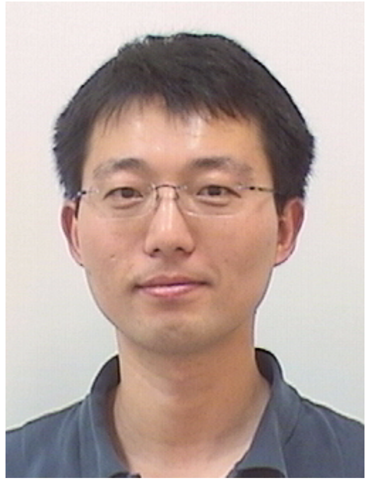
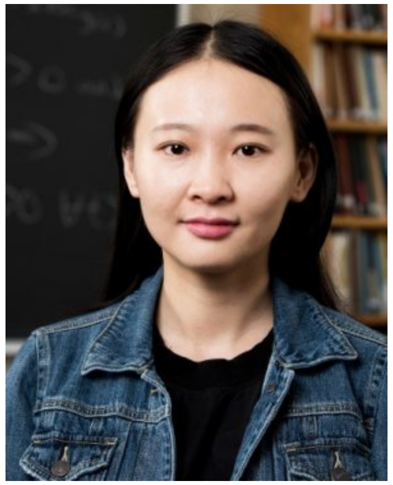
Learning Interaction kernels in agent-based systems

Mauro Maggioni

Joint work with Fei Lu, Sui Tang, Ming Zhong

Johns Hopkins University





Ki-Net/Duke workshop on Dimension reduction in physical and data sciences

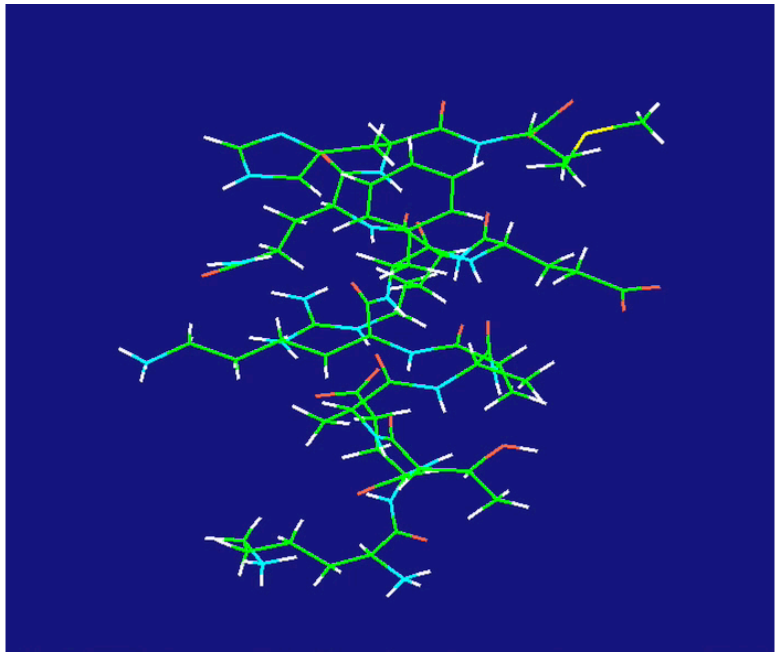
NSF, AFOSR

Machine Learning & Physical Systems

Many physical systems have very high-dimensional state spaces, are governed by a very large number of ODE's (or SDE's), which make them difficult to analyze. Homogenization, mean-field approximation, renormalization theory etc...are techniques to simplify such systems.

Physical systems with high-dimensional state spaces (e.g. many-particle systems) may exhibit behavior that is complex and high dimensional. Can Machine Learning help extract useful reductions?

Challenges for ML: learning principles that transfer across physical systems; incorporate existing physical knowledge or constraints.

