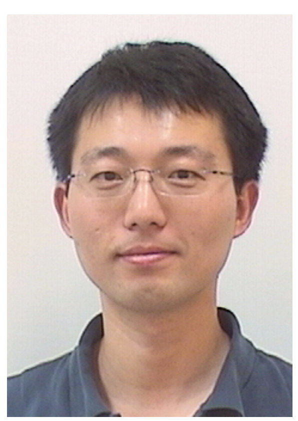
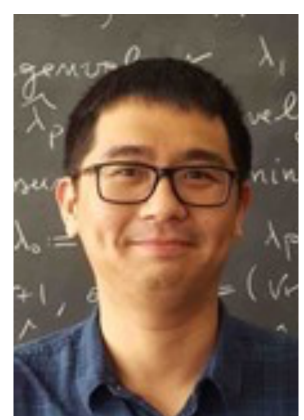
Learning Interaction laws in interacting particle systems

Brin MRC workshop on Scientific Machine Learning

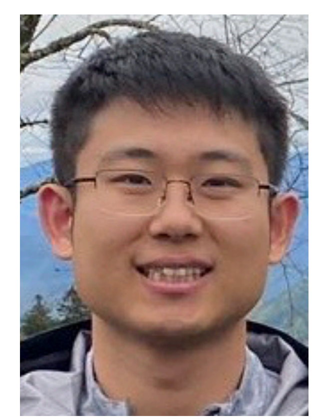
Mauro Maggioni
Johns Hopkins University


F. Lu


S. Tang


M. Zhong


X. Wang


Q. Lang

AFOSR
SIMONS FOUNDATION

https://mauromaggioni.duckdns.org

Learning Interaction Laws in Interacting Particle Systems

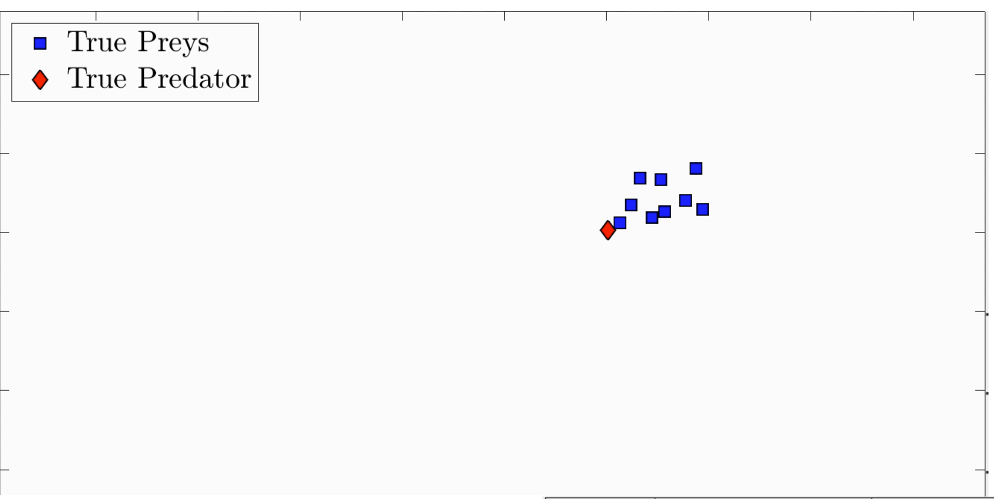
Problem: Given observations of trajectories of a dynamical system of interacting agents, learn the interaction rules.
Motivation: particle-/agent-based systems ubiquitous in Physics, Biology, social sciences, Economics, ... Beyond model-based interaction rules.
Further goals: hypothesis testing for agent-based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

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

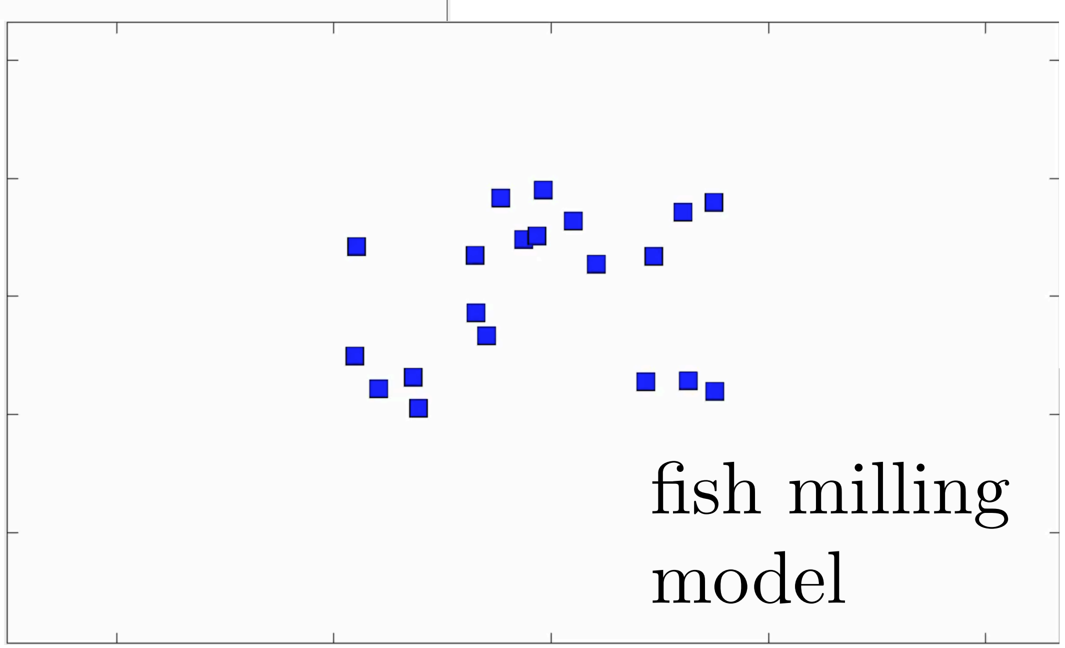
BBC Blue Planet (clip from YouTube)

Learning Interaction Laws in Interacting Particle Systems

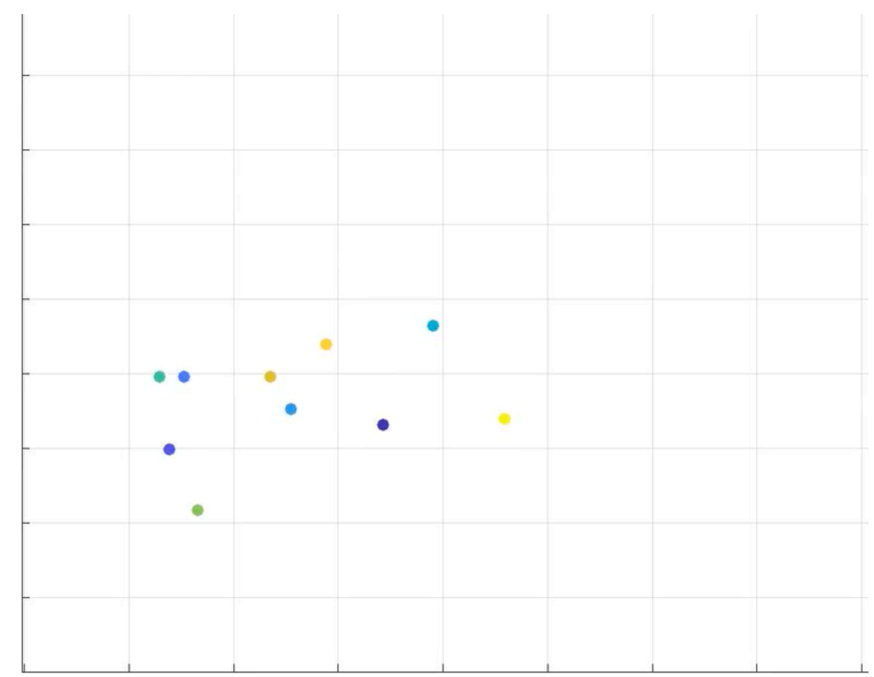
Problem: Given observations of trajectories of a dynamical system of interacting agents, learn the interaction rules.
Motivation: particle-/agent-based systems ubiquitous in Physics, Biology, social sciences, Economics, ... Beyond model-based interaction rules.
Further goals: hypothesis testing for agent-based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



prey/predator model



fish milling model

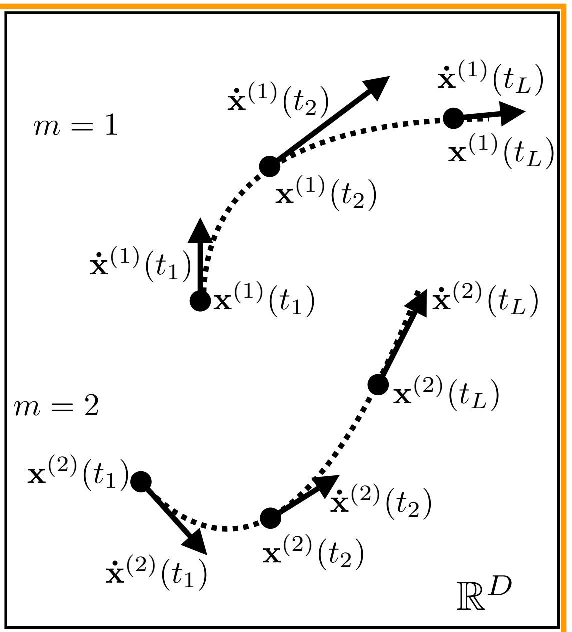


stochastic Lennard-Jones

Estimation/Learning for ODE systems

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

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



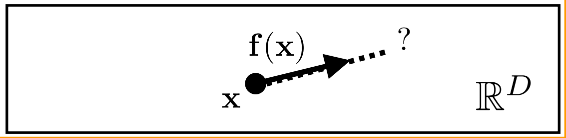
and we are given observations of positions and velocities

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

where:

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

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



Statistical learning version: $(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\dots,L; m=1,\dots,M}$, with $\mathbf{x}^{(m)}(t_1) \sim_{\text{i.i.d.}} \mu_0$; construct an estimator $\hat{\mathbf{f}}_n$ the unknown $\mathbf{f}$.

We are interested in the *nonparametric* setting, i.e. no assumptions on $\mathbf{f}$ except some regularity.

Nonparametric regression

Statistical learning version:
$(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\ldots,L; m=1,\ldots,M}$, with $\mathbf{x}^{(m)}(t_1) \sim_{\text{i.i.d.}} \mu_0$, we want to construct an estimator $\hat{\mathbf{f}}_n$ the unknown $\mathbf{f}$ in $\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t))$.

Possible approach: regression. In regression one is given pairs

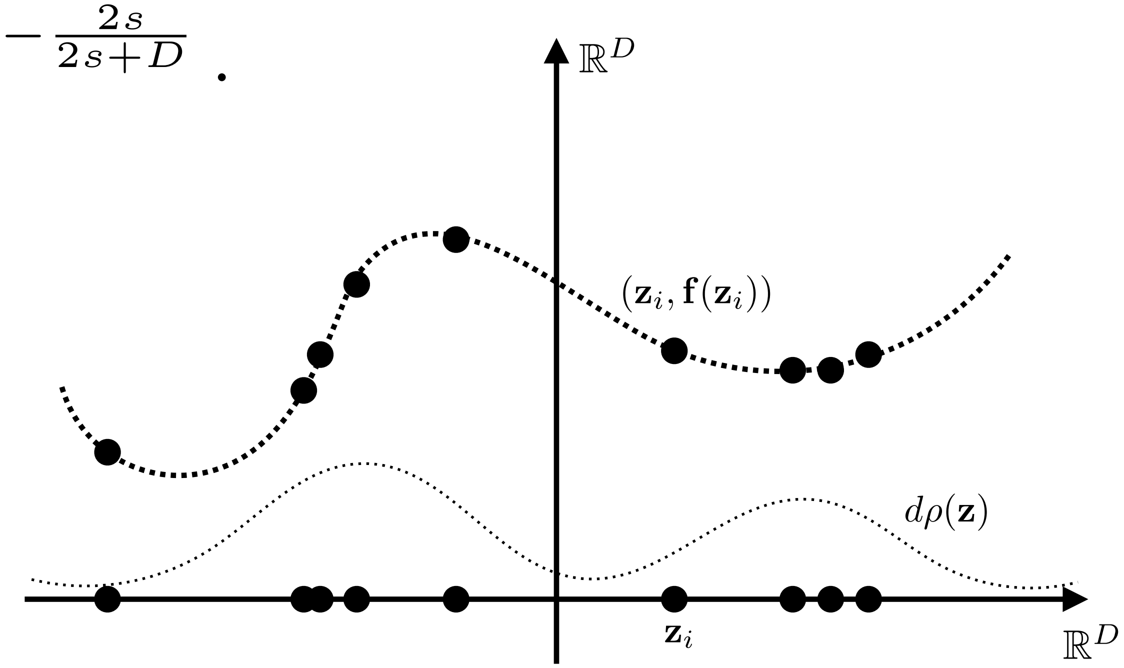
$$\{(\mathbf{z}_i, \mathbf{f}(\mathbf{z}_i) + \eta_i)\}_{i=1}^n \,, \text{with } \mathbf{z}_i \in \mathbb{R}^D, \mathbf{z}_i \sim_{\text{i.i.d.}} \rho\,,$$

with η independent noise, and outputs an estimator $\hat{\mathbf{f}}_n$.

Well-understood problem: estimators that, for $\mathbf{f}$ s-Hölder regular, satisfy

$$\mathbb{E}[||\hat{\mathbf{f}}_n - \mathbf{f}||^2_{L^2(\rho)}] \lesssim n^{-\frac{2s}{2s+D}}\,.$$

Moreover, this *learning rate* is optimal
(in the so-called min-max sense:
for any estimator one can find $\mathbf{f}$
for which the estimator does not converge
to $\mathbf{f}$ any faster than this).



Nonparametric estimation

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

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

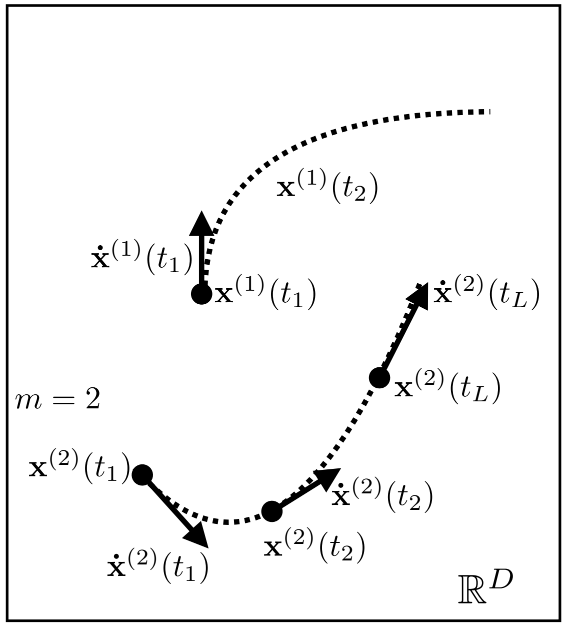
and we are given observations of positions and velocities

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

where:

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

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



$$(\mathbf{x}^{(m)}(t_l), \dot{\mathbf{x}}^{(m)}(t_l))_{l=1,\ldots,L; m=1,\ldots,M}, \text{ with } \mathbf{x}^{(m)}(t_1) \sim_{\text{i.i.d.}} \mu_0, \text{ construct } \hat{\mathbf{f}}_n.$$

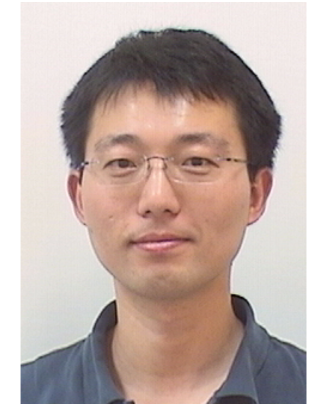
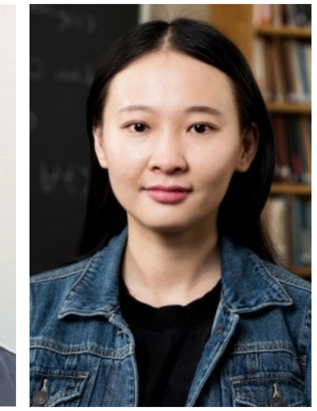
$\mathbf{z}_i$ $\quad$ $\mathbf{f}(\mathbf{z}_i)$

The observations are independent in m, but not in l.

Even if we pretended to have independence, without further assumptions on $\mathbf{f}$, besides s-Hölder regularity, the best attainable rate is $\mathbb{E}[||\hat{\mathbf{f}}_n - \mathbf{f}||_{L^2}] \lesssim n^{-\frac{s}{2s+D}}$, where $n = LM$ (L obs. in each of M traj.) 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 simple prototypical 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.

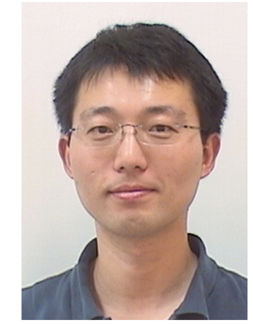
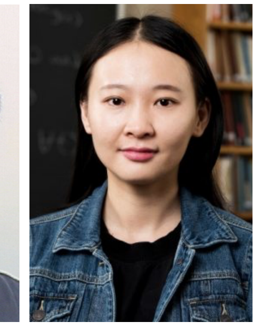
Different limits: $N \to +\infty$ (mean-field limit, joint work with M. Fornasier and M. Bongini), $M \to +\infty$ (joint work with F. Lu, S. Tang and M. Zhong).

