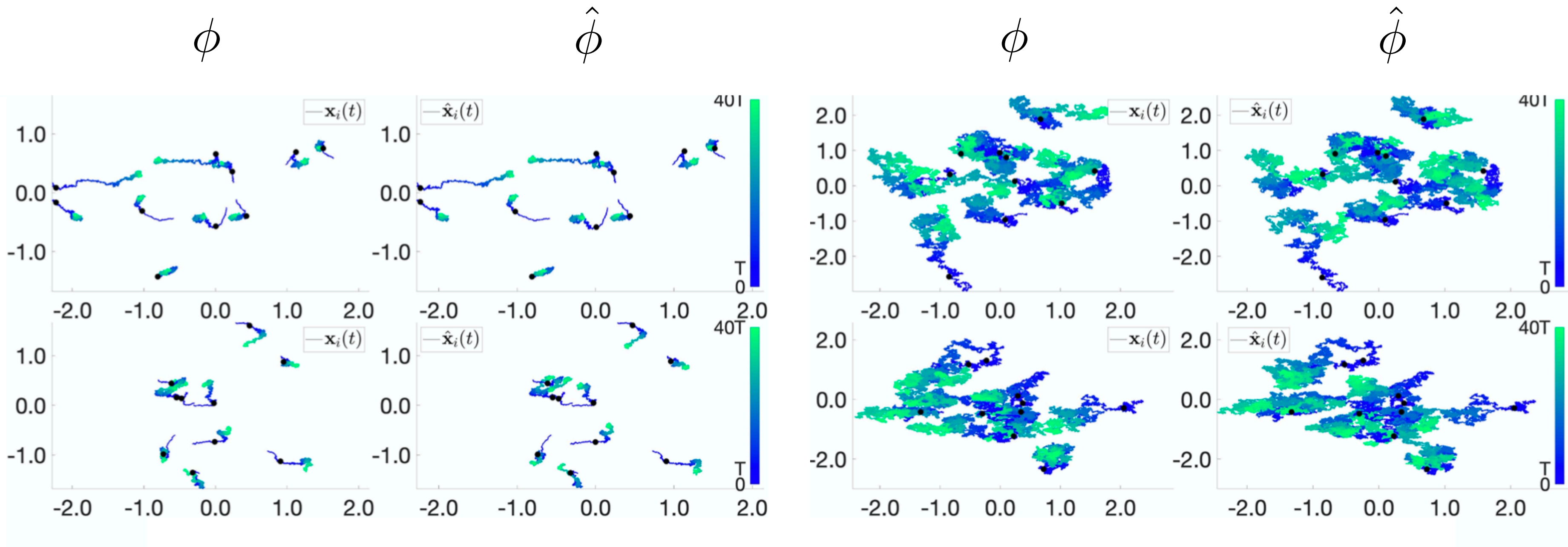


(c) $\sigma = 0.05, M = 1024$

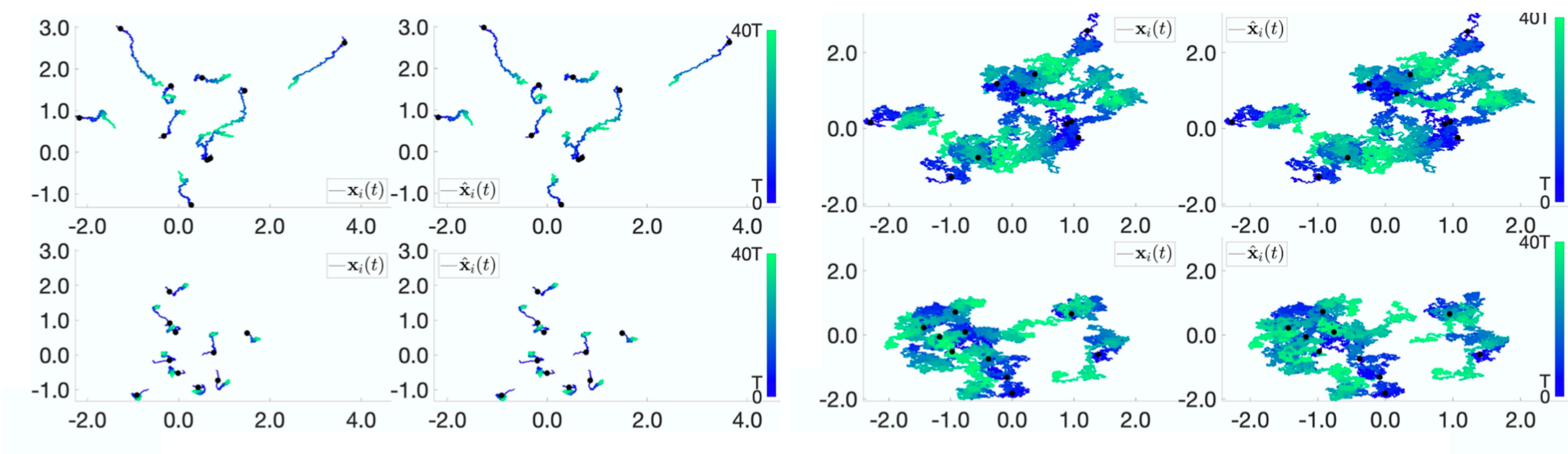


(d) $\sigma = 0.25, M = 1024$

Stochastic Lennard-Jones

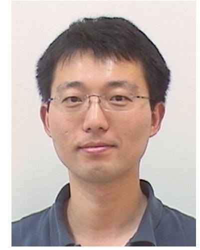
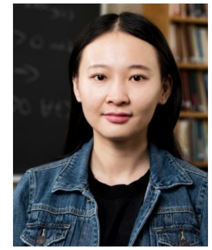
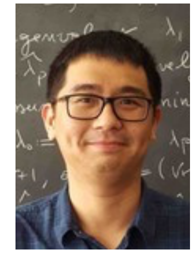
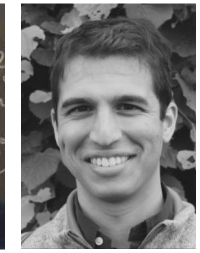