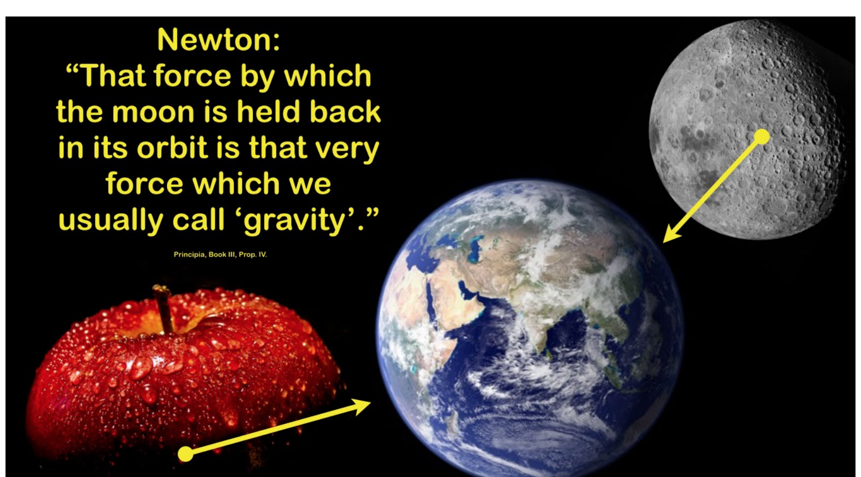


From https://www.youtube.com/watch?v=bb9ZTbYGRdc

Learning of Interaction Rules for agent-based systems

Given trajectories of dynamical system of interacting agents, learn the interaction rules. Applications: biological systems, particle systems.
Further goals: hypothesis testing for agent-based systems; transfer learning; agents on networks; collaborative and competitive games; learning dictionaries for complex dynamical systems.