- Strong model assumption on the form of the ODE system. Now the unknown is the function ϕ of 1 variable, r.

- We may be able avoid the curse of dimensionality.

- No value $\phi(r)$ is observed, so this is not regression, but an inverse problem.

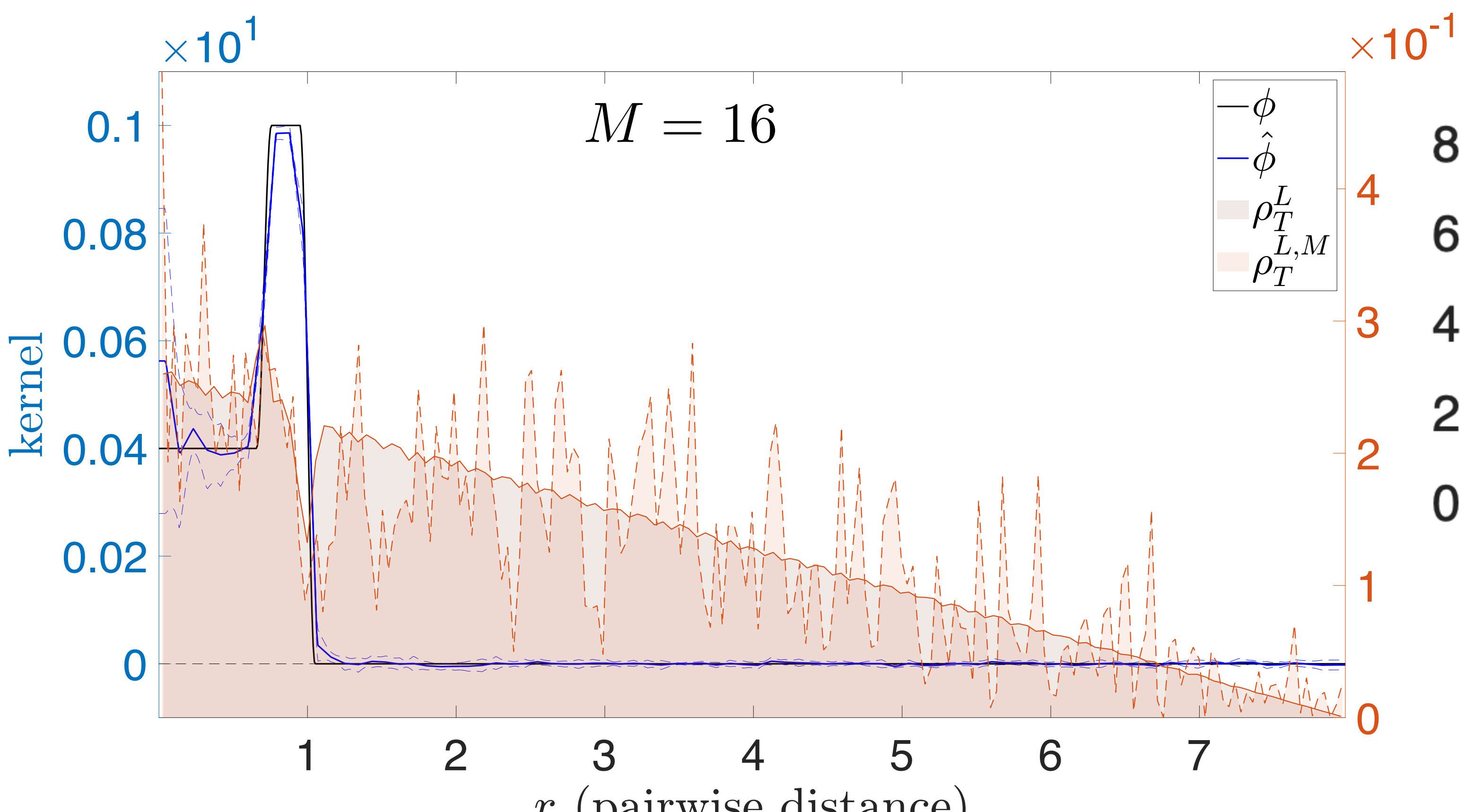
Agent-based systems

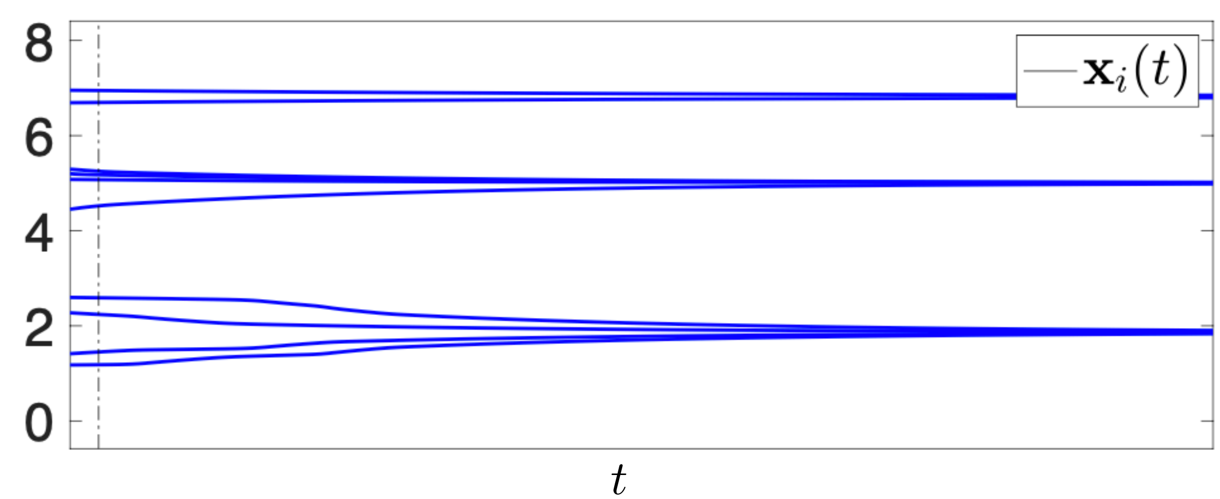




Particle- and agent-based systems are driven by ODEs with special structure. A simple prototypical model:

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

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 work with F. Lu, S. Tang and M. Zhong).





$\mathbf{x}_i > 0$ is the i-th opinion, $i = 1, \ldots, 10$.

The estimator for the interaction kernel

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

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

linear map (in ϕ) applied to unknown ϕ

Consider the empirical error functional

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

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

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

For $\mathcal{H}$ linear subspace, this is a least squares problem (Gauss, Legendre). We want a large $\mathcal{H}$ to reduce bias, but not to large as that increases the number of parameters to be estimated for a given amount of data.

Coercivity condition

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

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

We shall assume that the unknown interaction kernel ϕ is in the admissible class $\mathcal{K}_{R,S} := \{\psi \in C^1(\mathbb{R}_+) :\ \text{supp.}\psi \subset [0, R], \sup_{r\in[0,R]} |\psi(r)| + |\psi'(r)| \le S\}$.
Coercivity condition: $\forall \psi\ :\ \psi(\cdot)\cdot \in \mathcal{H}$, for $c_{L,N,\mathcal{H}}$, $\mathbf{r}_{ii'} := \mathbf{x}_i - \mathbf{x}_{i'}$, $r_{ii'} := ||r_{ii'}||$

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

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

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

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

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

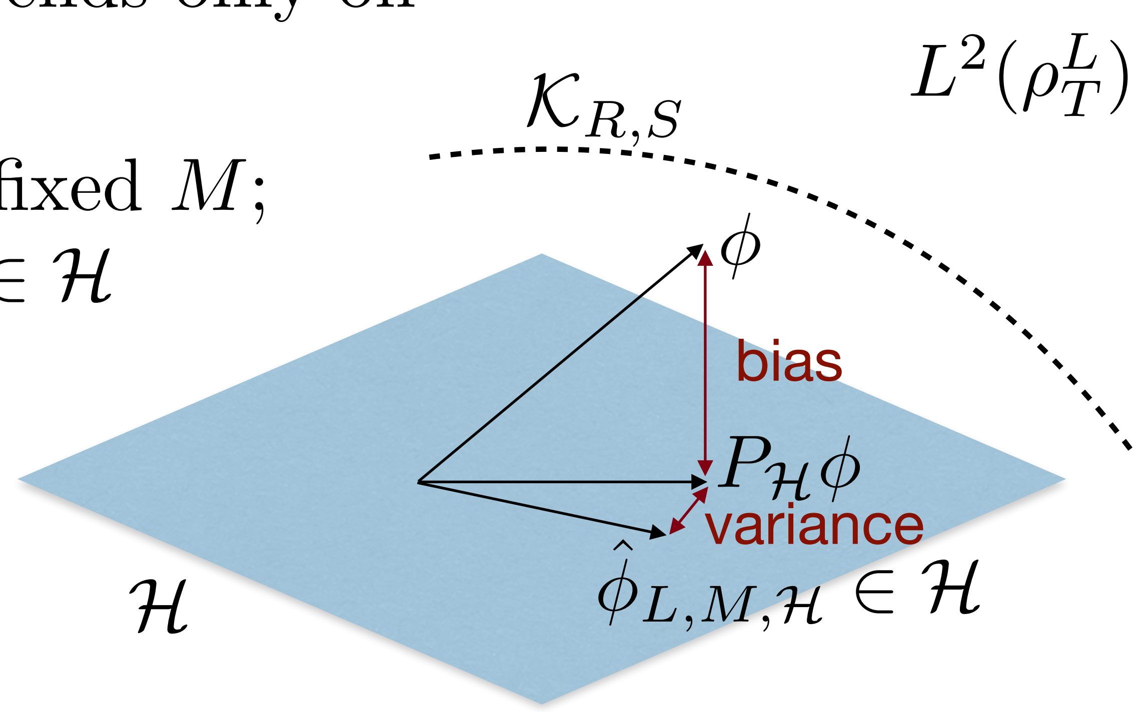
$$\hat{\phi}_{L,M,\mathcal{H}} := \arg\min_{\varphi\in\mathcal{H}} \mathcal{E}_{L,M}(\varphi)\,.$$

$+$ 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 randomness of $\hat{\phi}_{L,M,\mathcal{H}} \in \mathcal{H}$

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 $\dim(\mathcal{H}_n) \le c_0 n$ 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 is just the function that, evaluated at r, gives $\phi(r)r$

this is an occupancy measure for pairwise distances, over trajectories

coercivity constant: it is a crucial parameter controlling how well-conditioned the inverse problem is. Depends on the system.

- 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 that the effective sample size can be $LM = \#\text{obs}$; but that cannot be always true.

In the examples 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 $n = n_*$. Fourier, wavelets, etc...would be other natural choices.

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 $O(N^2 Ld \cdot M + M n_*^2)$ (online versions also possible).

Errors on trajectories

Standard arguments yield bounds on the distance 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:

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

distances between trajectories

quantity controlled by learning theorem

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

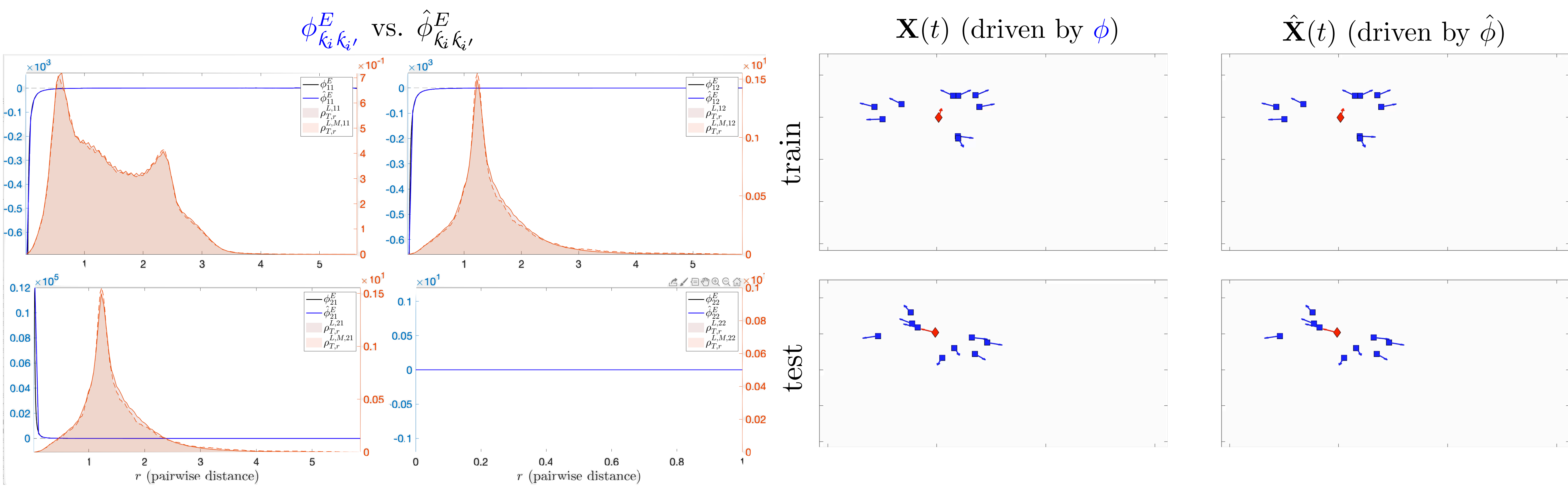
Example: 2nd order systems

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

simple environment (food, light, ...)

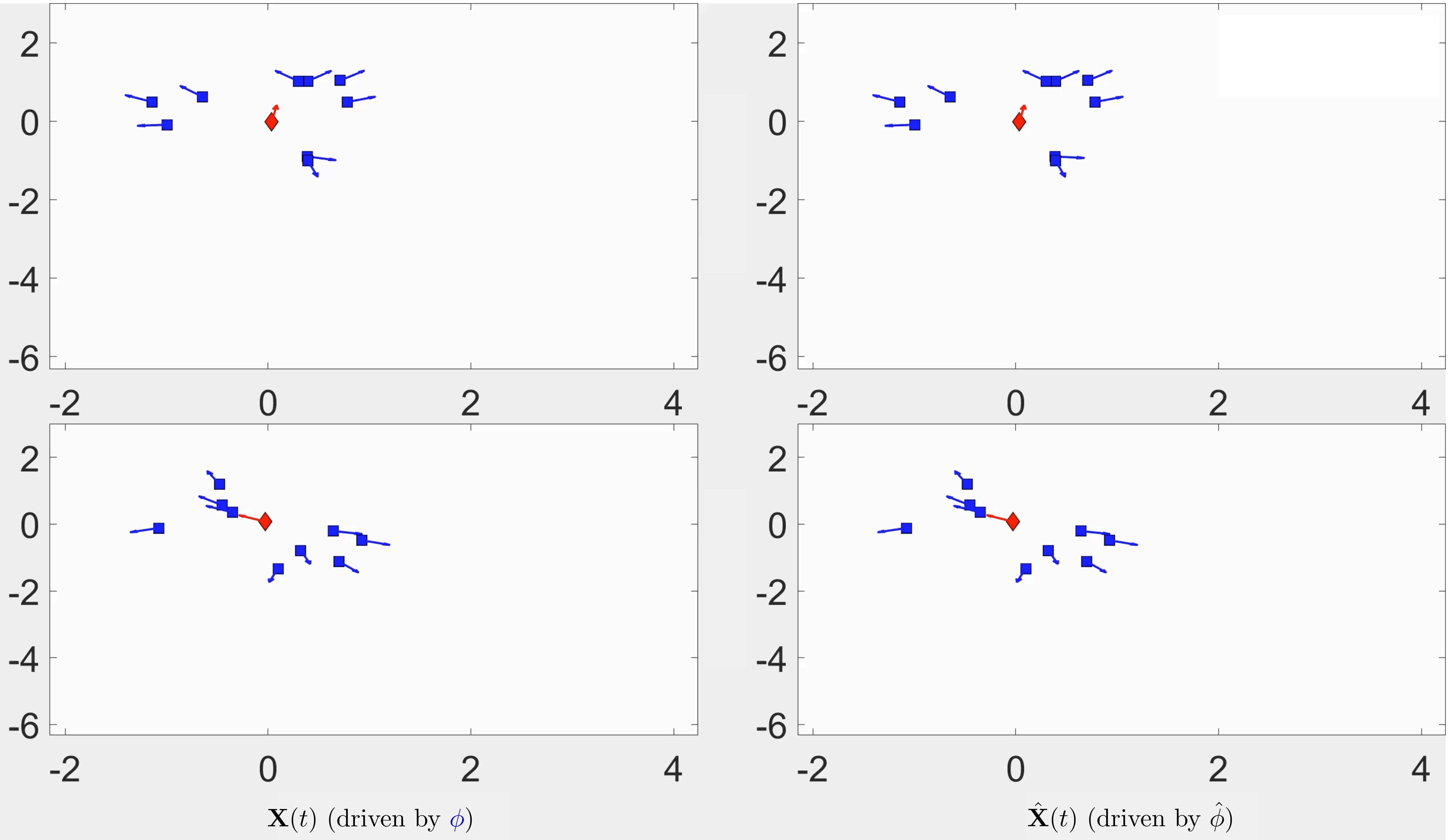
one kernel for each pair of interacting agent **types**