(a) $\sigma = 0.05, M = 128$ (b) $\sigma = 0.25, M = 128$



(c) $\sigma = 0.05, M = 1024$ (d) $\sigma = 0.25, M = 1024$

Conclusions

· Learning agent-based type system may be performed efficiently, nonparametrically, at least in special cases, notwithstanding the high-dimensional state space.
· Important generalizations: 1st- and 2nd-order, multi-type; more general interaction kernels.
· Hypothesis testing; transfer learning; dictionary learning for dynamical systems.
· Many open problems. E.g.: quantifying information needed for learning; stochasticity; hidden variables; general interaction kernels; ...
· Many applications: biological systems, particle systems, learning forces in molecular systems, ...

Nonparametric inference of interaction laws in systems of agents from trajectory data,
F. Lu, S. Tang, M. Zhong, MM, P.N.A.S., July 2019.

Data-driven Discovery of Emergent Behaviors in Collective Dynamics,
MM, J.Miller, M. Zhong, Physica D, 2020.

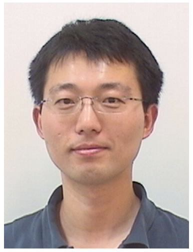
Learning interaction kernels in heterogeneous systems of agents from multiple trajectories,
F. Lu, MM, S. Tang, ArXiv 1910.04832, submitted.

Learning theory for inferring interaction kernels in second-order interacting agent systems,
J.Miller, M. Zhong, S. Tang, M. Maggioni, in preparation.

Links to code, papers: https://mauromaggioni.duckdns.org

AFOSR NSF
Bloomberg Philanthropies
SIMONS FOUNDATION

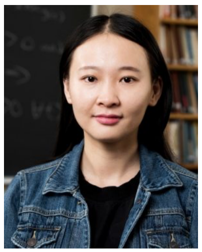
Team & Collaborators



Fei Lu, Asst. Prof.
Math, JHU.
Stochastics; learning; dynamics.

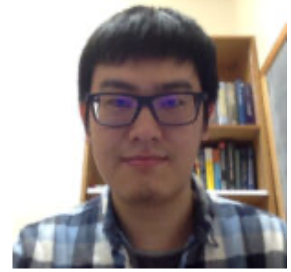
X. Felix Ye, Postdoc, Applied Math & Stats, JHU.
Stochastic dynamics; learning.

Marie-Jose Kuffner, Postdoc, Math, JHU.
Harmonic analysis; optimal transport.

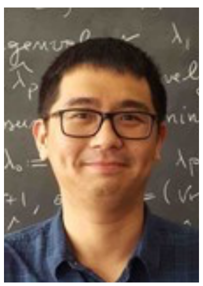


Sui Tang, Asst. Prof., Math, UCSB.
Learning; harmonic analysis.

Christian Kümmerle, Postdoc, Applied Math & Stats, JHU.
Optimization; learning; optimal transport.



Jinchao Feng, Postdoc, Applied Math & Stats, JHU.
Learning; optimization.

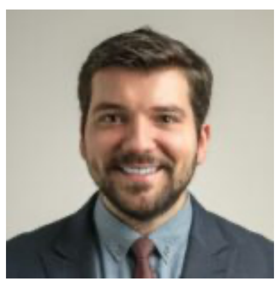


Ming Zhong, Postdoc, Applied Math & Stats, JHU.
Learning; dynamics.

Patrick Martin, Ph.D. Student, Math, JHU.
Learning; regression In high dimensions.

Sichen Yang, Ph.D. Student Applied Math & Stats, JHU.
Learning; PDEs; stochastics.

Jason Miller, Ph.D. Student, Applied Math & Stats, JHU.
Learning; dynamics.



Dan Popescu, Ph.D. Student Applied Math & Stats, JHU.
Neural networks; medical imaging; PDEs.

Links to code, papers: https://mauromaggioni.duckdns.org

AFOSR NSF Bloomberg Philanthropies
SIMONS FOUNDATION

If you have trouble with my web page…

Links to code, papers: https://mauromaggioni.duckdns.org

I not responsible for the contents of any other website, even those pretending to represent me or my work.

Certain internet providers censor all *.duckdns.org domains; if you wish to decide which websites to access instead of leaving this decision to your internet provider, you may consider to:

1) *change your DNS server* (e.g. to Cloudfare's 1.1.1.1); (nx)or
2) *use a reputable VPN service*; (nx)or
3) *use the (free) Tor browser*, which achieves 1+2, but only for the traffic associated to the Tor browser, unless you route all your traffic by installing and activating a Tor client (e.g. *brew install tor* on MacOS with *Homebrew*, or *yum install tor* on CentOS, etc…).

If you have concerns about the security of a site, you may check with www.ssllabs.com/ssltest and in 1-2 mins they will test basic security protocols for that web site, and use a browser that detect and stops tracking (duckduckgo browser, torBrowser, Safari,…)