Lots of recent interest in ML for learning ODE's and PDE's e.g. H. Shaeffer, N. Kutz, Y. Kevrekidis, R. Ward...



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

From https://www.youtube.com/watch?v=bb9ZTbYGRdc

Learning of Interaction Rules for agent-based systems

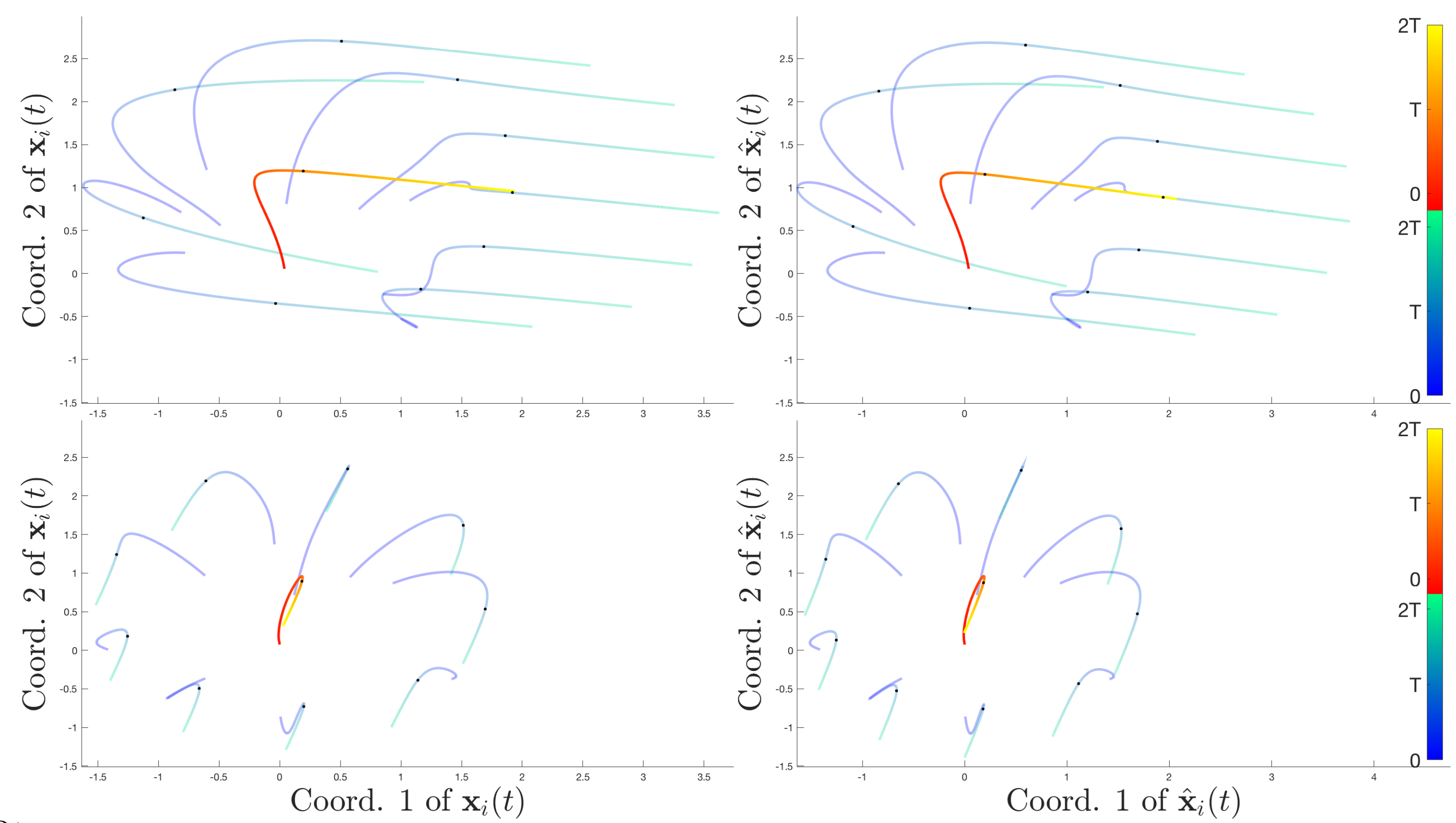
Given observations of the positions of agents $\{\mathbf{x}_i\}_{i=1}^N$ at different times $\{t_l\}_{l=1}^L$ and/or for different initial conditions $\{\mathbf{x}^{(m)}(0)\}_{m=1}^M$, evolving for example according to