energy and alignment interactions



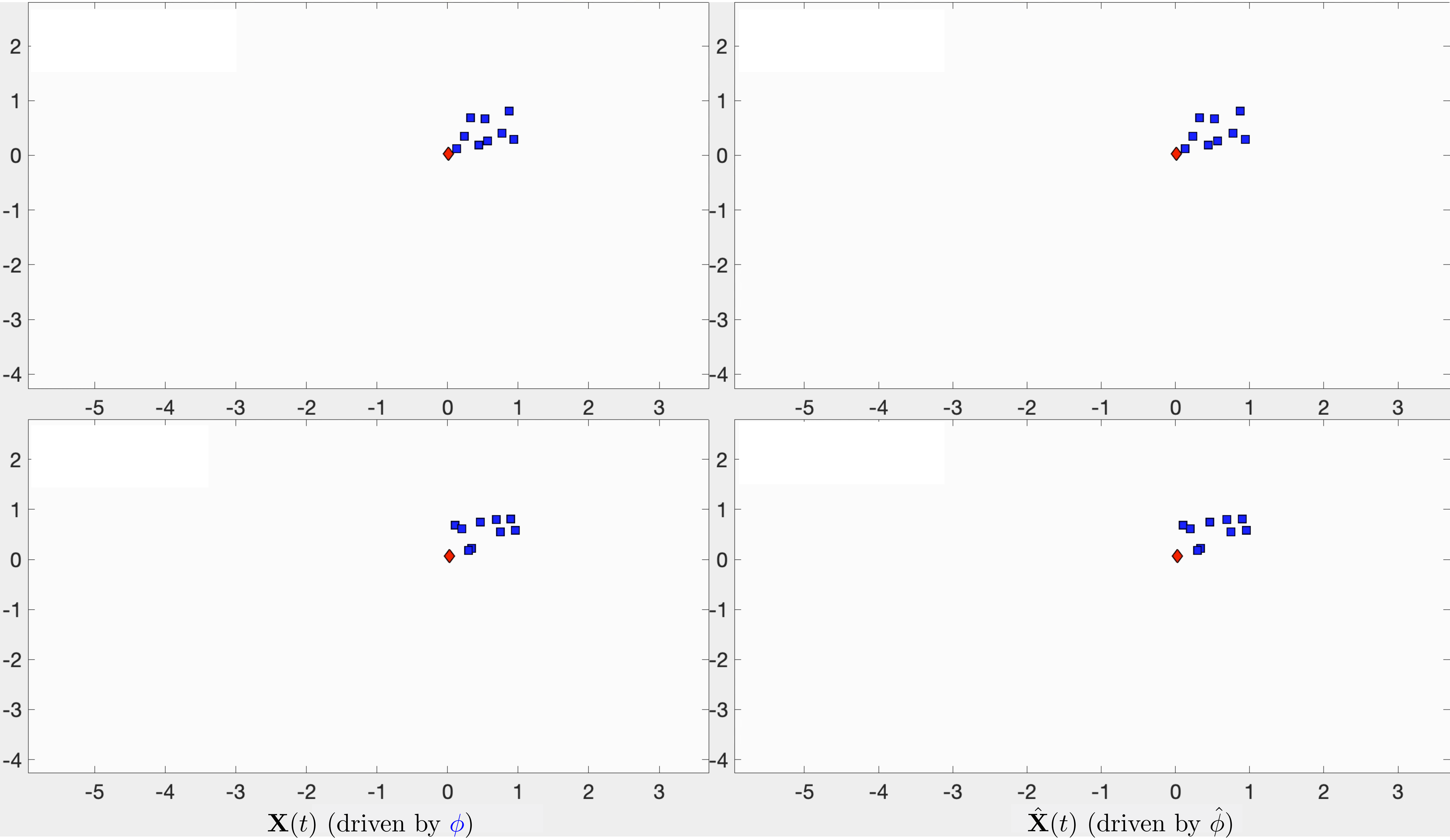
Example 2nd order Prey-Predator system. Left: the interaction kernels and ρ_L^T's. Right: trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

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).

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, with good probability of success.

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

BBC Blue Planet

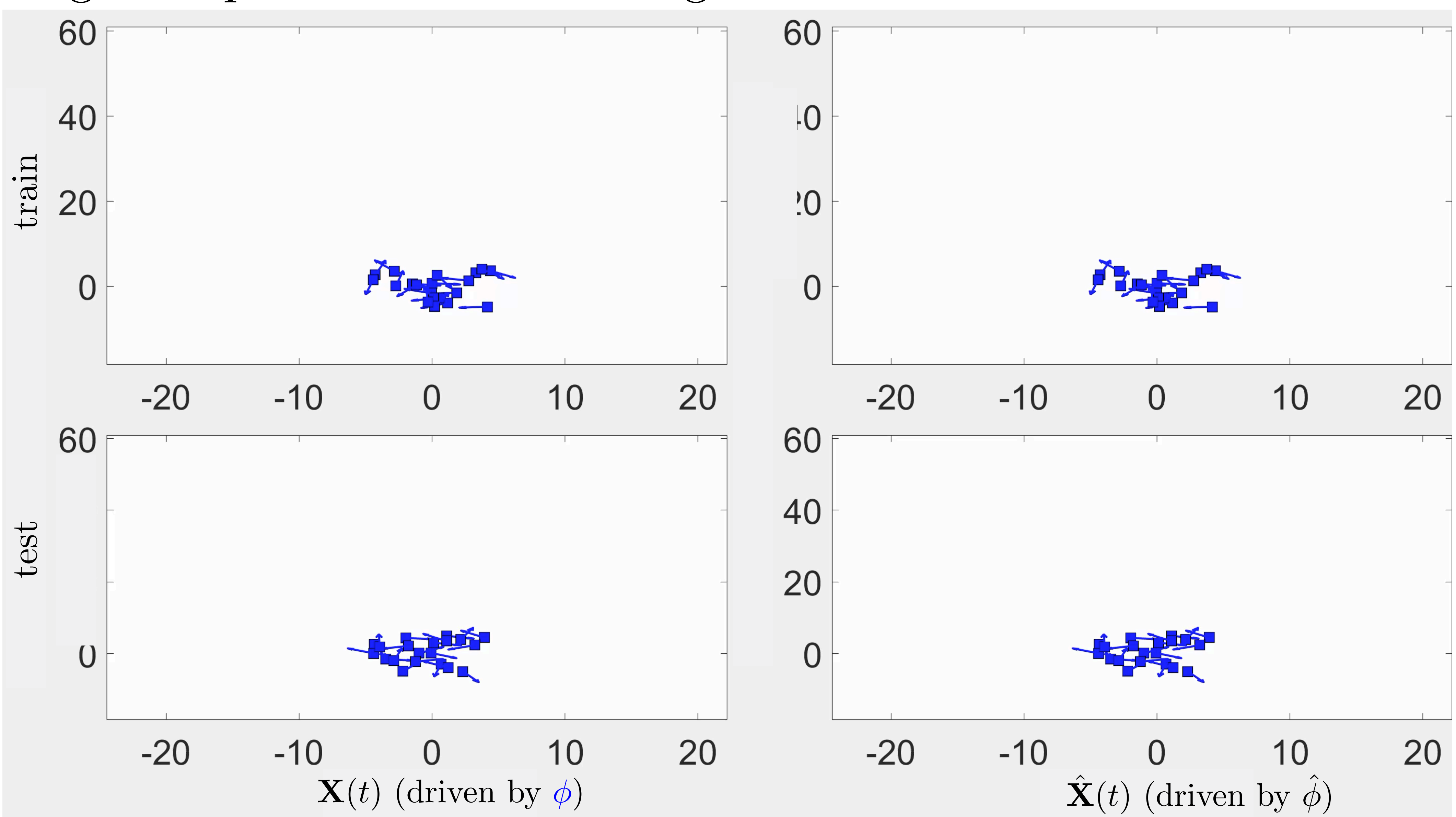
Emerging behaviors: flocking

(*) 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 $+\infty$.

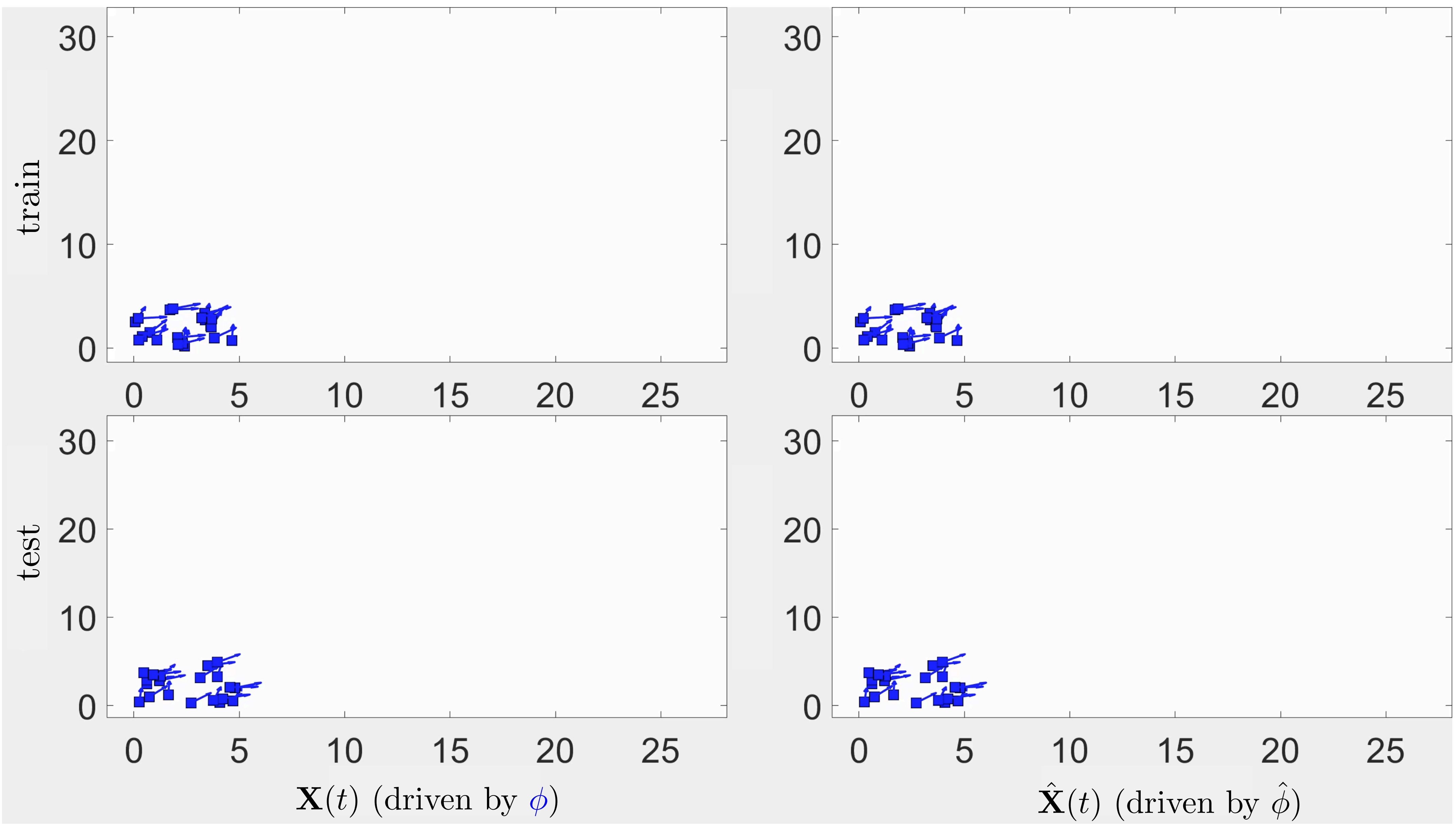


Emerging behaviors: anticipation & flocking

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

$$\ddot{\mathbf{x}}_i = \frac{1}{N}\sum_{i'=1, i'\neq i}^{N} \frac{\tau U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)}{||\mathbf{x}_{i'}-\mathbf{x}_i||}(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)$$
$$+\frac{1}{N}\sum_{i'=1, i'\neq i}^{N}\left\{\frac{-\tau U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)(\mathbf{x}_{i'}-\mathbf{x}_i)\cdot(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)}{||\mathbf{x}_{i'}-\mathbf{x}_i||^3}\right.$$
$$\left.+\frac{\tau U''(||\mathbf{x}_{i'}-\mathbf{x}_i||)(\mathbf{x}_{i'}-\mathbf{x}_i)\cdot(\dot{\mathbf{x}}_{i'}-\dot{\mathbf{x}}_i)}{||\mathbf{x}_{i'}-\mathbf{x}_i||^2}+\frac{U'(||\mathbf{x}_{i'}-\mathbf{x}_i||)}{||\mathbf{x}_{i'}-\mathbf{x}_i||}\right\}(\mathbf{x}_{i'}-\mathbf{x}_i).$$

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



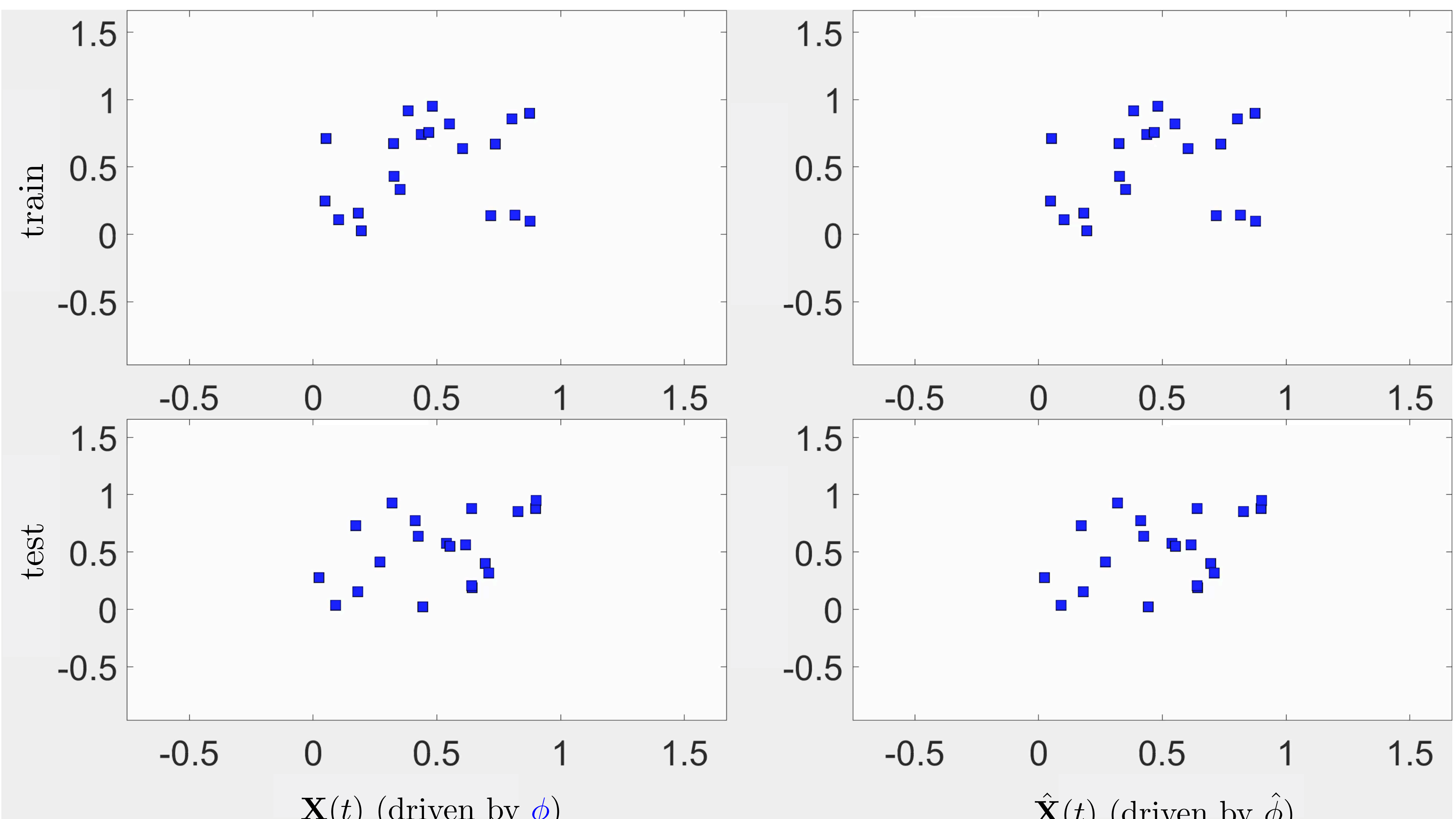
Emerging behaviors: Fish mill patterns

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

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

$$m_i \ddot{\mathbf{x}}_i = \alpha \dot{\mathbf{x}}_i - \beta ||\dot{\mathbf{x}}_i||^2 \dot{\mathbf{x}}_i - \sum_{i'} \nabla_2 U(\mathbf{x}_i, \mathbf{x}_{i'}) \,,$$