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

we want to learn 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.

This was studied in *Inferring Interaction Rules from Observations of Evolutive Systems I: The Variational Approach*, M. Bongini, M. Fornasier, M. Hansen, and MM, M3S, 2017

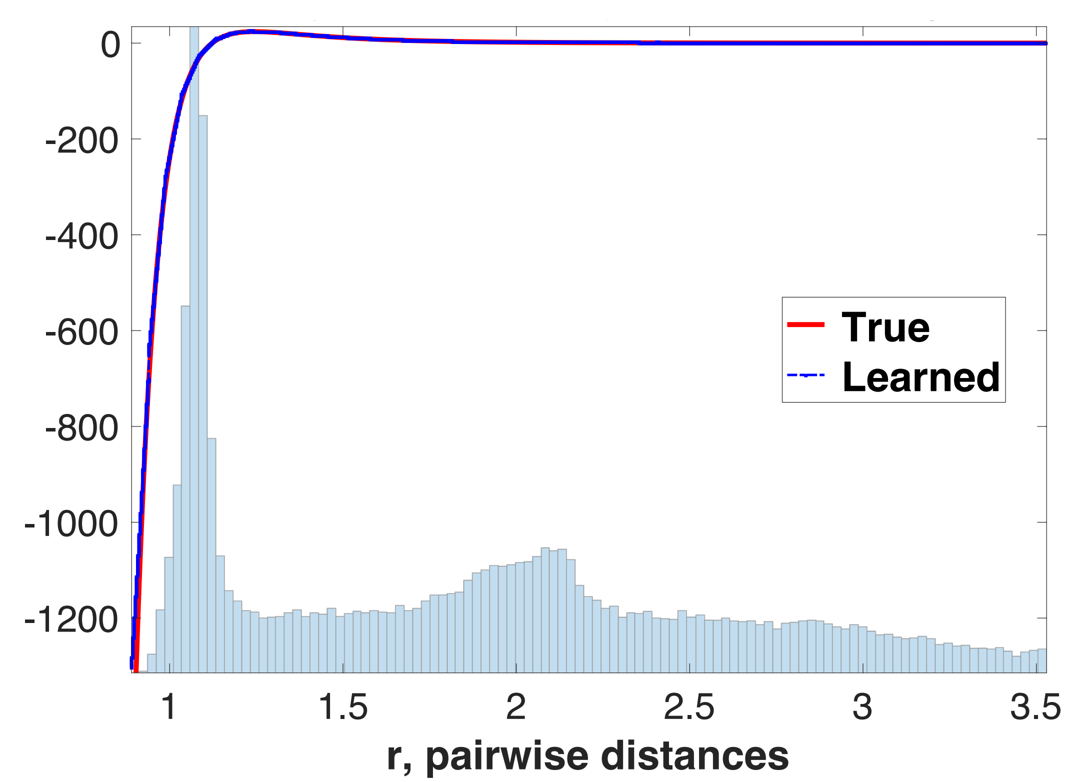
Learning the Interaction Kernel

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

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

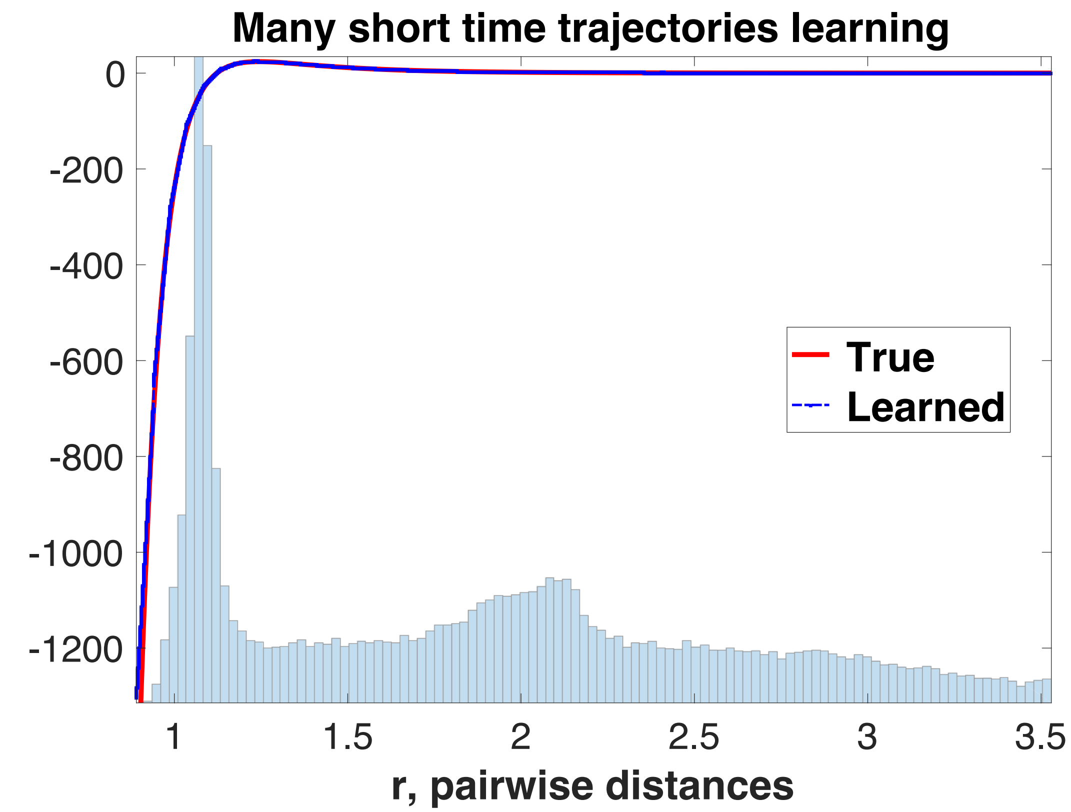
Fix the time scale $[0, T]$. We introduce the measure on $\mathbb{R}_+$ defined by $\rho_T^L(r) := \frac{1}{\binom{N}{2}L} \sum_{l=1}^{L} \mathbb{E}_{X(0)\sim\mu_0}\left[\sum_{i,i'=1,i<i'}^{N} \delta_{r_{ii'}(t_l)}(r)\right]$.