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



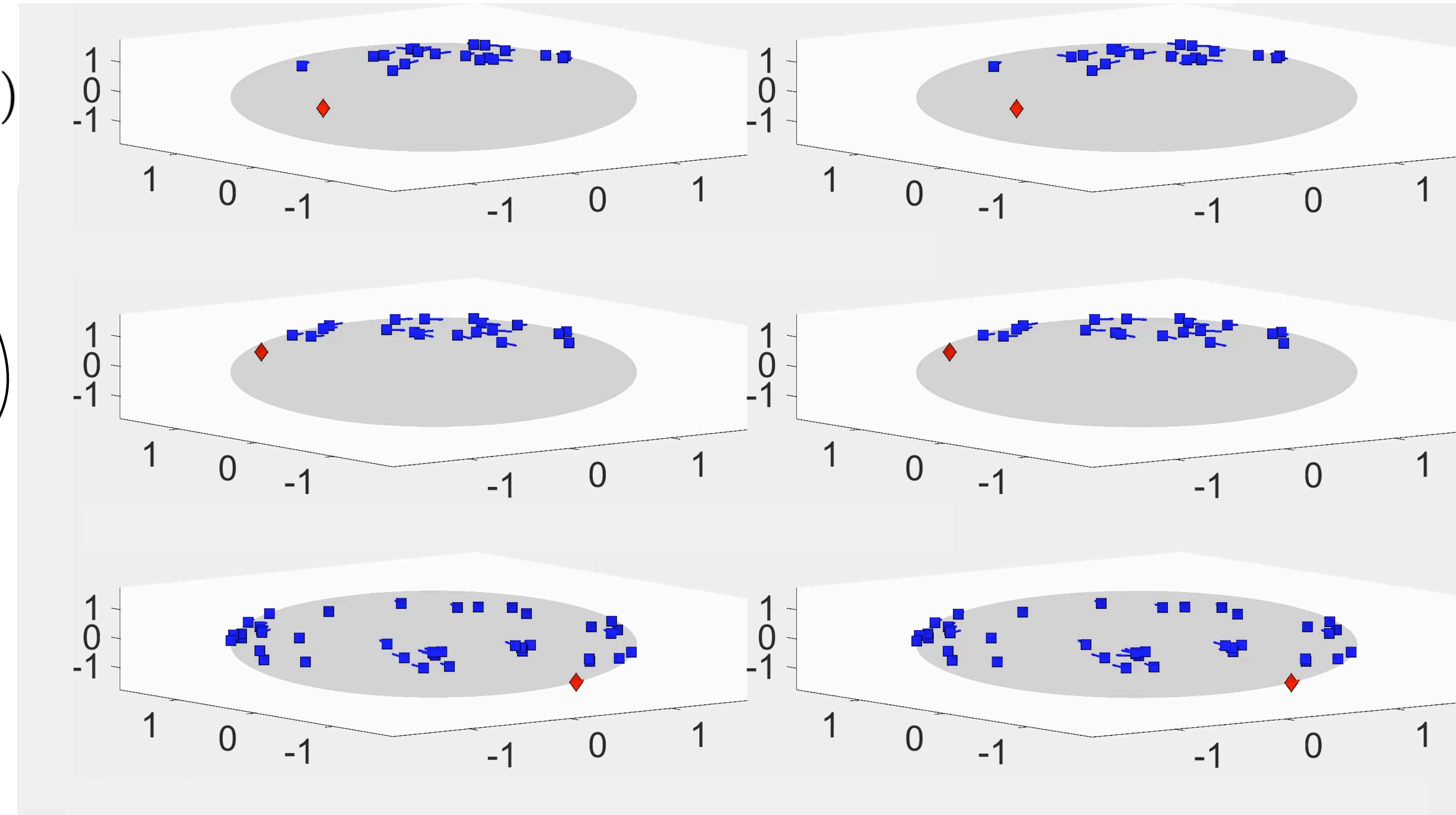
Interacting particles on manifolds

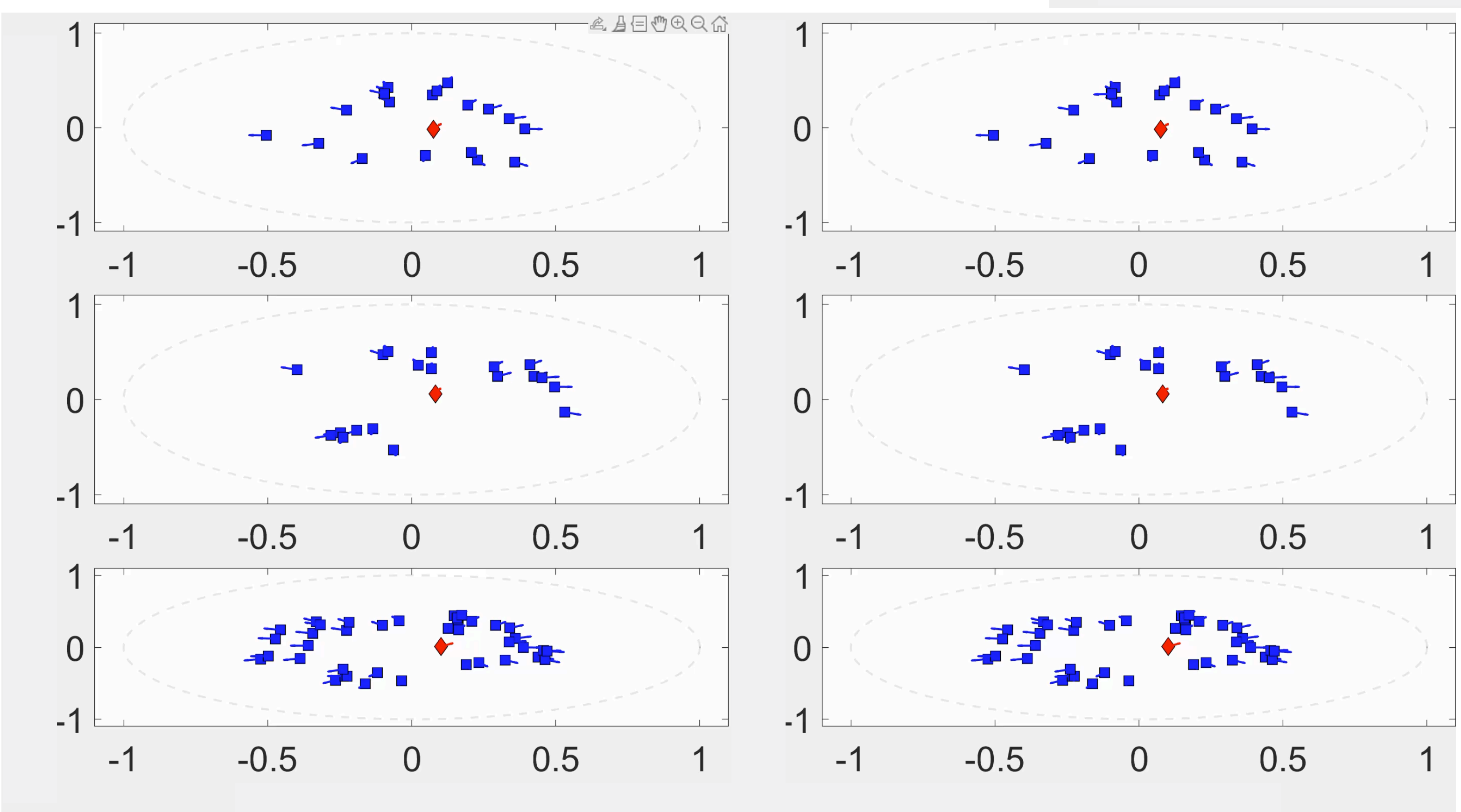
$$\dot{\mathbf{x}}_i^{(m)} = \frac{1}{N}\sum_{i'=1}^{N} \phi(||\mathbf{x}_i^{(m)} - \mathbf{x}_{i'}^{(m)}||)(\mathbf{x}_{i'}^{(m)} - \mathbf{x}_i^{(m)})$$

Generalization to manifolds:

· distances → geodesic distances

· $(\mathbf{x_{i'}} - \mathbf{x_i})/||\mathbf{x_{i'}} - \mathbf{x_i}||$ → direction of tangent to geodesic from $\mathbf{x_i}$ to $\mathbf{x_{i'}}$ at $\mathbf{x_i}$.





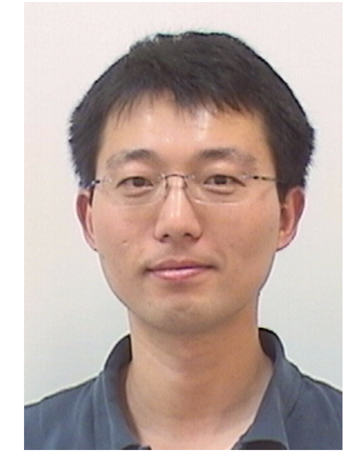
$\mathbb{S}^2$

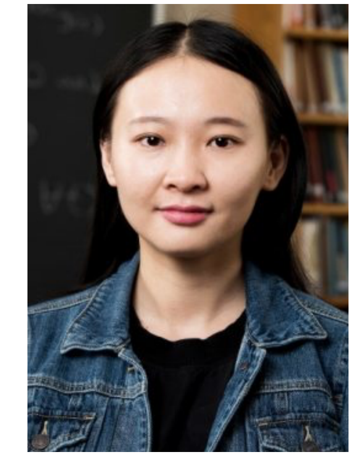
Poincaré disk

Prey-predator system on...

MM, J. Miller, H. Qiu, M. Zhong, *Learning Interaction Kernels for Agent Systems on Riemannian Manifolds*, ICML 2021

The Stochastic case





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

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

Joint work with F. Lu and S. Tang, *Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories*, FOCM, 2021.

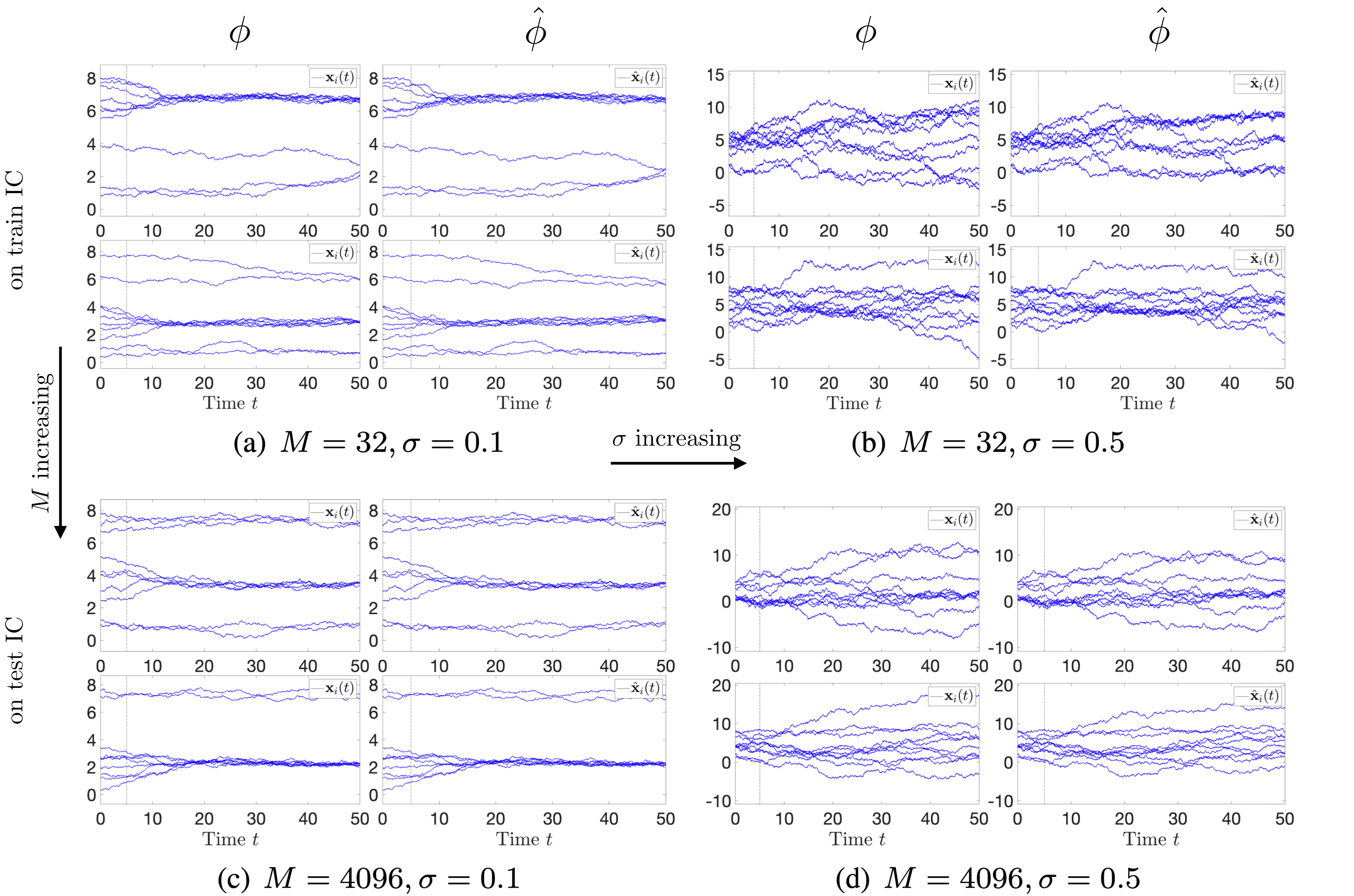
Note that in the stochastic case we do not (cannot!) observe velocities, but only positions. We have studied carefully the dependence on the observation time gap $\Delta t := t_{l+1} - t_l = T/L$:

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

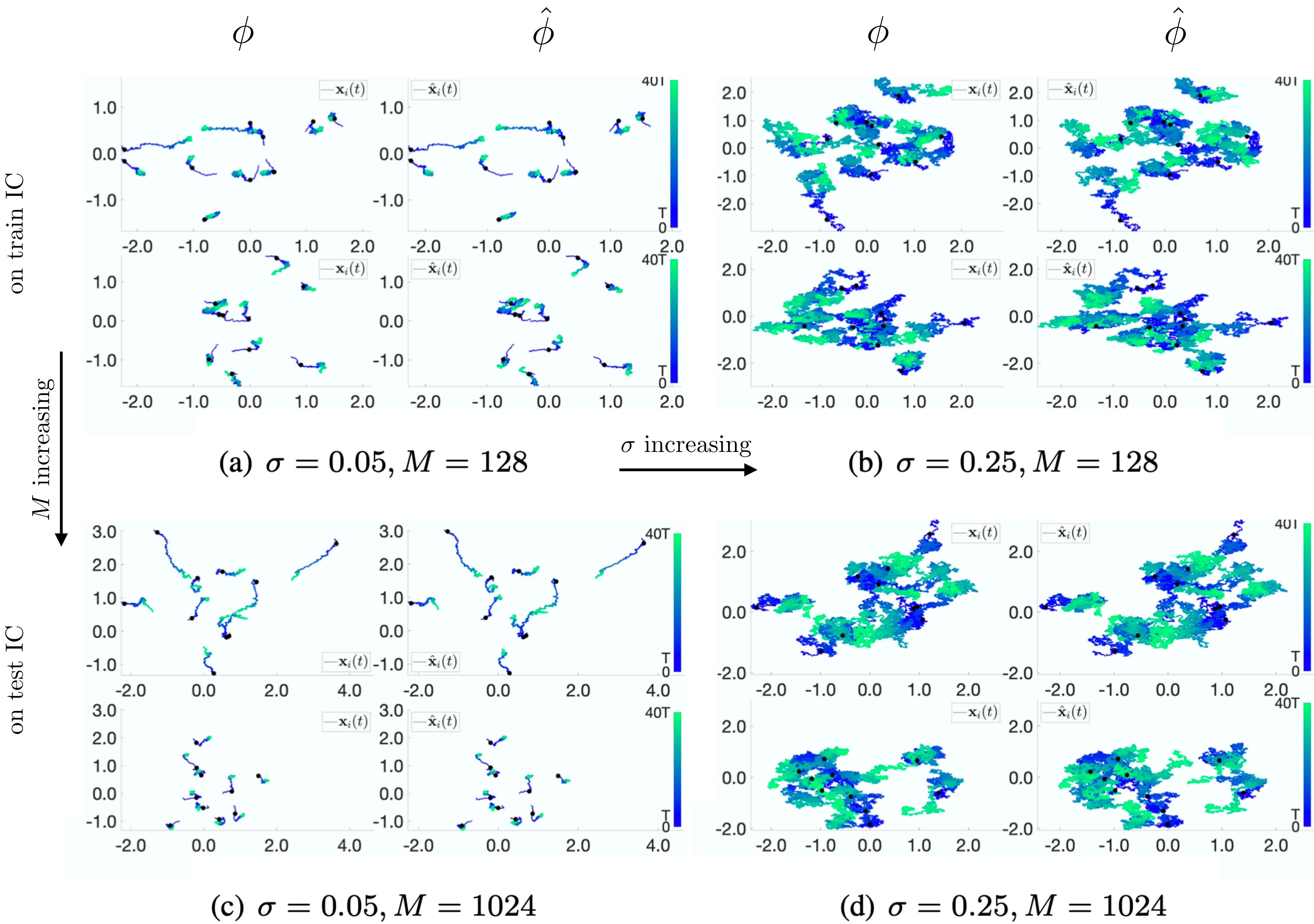
approximation error | statistical error | discretization error

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

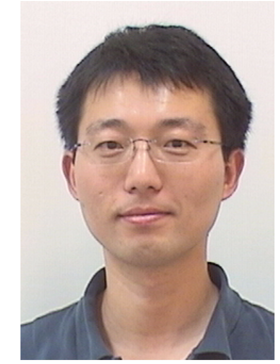
Stochastic opinion dynamics



Stochastic Lennard-Jones



Interacting Particle Systems on Networks


F. Lu


X. Wang


Q. Lang

We consider a heterogeneous dynamical system with N interacting particles on a graph: $G = (V, E, \mathbf{a})$ a graph, $\mathbf{a} = (\mathbf{a}_{ij}) \in [0,1]^{N\times N}$, $\mathbf{a}_{ij} > 0$ iff $(i,j) \in E$. At each vertex $i \in \{1, \ldots, N\}$ there is a particle $X_t^i \in \mathbb{R}^d$, with dynamics

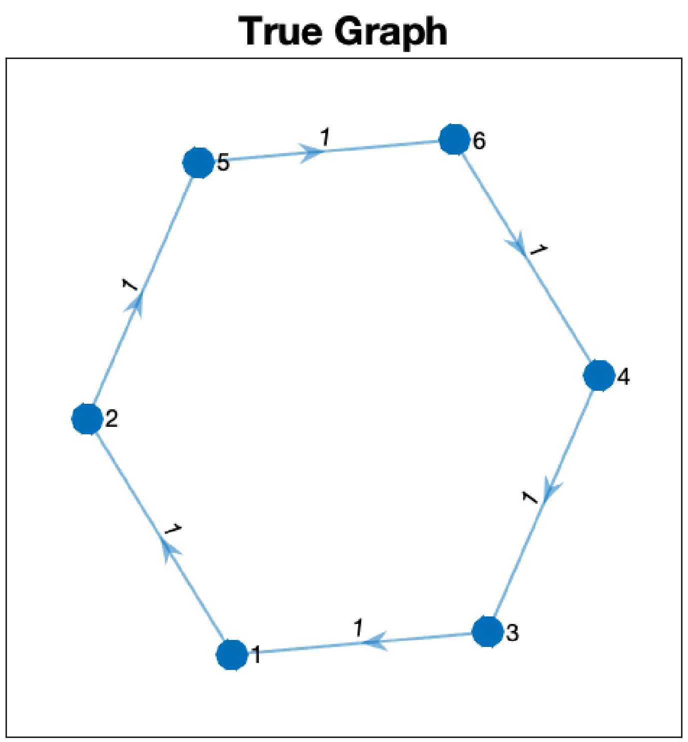
$$\mathcal{S}_{\mathbf{a},\Phi} \; : \qquad dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi(X_t^j - X_t^i)dt + \sigma dW_t^i, \qquad i = 1, \ldots, N$$

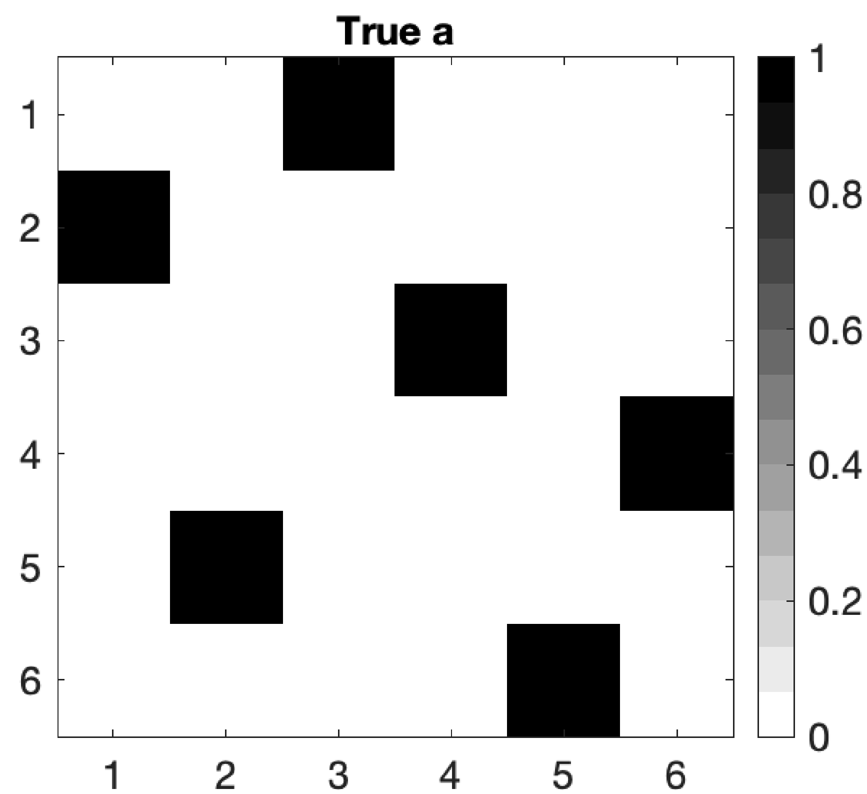
Observations: $\{\mathbf{X}_{t_l}^{(m)}\}_{l\in[L],m\in[M]}$ + noise, where $\mathbf{X} = (X_i)_{i\in[N]} \in \mathbb{R}^{N\times d}$.

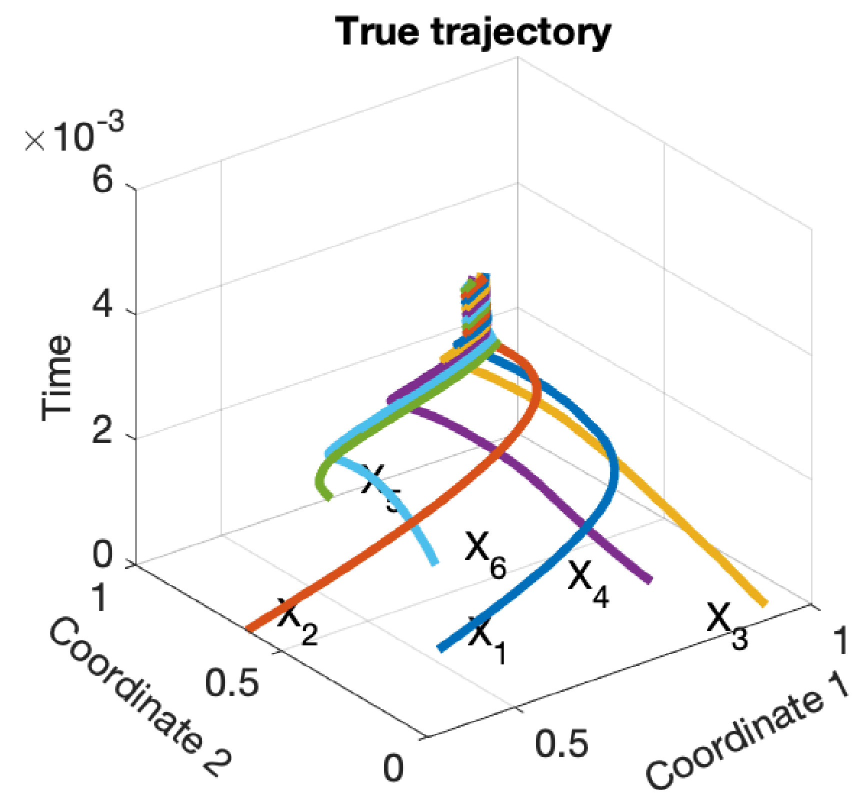
Want to estimate both $\mathbf{a} \in [0,1]^{N\times N}$ and $\Phi : \mathbb{R}^d \to \mathbb{R}^d$.

Lennard-Jones interactions on a network

$$\phi(x) = \left(-\tfrac{1}{3}x^{-9} + \tfrac{4}{3}x^{-3}\right)\mathbf{1}_{x\geq 0.5} - 160\,\mathbf{1}_{0\leq x<0.5}$$







M	N	p	L	T	σ	σ_{obs}
10^3	6	10	50	$5\cdot 10^{-3}$	10^{-3}	10^{-3}

Interacting Particle Systems on Networks

We consider a heterogeneous dynamical system with N interacting particles on a graph: $G = (V, E, \mathbf{a})$ a graph, $\mathbf{a} = (\mathbf{a}_{ij}) \in [0,1]^{N\times N}$, $\mathbf{a}_{ij} > 0$ iff $(i,j) \in E$. At each vertex $i \in \{1, \ldots, N\}$ there is a particle $X_t^i \in \mathbb{R}^d$, with dynamics

$$\mathcal{S}_{\mathbf{a},\Phi} \; : \qquad dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi(X_t^j - X_t^i)dt + \sigma dW_t^i, \qquad i = 1, \ldots, N$$

Observations: $\{\mathbf{X}_{t_l}^{(m)}\}_{l\in[L],m\in[M]}$ + noise, where $\mathbf{X} = (X_i)_{i\in[N]} \in \mathbb{R}^{N\times d}$.

Want to estimate both $\mathbf{a} \in [0,1]^{N\times N}$ and $\Phi : \mathbb{R}^d \to \mathbb{R}^d$.

Parametric setting for simplicity: $\Phi \in \mathcal{H}$, for some given finite-dimensional hypothesis space $\mathcal{H} = \mathrm{span}\{\psi_k\}_{k\in[p]}$; then $\Phi = \sum_{k\in[p]} c_k\psi_k$.

$$(\widehat{\mathbf{a}}, \widehat{c}) = \mathrm{argmin}_{(\mathbf{a},c)}\mathcal{E}_{L,M}(\mathbf{a}, c)$$

$$\mathcal{E}_{L,M}(\mathbf{a}, c) := \frac{1}{MT}\sum_{l=0,m=1}^{L-1,M} \left\|\Delta\mathbf{X}_{t_l}^m - \mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)c\Delta t\right\|_F^2$$

where $\mathbf{B}(\mathbf{X}_t)_i := \left(\psi_k(X_t^j - X_t^i)\right)_{j,k} \in \mathbb{R}^{N\times 1\times d\times p}$ for each $i \in [N]$.

Interacting Particle Systems on Networks

$$\mathcal{S}_{\mathbf{a},\Phi} \ : \qquad dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi(X_t^j - X_t^i)dt + \sigma dW_t^i, \qquad i = 1,\ldots,N$$

Observations: $\{\mathbf{X}_{t_l}^{(m)}\}_{l\in[L],m\in[M]}$ + noise, where $\mathbf{X} = (X_i)_{i\in[N]} \in \mathbb{R}^{N\times d}$.

Parametric setting for simplicity: $\Phi \in \mathcal{H}$, for some given finite-dimensional hypothesis space $\mathcal{H} = \mathrm{span}\{\psi_k\}_{k\in[p]}$; then $\Phi = \sum_{k\in[p]} c_k\psi_k$.

$$(\widehat{\mathbf{a}}, \widehat{c}) = \mathrm{argmin}_{(\mathbf{a},c)}\mathcal{E}_{L,M}(\mathbf{a}, c)$$

$$\mathcal{E}_{L,M}(\mathbf{a}, c) := \frac{1}{MT}\sum_{l=0,m=1}^{L-1,M} \left\|\Delta\mathbf{X}_{t_l}^m - \mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)c\Delta t\right\|_F^2$$

Normalization: $||\mathbf{a}_{i,\cdot}||_2 = 1$, defining the set $\mathcal{M}$ of admissible weights.

$\mathcal{E}$ nonlinear, non-convex, but separately convex in each of the two arguments.

Alternating Least Squares

$$(\widehat{\mathbf{a}}, \widehat{c}) = \operatorname{argmin}_{(\mathbf{a},c)} \mathcal{E}_{L,M}(\mathbf{a}, c)$$

$$\mathcal{E}_{L,M}(\mathbf{a}, c) := \frac{1}{MT} \sum_{l=0,m=1}^{L-1,M} \left\| \Delta \mathbf{X}_{t_l}^m - \mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m) c \Delta t \right\|_F^2$$

1. Given c, estimate $\mathbf{a}$ by directly solving the minimizer of the quadratic loss function with c fixed, which solves

$$\widehat{\mathbf{a}}_{i\cdot} \mathcal{A}_{c,M,i}^{\mathrm{ALS}} := \widehat{\mathbf{a}}_{i\cdot}([\mathbf{B}(\mathbf{X}_{t_l}^m)_i]_{l,m} c) = [(\Delta \mathbf{X}_{t_l}^m)_i]_{l,m} / \Delta t$$

with $[\mathbf{B}(\mathbf{X}_{t_l}^m)_i]_{l,m} \in \mathbb{R}^{N \times (dLM) \times p}$, $\mathcal{A}_{c,M,i}^{\mathrm{ALS}} := [\mathbf{B}(\mathbf{X}_{t_l}^m)_i]_{l,m} c \in \mathbb{R}^{N \times (dLM)}$ and $[\Delta \mathbf{X}_{t_l}^m]_{l,m} \in \mathbb{R}^{N \times dLMN}$ obtained by multiplying appropriate tensor slices by c.

2. Given $\mathbf{a}$, estimate c by minimizing the loss function with fixed $\mathbf{a}$ by solving

$$\mathcal{A}_{\mathbf{a},M}^{\mathrm{ALS}} \widehat{c} := [\mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)]_{l,m} \widehat{c} = [\Delta \mathbf{X}_{t_l}^m]_{l,m} / \Delta t \,,$$

where $\mathcal{A}_{\mathbf{a},M}^{\mathrm{ALS}} := [\mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)]_{l,m} \in \mathbb{R}^{dLMN \times p}$ is again obtained by stacking in a block-row fashion and $\mathcal{A}_{\mathbf{a},M,i}^{\mathrm{ALS}} := [\mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)_i]_{l,m}$.

Operation Regression + ALS

$$(\widehat{\mathbf{a}}, \widehat{c}) = \operatorname{argmin}_{(\mathbf{a},c)} \mathcal{E}_{L,M}(\mathbf{a}, c)$$

$$\mathcal{E}_{L,M}(\mathbf{a}, c) := \frac{1}{MT} \sum_{l=0,m=1}^{L-1,M} \left\| \Delta \mathbf{X}_{t_l}^m - \mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m) c \Delta t \right\|_F^2$$

Operator Regression. Consider $\{\mathbf{Z}_i = \mathbf{a}_{i,\cdot}^\top c^\top \in \mathbb{R}^{(N-1)\times p}\}_{i=1}^N$ treated as vectors $z_i \in \mathbb{R}^{(N-1)p\times 1}$; they solve

$$\mathcal{A}_{i,M} z_i = [\mathcal{A}_i]_{l,m} z_i := [(\mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m) c \Delta t)_i]_{l,m} = [(\Delta \mathbf{X}_{t_l}^m)_i]_{l,m}, \quad i \in [N],$$

where $\mathcal{A}_{i,M} = [\mathcal{A}_i]_{l,m} \in \mathbb{R}^{dML\times(N-1)p}$, since the loss function can be written as $\frac{1}{ML} \sum_{l,m,i=1}^{L,M,N} \left| [(\Delta \mathbf{X}^m)_i]_{l,m} - [\mathcal{A}_i]_{l,m} z_i \right|^2$.

Deterministic ALS stage. The rows of $\mathbf{a}$ and the vector c are estimated via a joint factorization of the matrices of the estimated vectors $\{\widehat{z}_{i,M}\}$, denoted by $\widehat{\mathbf{Z}}_{i,M}$, with a shared vector c:

$$(\widehat{\mathbf{a}}^M, \widehat{c}^M) = \operatorname{argmin}_{\mathbf{a}\in\mathcal{M},\, c\in\mathbf{R}^p} \mathcal{E}(\mathbf{a}, c) := \sum_{i=1}^N \left\| \widehat{\mathbf{Z}}_{i,M} - \mathbf{a}_{i,\cdot}^\top c^\top \right\|_F^2$$

Theoretical results

The system satisfies a **rank-2 joint coercivity condition** on $\mathcal{H}$ if $\exists\, c_{\mathcal{H}} > 0$ s.t. $\forall\, \Phi_1, \Phi_2 \in \mathcal{H}$ with $\langle \Phi_1, \Phi_2 \rangle_{L^2(\rho_L)} = 0$, $\forall\, \mathbf{a}^{(1)}, \mathbf{a}^{(2)} \in \mathcal{M}$ and $\forall\, i \in [N]$

$$\frac{1}{L}\sum_{l=0}^{L-1} \mathbb{E}\left[\Big|\sum_{j\neq i}[\mathbf{a}_{ij}^{(1)}\Phi_1(\mathbf{r}_{ij}(t_l)) + \mathbf{a}_{ij}^{(2)}\Phi_2(\mathbf{r}_{ij}(t_l))]\Big|^2\right] \geq c_{\mathcal{H}}\left[|\mathbf{a}_{i\cdot}^{(1)}|^2\|\Phi_1\|_{\rho_L}^2 + |\mathbf{a}_{i\cdot}^{(2)}|^2\|\Phi_2\|_{\rho_L}^2\right]$$

uniqueness of the minimizer for $M = \infty$, solution of $\mathcal{E}_{L,\infty}(\mathbf{a}, \Phi) = 0$.

matrices in the least squares steps of ALS are well-conditioned.