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, superimposed to an empirical estimate of ρ_T^L, for a system of $N = 7$ agents, and L, T small.

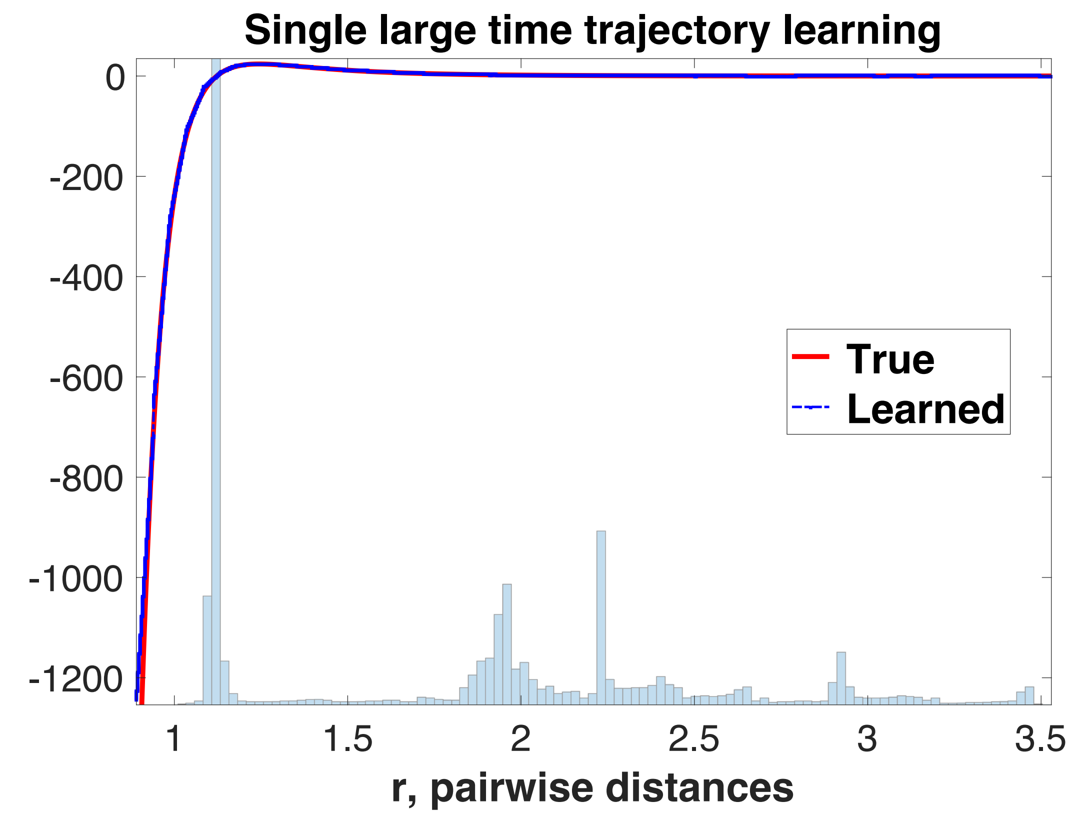


Example: L-J kernel and ρ_L^T

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, superimposed to an empirical estimate of ρ_T^L, for a system of $N = 7$ agents, and L, T small.



Example (cont'd). The measure ρ_T^L does depend on L and T.
With the same system as above, we consider here L and T large.
ρ_T^L is much more concentrated, due to the system approaching equilibrium..



The estimator

Observations: $\{(\mathbf{x}_i, \dot{\mathbf{x}}_i)(t_l)\}_{I=1,l=1}^{N,L}$, for M different IC's, from

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

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 simple hypothesis space of functions on $\mathbb{R}_+$, with dimension n (which will be chosen dependent on 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: $\forall \varphi \ : \ \varphi(\cdot)\cdot \in L^2(\rho_T^L)$, for $c_L > 0$

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

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

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

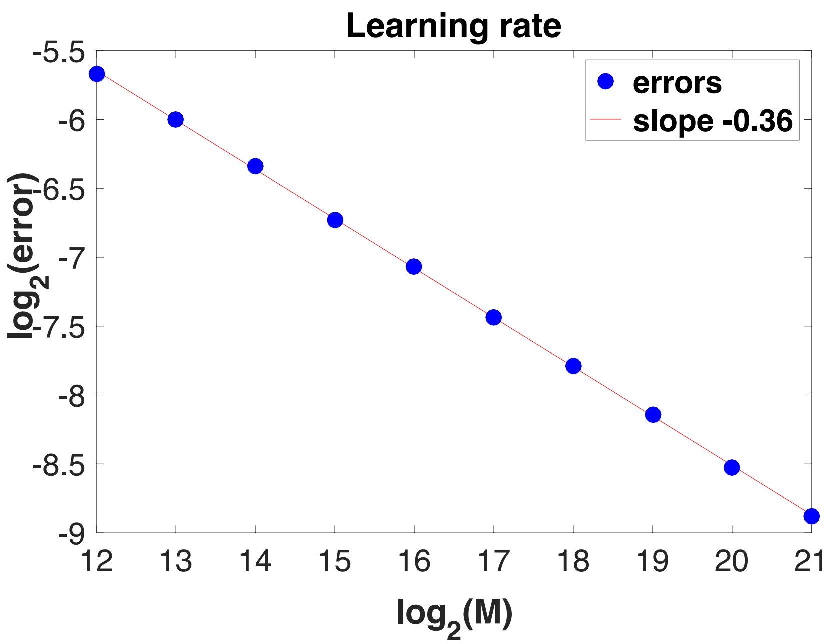
Main Theorem

Theorem. Let $\{\mathcal{H}_n\}_n$ 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 $C = C(c_0, c_1, R, S)$

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

- 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.



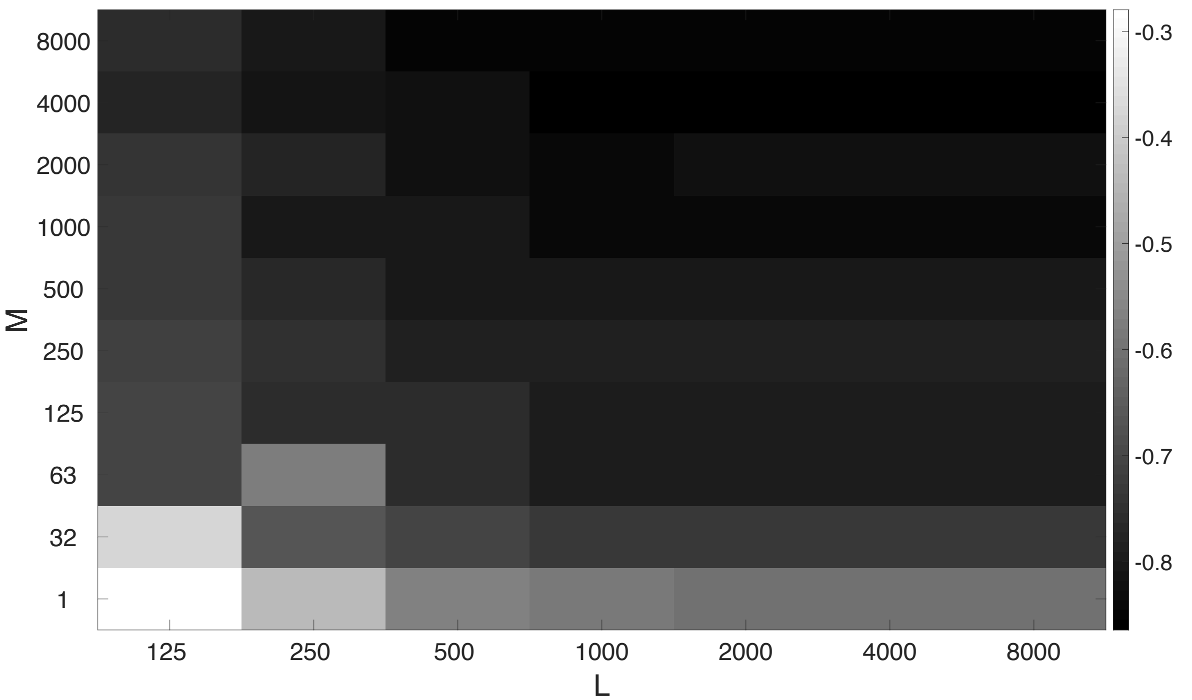
Main Theorem

Theorem. Let $\{\mathcal{H}_n\}_n$ 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 $C = C(c_0, c_1, R, S)$

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

- 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.