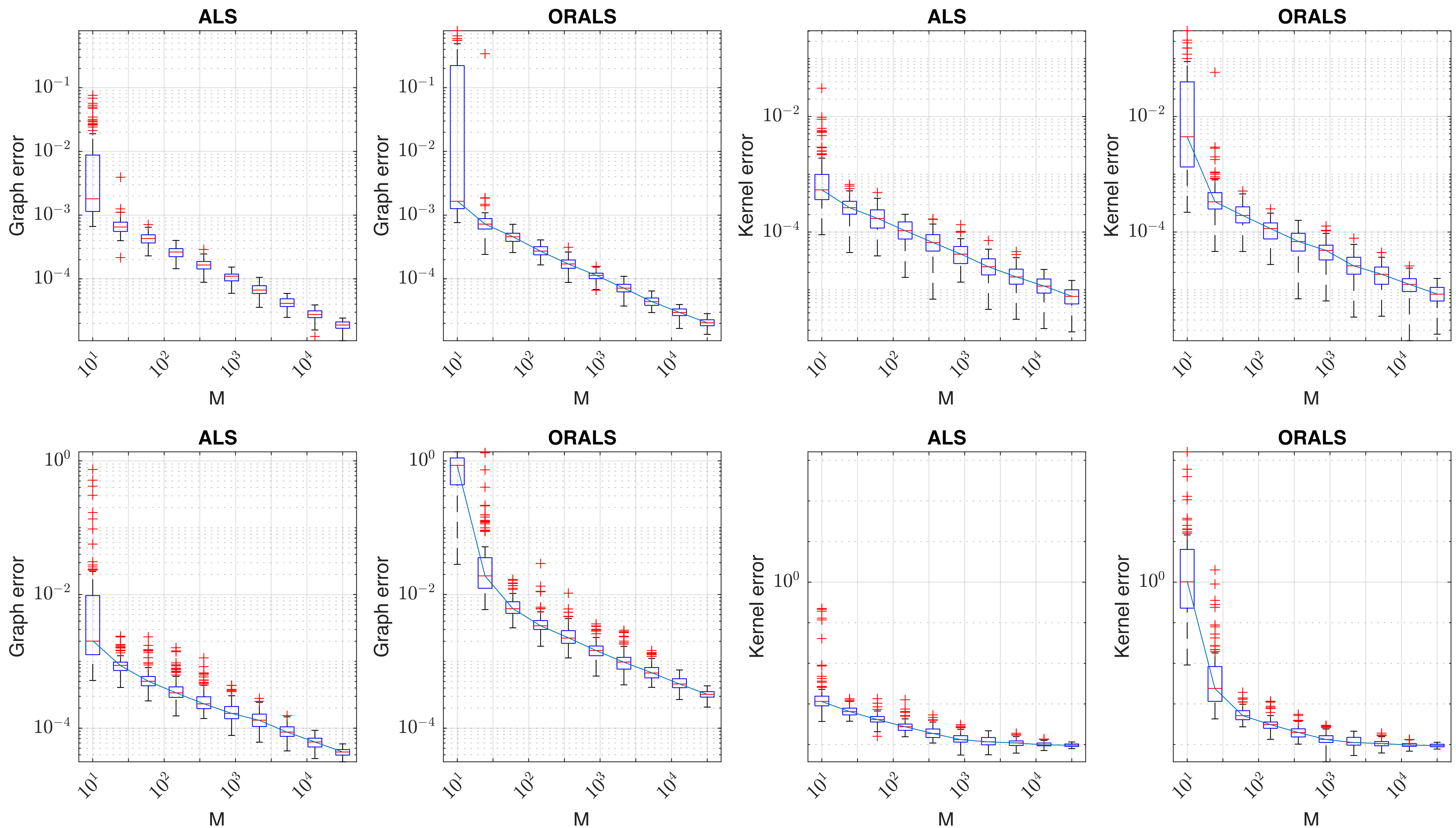
The system satisfies an **interaction kernel coercivity condition** in $\mathcal{H}$ if $\exists c_{0,\mathcal{H}} \in (0,1)$ s.t. $\forall \Phi \in \mathcal{H}$ and $i \in [N]$ $\frac{1}{L(N-1)}\sum_{l=0}^{L-1}\sum_{j\neq i}\mathbb{E}[\mathrm{tr}\,\mathrm{Cov}(\Phi(\mathbf{r}_{ij}(t_l)) \mid \mathcal{F}_l^i)] \geq c_{0,\mathcal{H}}\|\Phi\|_{\rho_L}^2$ where $\mathcal{F}_l^i$ is the σ-algebra generated by $(\mathbf{X}_{t_{l-1}}, X_{t_l}^i)$.

rank-2 joint coercivity

ORALS yields consistent and asymptotically normal estimator

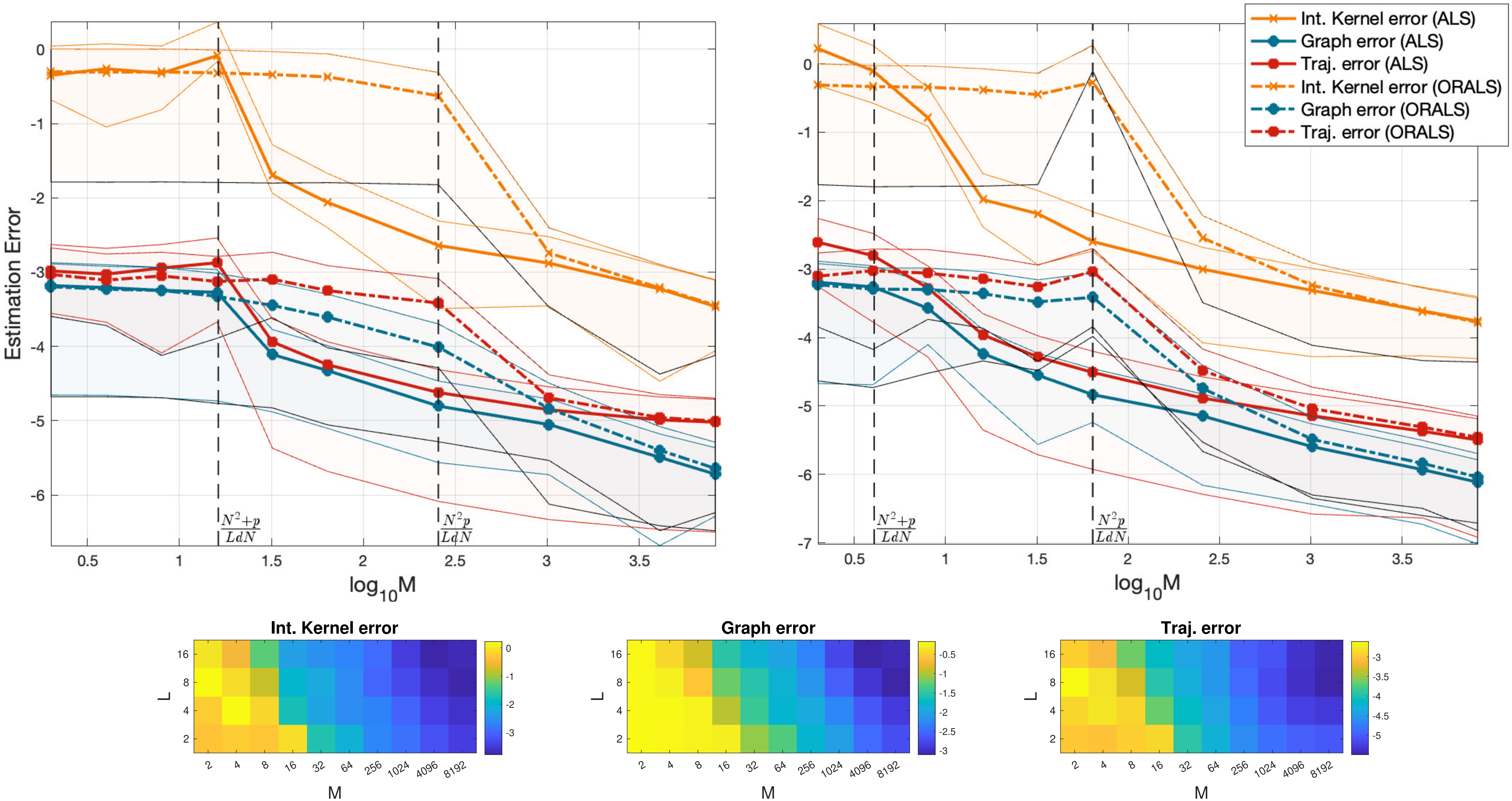
matrices in ORALS are well-conditioned

Convergence for large M



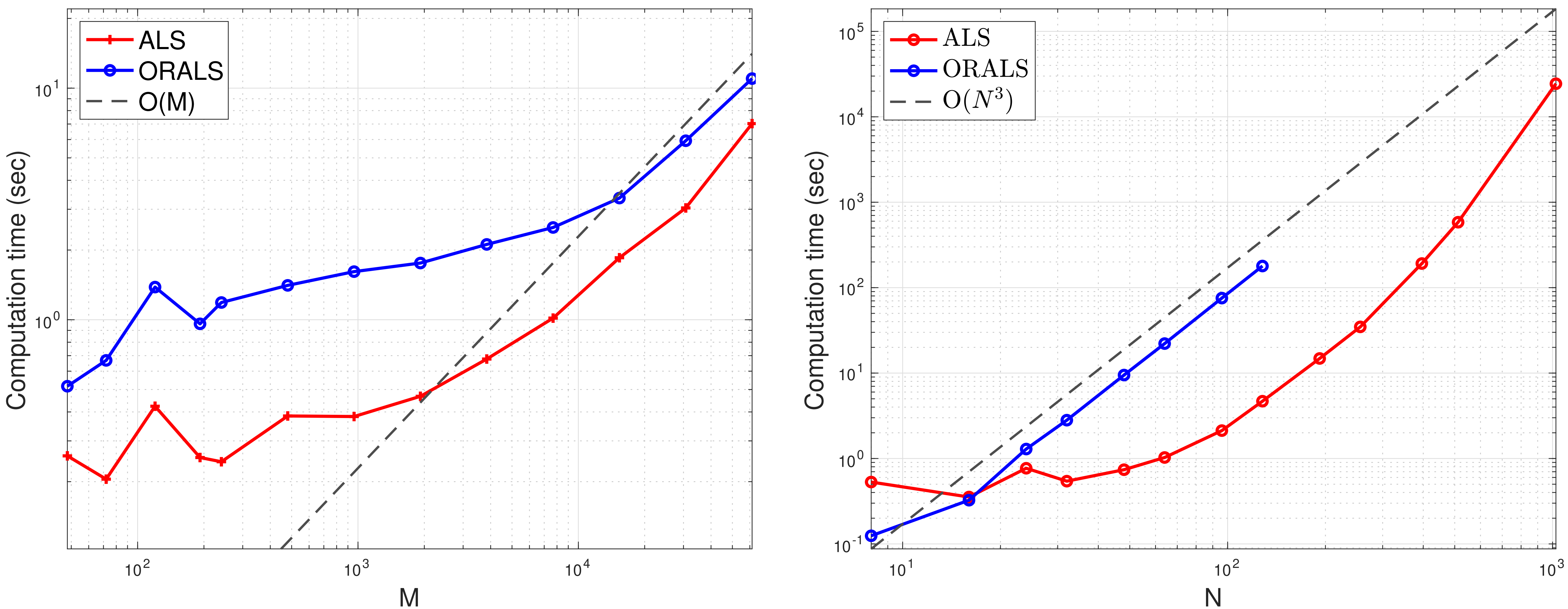
Convergence with sample size M increasing in 100 independent experiment runs. The top row shows almost perfect rates of $M^{-1/2}$ for both algorithms for the case of noiseless data and a well-specified basis. For the case of noisy observations, the bottom row shows robust convergence with the errors decaying until they reach 10^{-4}, the variance of observation noise.

Convergence & sampling



Top: Estimation errors as a function of M (all other parameters fixed), for ALS and ORALS, for a random Fourier interaction kernel with $p = 16$, $N = 32$, $L = 2$ (left) and $L = 8$ (right). In the small and medium sample regime, between the two vertical bars, ALS significantly and consistently outperforms ORALS; for large sample sizes, the two estimators have similar performance. Bottom: The performance of the ALS estimator improves not only as M increases but also as L increases.

Computational cost



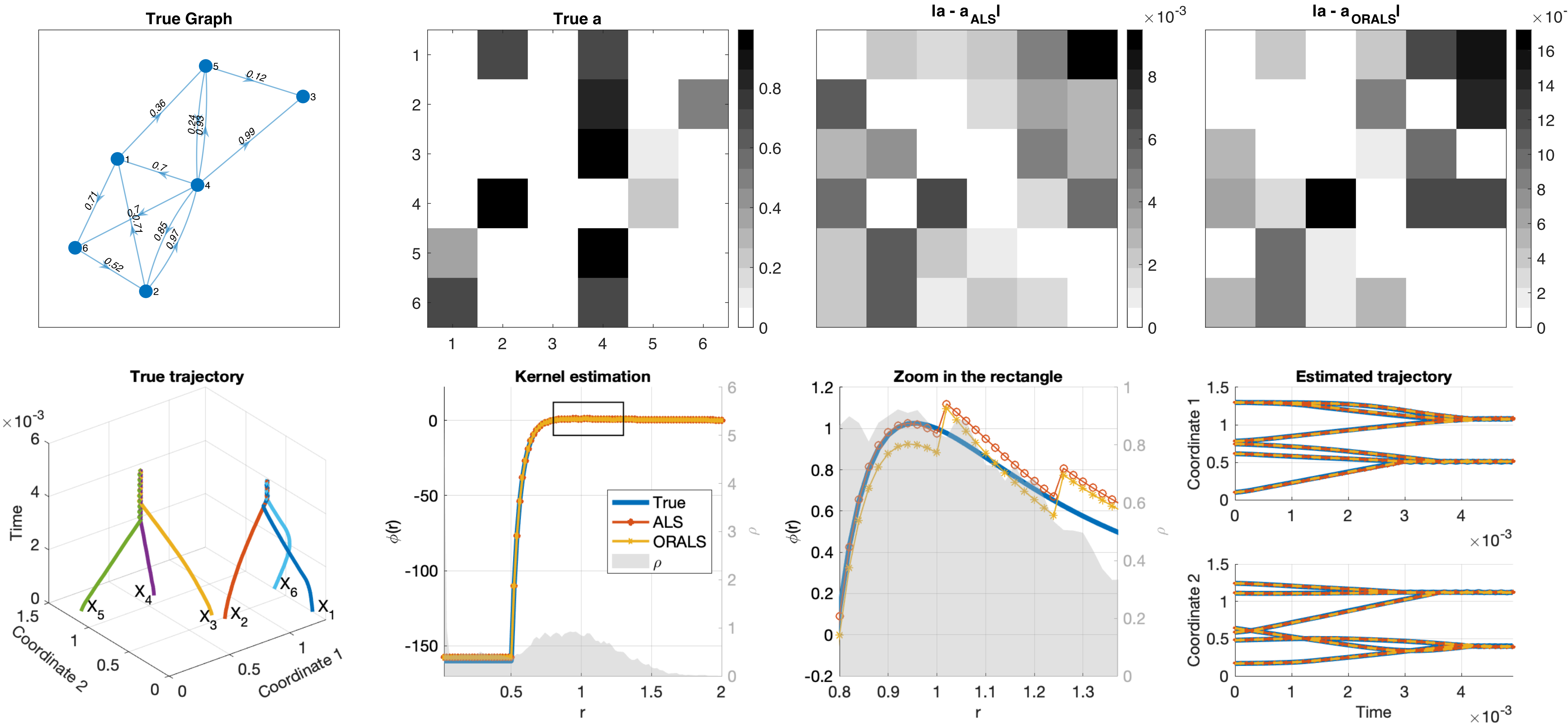
	ALS	ORALS
Assembling mats/vecs Solving	$O(MLdN^2p)$ $O(MLdN(p^2+N^2))$	$O(MLdN^3p^2)$ $O(MLdN^3+N^4p^3)$
Total (if $MLd > N$)	$O(MLdN(p^2+Np+N^2))$	$O(MLdN^3+N^4p^3)$

Lennard-Jones interactions on a network

$$\phi(x) = \left(-\tfrac{1}{3}x^{-9} + \tfrac{4}{3}x^{-3}\right)\mathbf{1}_{x\geq 0.5} - 160\,\mathbf{1}_{0\leq x<0.5}$$

M	N	p	L	T	σ	σ_{obs}
10^3	6	10	50	$5\cdot 10^{-3}$	10^{-3}	10^{-3}

$$\{\psi_{1+k} = x^{-9}\mathbf{1}_{[0.25k+0.5,+\infty]}\}_{k=0}^{2} \cup \{\psi_{4+k} = x^{-3}\mathbf{1}_{[0.25k+0.5,+\infty]}\}_{k=0}^{2} \cup \{\psi_{7+k} = \mathbf{1}_{[0,0.25k+0.5]}\}_{k=0}^{3}$$



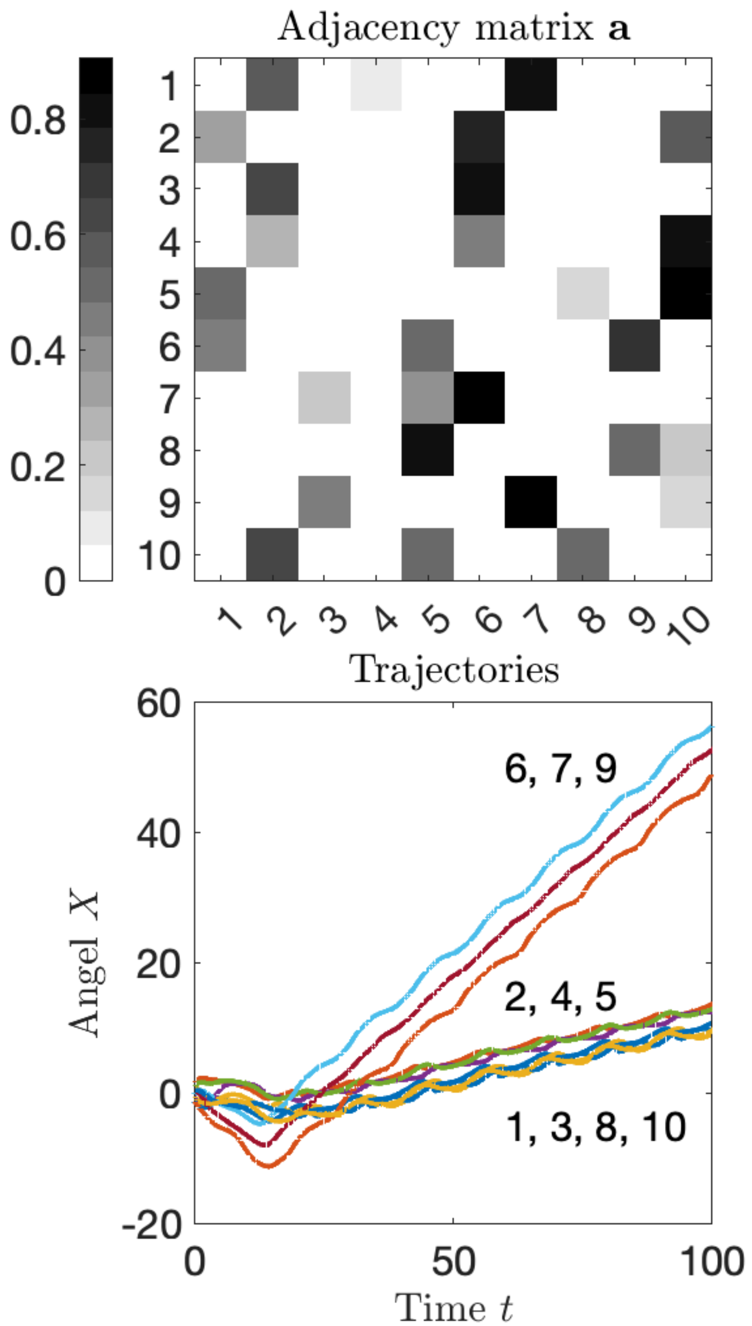
	Graph error $\varepsilon_{\mathbf{a}}$	Kernel error ε_K	Traj. error $\varepsilon_{\mathbf{X}}$	Exp. traj. error $\varepsilon_{\mathbf{X}}$
ALS	8.47×10^{-3}	1.45×10^{-2}	6.1×10^{-3}	$6.19\times10^{-3}\pm 8.12\times10^{-4}$
ORALS	1.67×10^{-2}	1.47×10^{-2}	6.6×10^{-3}	$7.41\times10^{-3}\pm 1.07\times10^{-3}$

Kuramoto interactions on a network

$$dX_t^i = \kappa \sum_{j \in \mathcal{N}_i} \mathbf{a}_{ij} \sin(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
$8, 64, 512$	10	100	$1 \cdot 10^{-1}$	10^{-4}	10^{-3}

$\mathcal{H} = \mathrm{span}\{\cos(x), \sin(2x), \cos(2x), \ldots, \cos(7x), \sin(7x)\}$, which does not contain Φ, and $\mathcal{H}_\phi := \mathrm{span}\{\mathcal{H}, \Phi\}$.

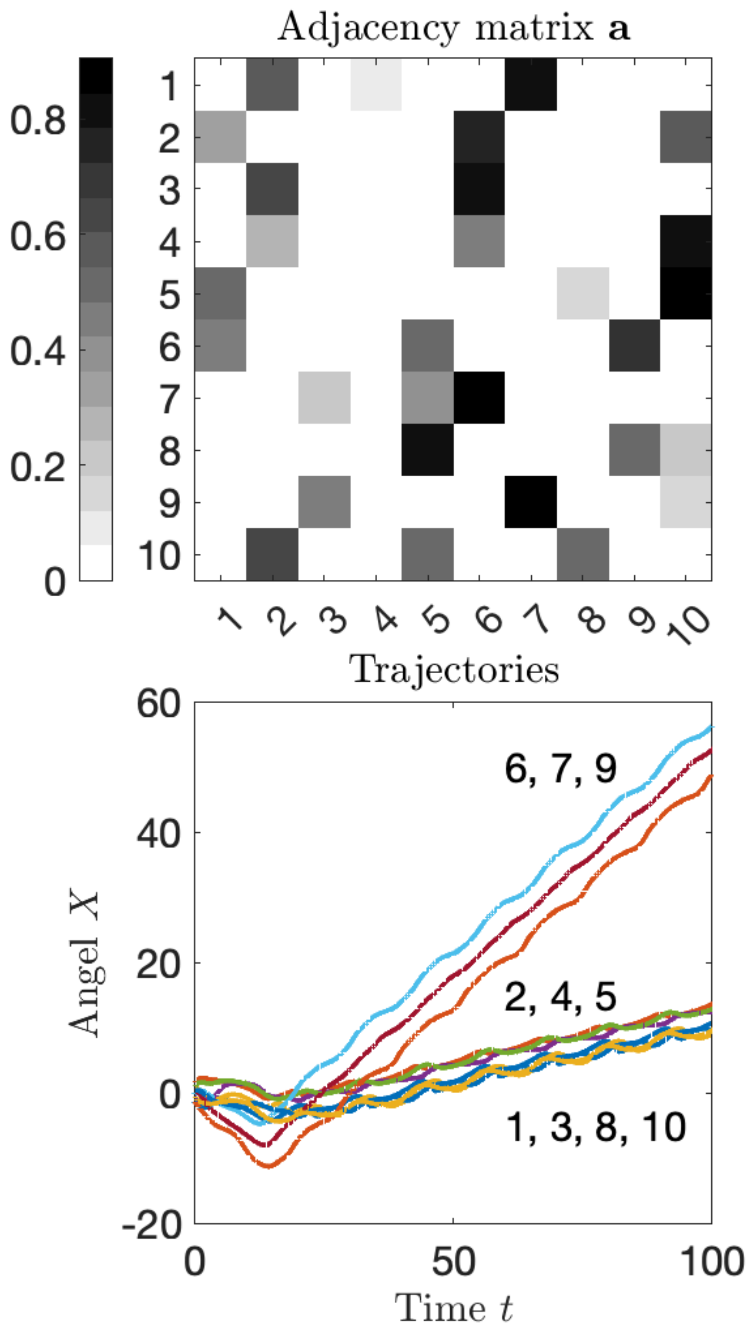


Kuramoto interactions on a network

$$dX_t^i = \kappa \sum_{j \in \mathcal{N}_i} \mathbf{a}_{ij} \sin(X_t^j - X_t^i) dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
$8, 64, 512$	10	100	$1 \cdot 10^{-1}$	10^{-4}	10^{-3}

$\mathcal{H} = \text{span}\{\cos(x), \sin(2x), \cos(2x), \ldots, \cos(7x), \sin(7x)\}$, which does not contain Φ, and $\mathcal{H}_\phi := \text{span}\{\mathcal{H}, \Phi\}$.

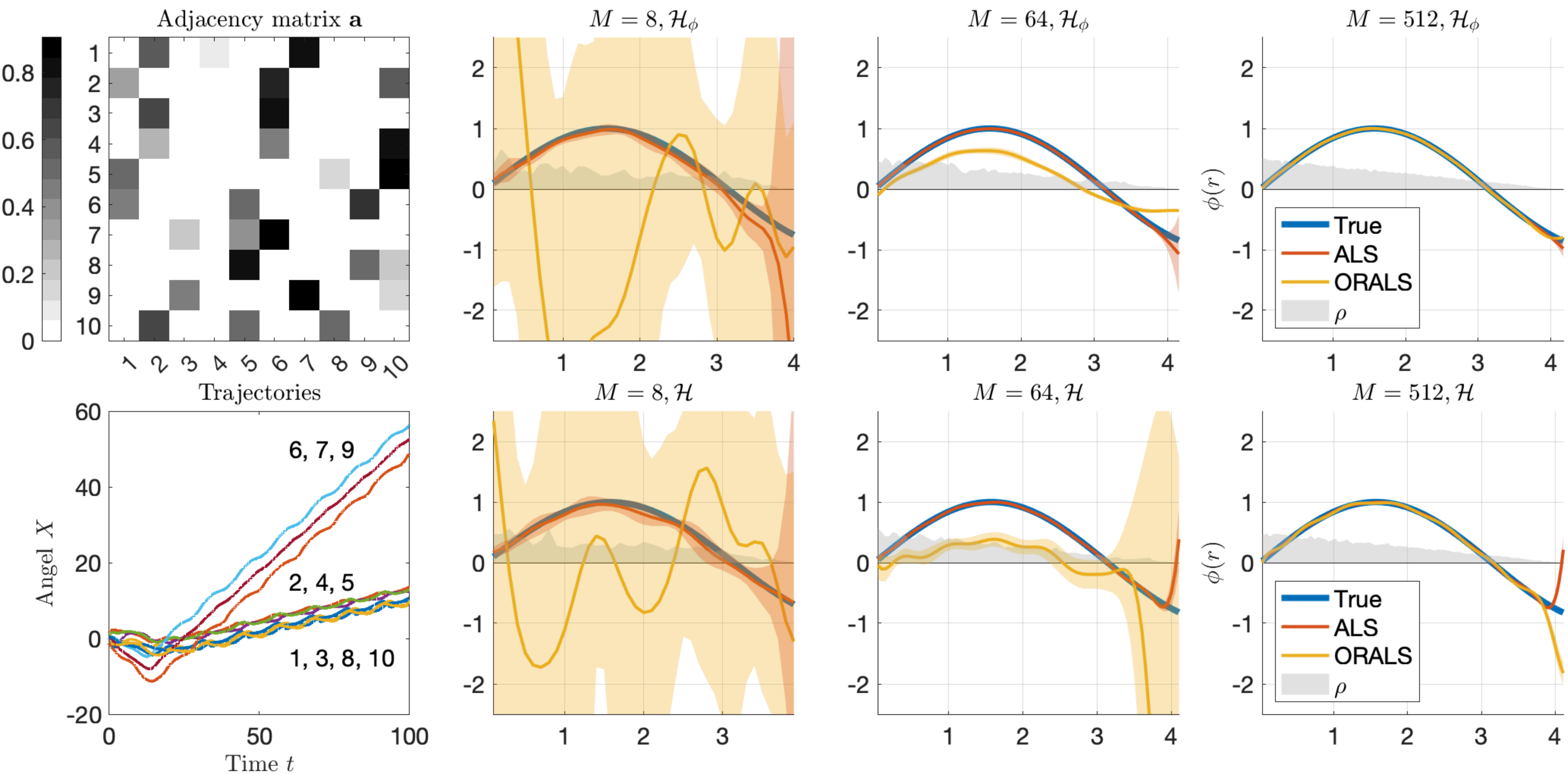


Kuramoto interactions on a network

$$dX_t^i = \kappa \sum_{j \in \mathcal{N}_i} \mathbf{a}_{ij} \sin(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
$8, 64, 512$	10	100	$1 \cdot 10^{-1}$	10^{-4}	10^{-3}

$\mathcal{H} = \text{span}\{\cos(x), \sin(2x), \cos(2x), \ldots, \cos(7x), \sin(7x)\}$, which does not contain Φ, and $\mathcal{H}_\phi := \text{span}\{\mathcal{H}, \Phi\}$.



Leader-follower opinion dynamics

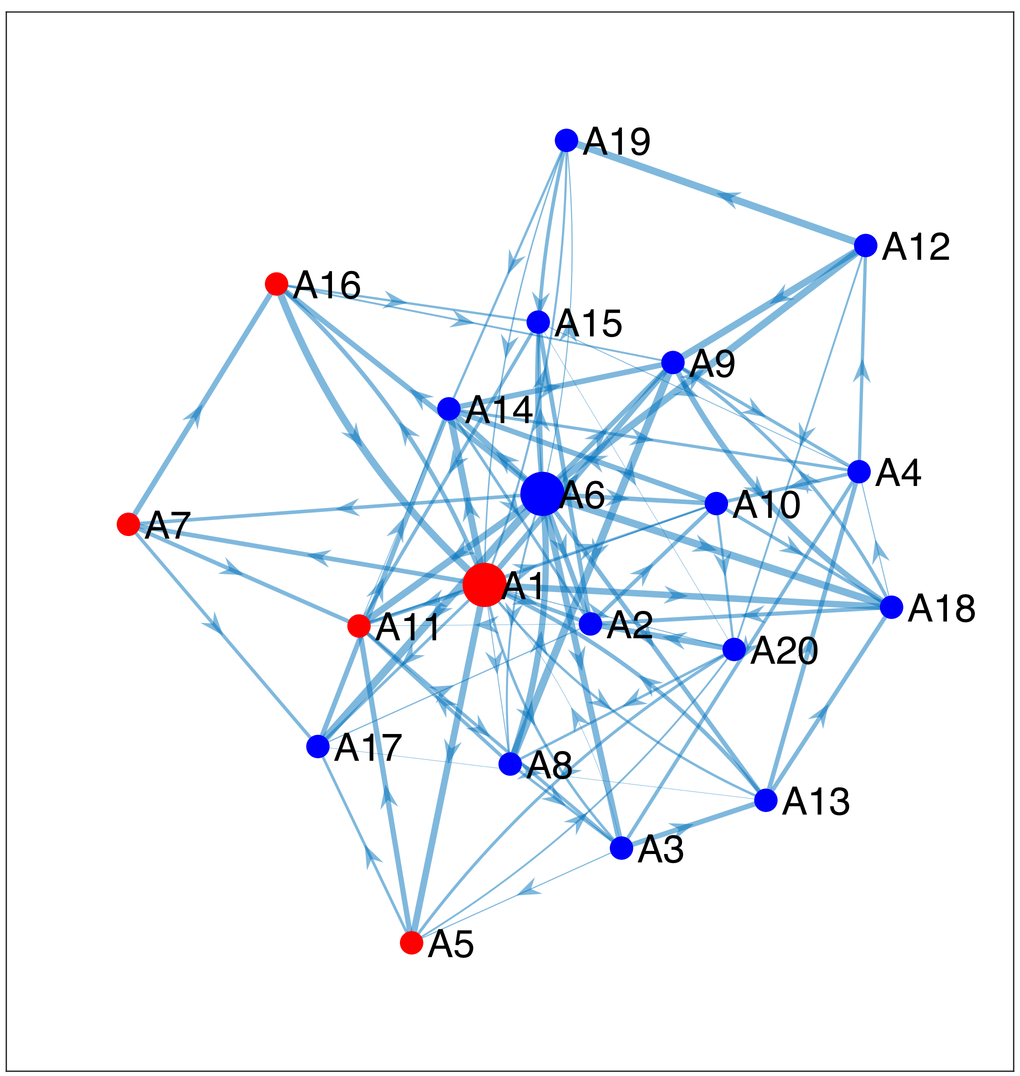
$$dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
$15, 30, 100$	20	100	1	0	0

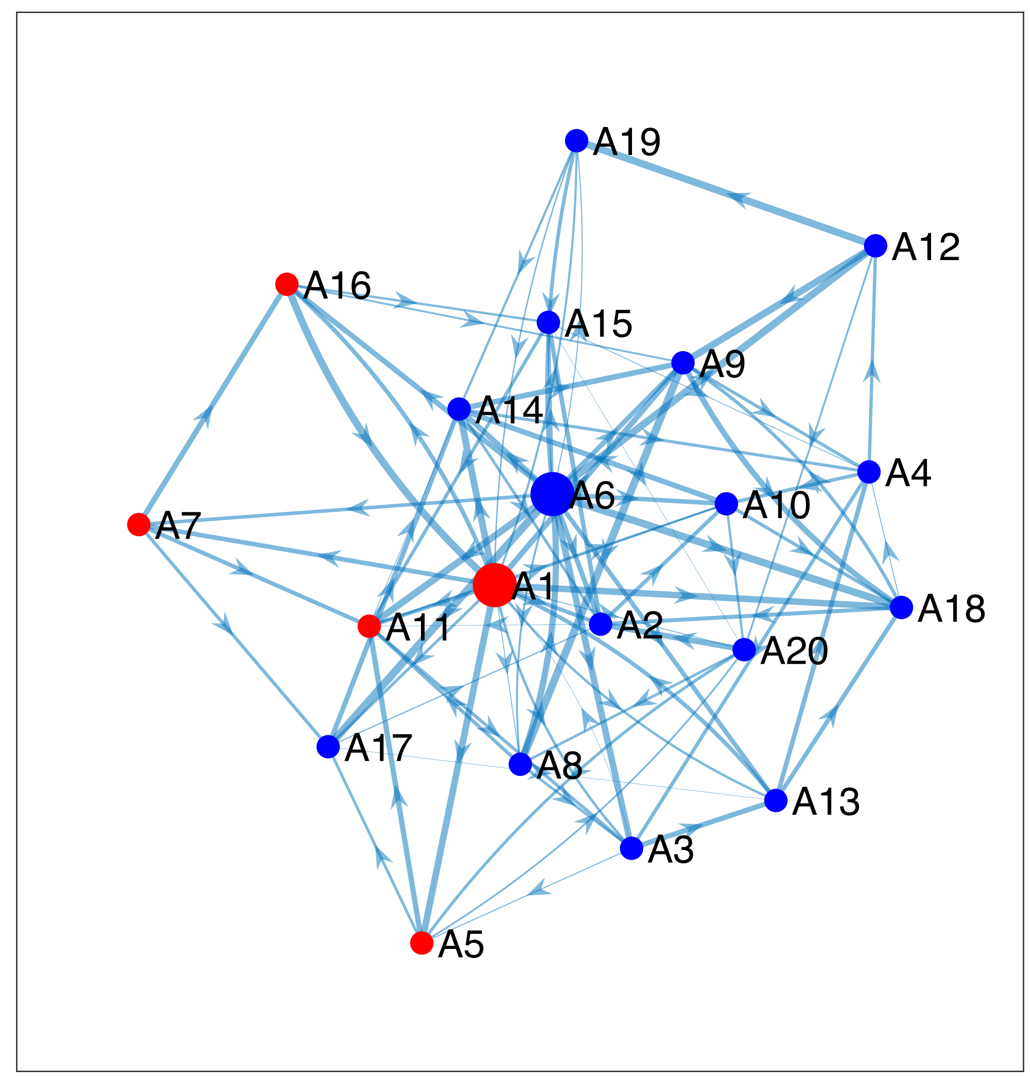
with $\Phi(x) = -\psi_1(x) - 0.1\psi_2(x)$, where $\psi_1(x) = \mathbf{1}_{\{x\leq 1\}}$, $\psi_2(x) = \mathbf{1}_{\{1<x\leq 1.5\}}$.

"Leaders": consider the feature $L_i = \alpha\|\mathbf{a}_{i\cdot}\|_{\ell_1} + \beta\|\mathbf{a}_{\cdot i}\|_{\ell_1}$, with $\alpha+\beta = 1$. With $\alpha \gg \beta$, clustering yields the set of "leaders".
"Followers": we group them by assigning them to leaders based on a score $\widetilde{L}_j^k = \alpha\sum_{i\in G^k}|\mathbf{a}_{ij}| + \beta\sum_{i\in G^k}|\mathbf{a}_{ji}|$ from groups G^k of "followers" to "leaders".

Leader-follower network M=100



Leader-follower network True



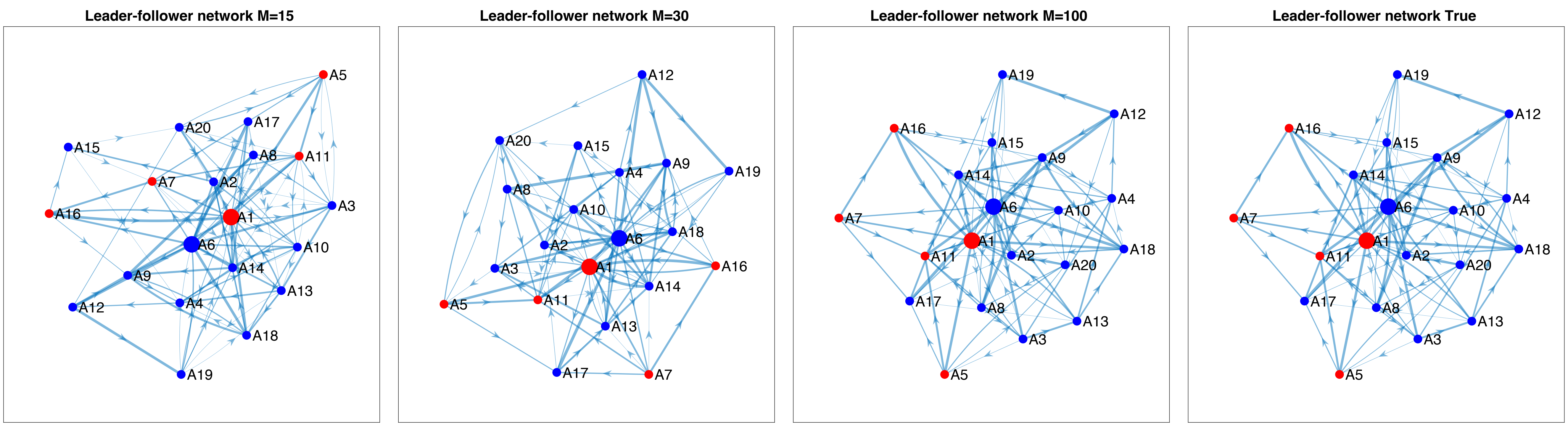
Leader-follower opinion dynamics

$$dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
$15, 30, 100$	20	100	1	0	0

with $\Phi(x) = -\psi_1(x) - 0.1\psi_2(x)$, where $\psi_1(x) = \mathbf{1}_{\{x\leq 1\}}$, $\psi_2(x) = \mathbf{1}_{\{1<x\leq 1.5\}}$.

"Leaders": consider the feature $L_i = \alpha\|\mathbf{a}_{i\cdot}\|_{\ell_1} + \beta\|\mathbf{a}_{\cdot i}\|_{\ell_1}$, with $\alpha+\beta = 1$. With $\alpha \gg \beta$, clustering yields the set of "leaders".
"Followers": we group them by assigning them to leaders based on a score $\widetilde{L}_j^k = \alpha\sum_{i\in G^k}|\mathbf{a}_{ij}| + \beta\sum_{i\in G^k}|\mathbf{a}_{ji}|$ from groups G^k of "followers" to "leaders".



Estimated networks of leaders and followers, with training sample sizes $M \in \{15, 30, 100\}$. When $M = 100$, the estimated network is accurate. When $M = 30$, the leaders-follower network is correctly identified, though the weight matrix is less accurate. When $M = 15$, the sample size is too small for a meaningful inference; but the clustering is still reliable.

Particles of different types

$$\mathcal{S}_{\mathbf{a},(\Phi_q)_{q=1}^{Q},\kappa}: \qquad dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi_{\kappa(i)}(X_t^j - X_t^i)dt + \sigma dW_t^i, \qquad i = 1,\ldots,N$$

$$\Phi_{\kappa(i)}(x) = \textstyle\sum_{k=1}^{p} \mathbf{c}_{ki}\psi_k(x) \qquad \mathbf{c}\in\mathbb{R}^{p\times N} \qquad \mathbf{c}_{\cdot,i} = \mathbf{c}_{\cdot,\kappa(i)}$$

$$\dot{\mathbf{X}}_t = \mathbf{a}\mathbf{B}(\mathbf{X}_t)\mathbf{c} + \sigma\dot{\mathbf{W}} = \big(\mathbf{a}_{i\cdot}\mathbf{B}(\mathbf{X}_t)_i\mathbf{c}_{\cdot i}\big)_{i\in[N]} + \sigma\dot{\mathbf{W}}$$

$$\mathbf{a}_{i\cdot}\mathbf{B}(\mathbf{X}_t)_i\mathbf{c}_{\cdot i} = \textstyle\sum_{j\neq i}\mathbf{a}_{ij}\sum_{k=1}^{p}\psi_k(X_t^j - X_t^i)c_{ki} \in \mathbb{R}^d\,, i = 1,\ldots,N$$

Writing $\mathbf{c} = \mathbf{u}\mathbf{v}^T$, with $\mathbf{u}\in\mathbb{R}^{p\times Q}$ the coefficient matrix, and $\mathbf{v}\in\mathbb{R}^{N\times Q}$ the type matrix, both orthogonal, we relax the problem to

$$\operatorname{argmin}_{\substack{(\mathbf{a},\mathbf{u},\mathbf{v})\in\mathcal{M}\times\mathbb{R}^{p\times Q}\times\mathbb{R}^{N\times Q}\\ \mathbf{v}^\top\mathbf{v}=I_Q}} \frac{1}{MT}\sum_{l=1,m=1}^{L,M}\left\|\Delta\mathbf{X}_{t_l}^m - \mathbf{a}\mathbf{B}(\mathbf{X}_{t_l}^m)\mathbf{u}\mathbf{v}^\top\Delta t\right\|_F^2$$

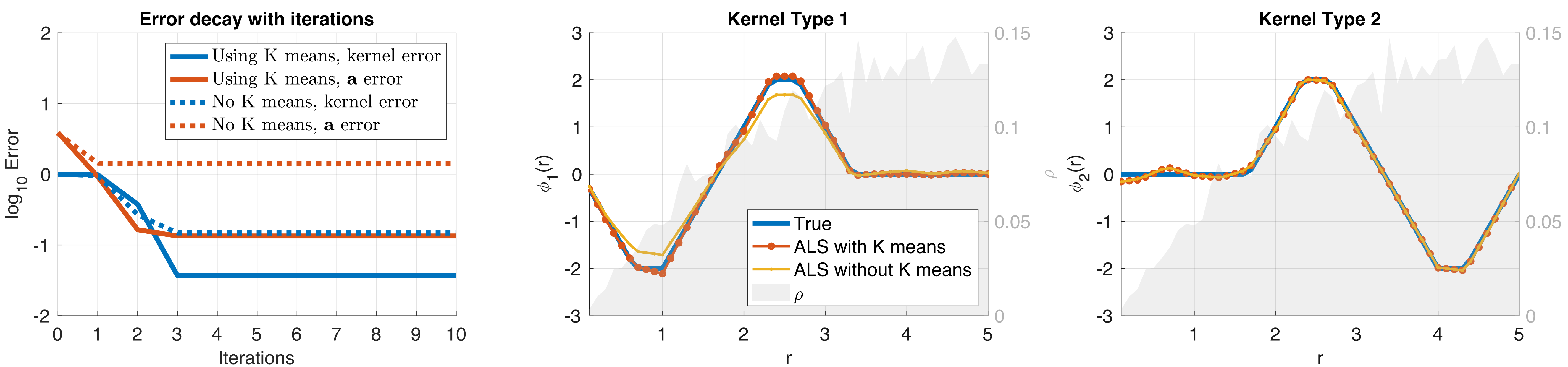
We use 3-way ALS to solve this problem; to enforce that $\mathbf{c}$ is not just low rank, but has only Q different columns, i.e. $\mathbf{c}_{\cdot i} = \mathbf{c}_{\cdot\kappa(i)}$, perform K-means on the cols($\mathbf{c}$) at every iteration.

Particles of different types: example

$$dX_t^i = \sum_{j \neq i} \mathbf{a}_{ij} \Phi_{\kappa(i)}(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
400	50	50	$5 \cdot 10^{-2}$	10^{-3}	10^{-3}

$\kappa : [N] \to [Q]$, with $Q = 2$, with Φ_1 short-range, and Φ_2 long-range.



Estimation of two types of kernels: short range and long range. The first panel shows the error decay with respect to iteration numbers. The algorithm using K-means decays faster and reaches lower errors than the algorithm without K-means. The right two columns show the estimation result of the two kernels. The classification is correct for both of the algorithms, and the one with K-means yields more accurate estimators, particularly for the kernel Type 1.

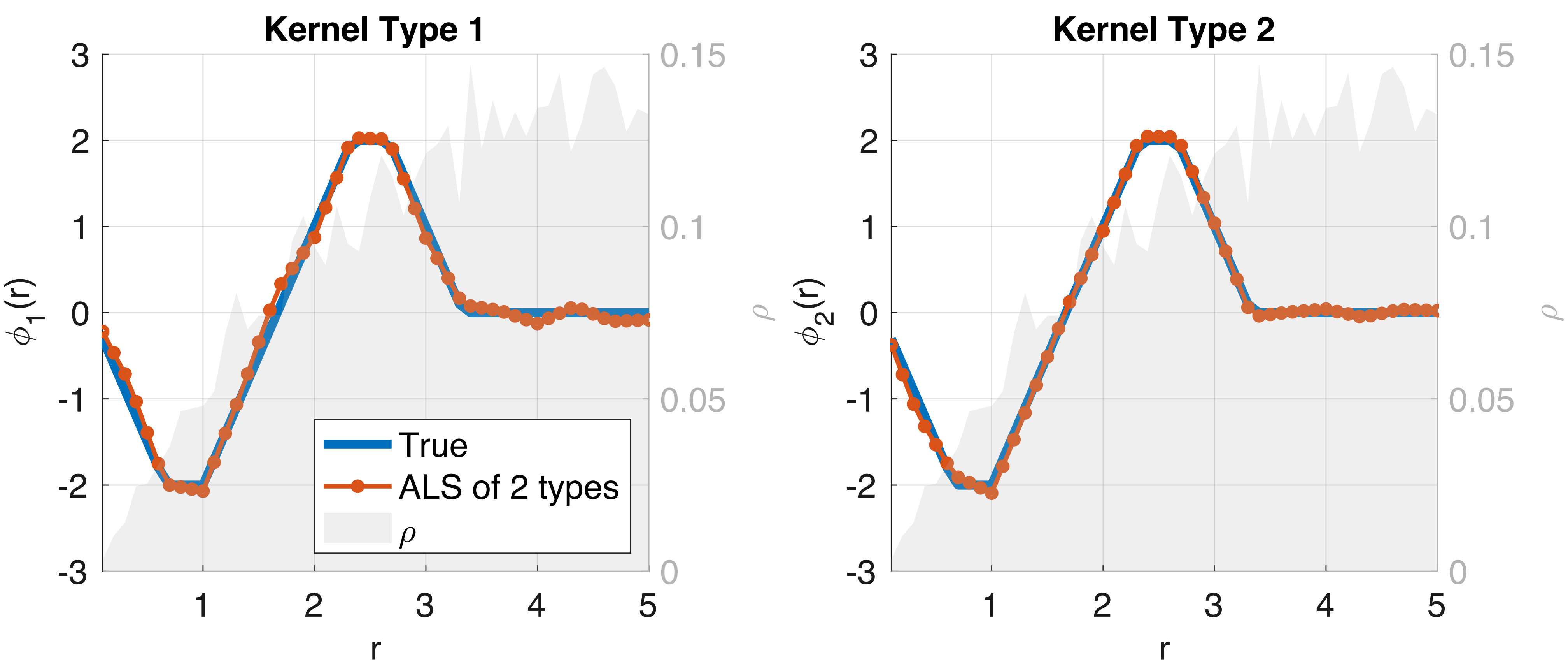
Particles of different types: model selection

$$dX_t^i = \sum_{j\neq i} \mathbf{a}_{ij}\Phi_{\kappa(i)}(X_t^j - X_t^i)dt + \sigma dW_t^i$$

M	N	L	T	σ	σ_{obs}
400	50	50	$5\cdot 10^{-2}$	10^{-3}	10^{-3}

$\kappa : [N] \to [Q]$, with $Q \in \{1,2\}$ interaction kernels. Estimators are constructed, in either case, with both $Q = 1$ and $Q = 2$. Their performance is evaluated with prediction error on long trajectories.

	$Q_{true} = 1$	$Q_{true} = 2$
Estimated with $Q = 1$	$\mathbf{1.2 \pm 0.8 \times 10^{-2}}$	$2.1 \pm 0.7 \times 10^{-1}$
Estimated with $Q = 2$	$1.4 \pm 0.7 \times 10^{-2}$	$\mathbf{1.1 \pm 0.3 \times 10^{-2}}$



Estimated kernels in a misspecified case: estimating two types of kernel when data is generated using a single kernel. Output consists of two types of kernels, but both are close to the true kernel.

Conclusions

· Learning interaction kernels in particle systems may be performed efficiently, nonparametrically, without curse of dimensionality of the state space...
· ...also on networks, with particles of different types, with interaction kernels, networks and types all unknown.
· Generalizations: 1st- and 2nd-order, multi-type, stochastic; learning variables; more general interaction kernels.

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

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

Learning Interaction Kernels for Agent Systems on Riemannian Manifolds, MM, J. Miller, H. Qui, M. Zhong, Proc. ICML 2021

Learning Interaction Variables and Kernels from Observations of Agent-Based Systems, J. Feng, MM, P. Martin, M. Zhong, Inter. Symp. Math. Th. Net. Sys., 2022

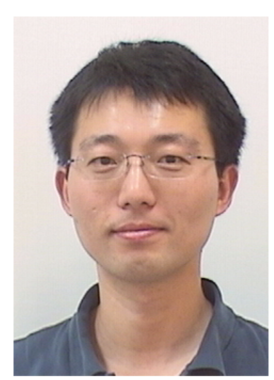
Learning interaction kernels in heterogeneous systems of agents from multiple trajectories,
F. Lu, MM, S. Tang, JMLR '21.

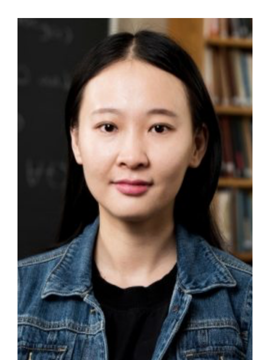
Learning theory for inferring interaction kernels in second-order interacting agent systems,
J.Miller, M. Zhong, S. Tang, M. Maggioni, arXiv 2010.03729

Learning interaction kernels in stochastic systems of interacting particles from multiple trajectories, F. Lu, M M. Maggioni, S. Tang, Found. Comp. Math., 2021

Interacting Particle Systems on Networks: joint inference of the network and the interaction kernel, Q. Lang, X. Wang, F. Lu, M. Maggioni, arXiv, 2024

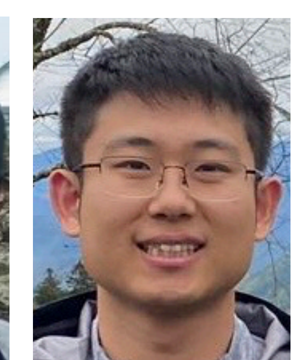
Links to code, papers:
https://mauromaggioni.duckdns.org


F. Lu


S. Tang


M. Zhong


X. Wang


Q. Lang