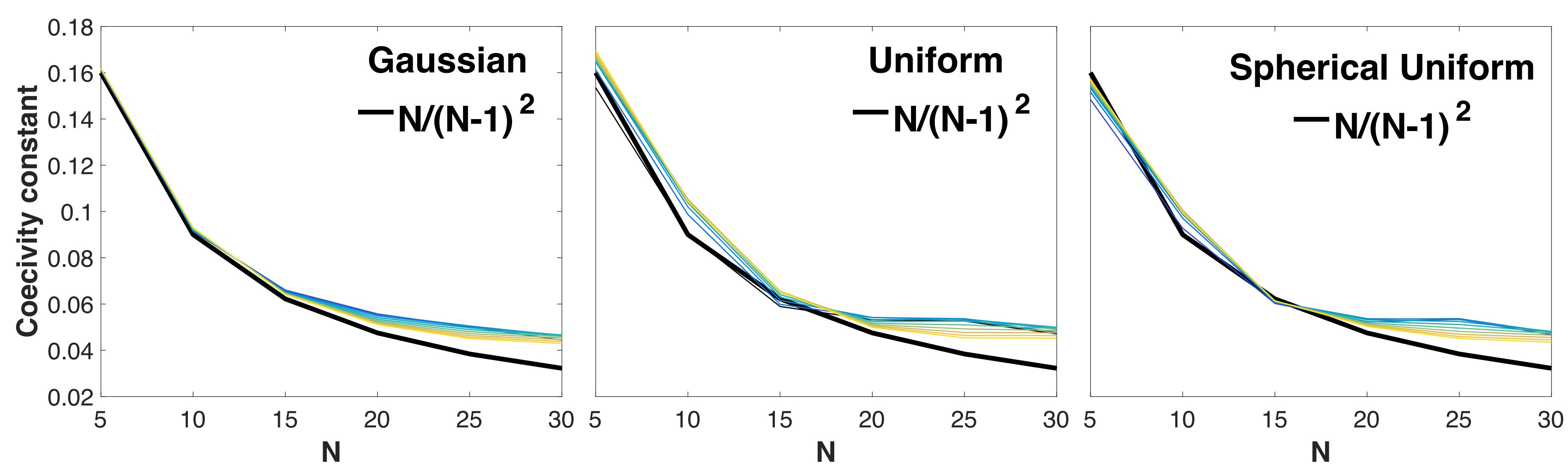


Main Theorem

Theorem. Let $\{\mathcal{H}_n\}_n$ 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 $C = C(c_0, c_1, R, S)$

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

c_L 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.



Errors on trajectories

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}.

On the bias variance tradeoff

Theorem. Let the true kernel $\phi \in \mathcal{K}_{R,S}$, and let $\mathcal{H}$ be a compact convex subset of $L^\infty([0,R])$, containing only functions bounded above by $S_0 \geq S$ a.e.. Assume that the coercivity condition. Then for any $\epsilon > 0$, the estimate

$$c_L \|\hat{\phi}_{L,M,\mathcal{H}}(\cdot)\cdot - \phi(\cdot)\cdot\|^2_{L^2(\rho_L^T)} \leq 2 \inf_{\varphi \in \mathcal{H}} \|\varphi(\cdot)\cdot - \phi(\cdot)\cdot\|^2_{L^\infty([0,R])} + 2\epsilon$$

holds true with probability at least $1-\delta$, provided that

$$M \geq \frac{1152 S_0^2 R^2}{c_T \epsilon} \big(\log(\mathcal{N}(\mathcal{H}, \frac{\epsilon}{48 S_0 R^2})) + \log(\frac{1}{\delta}) \big) \,,$$

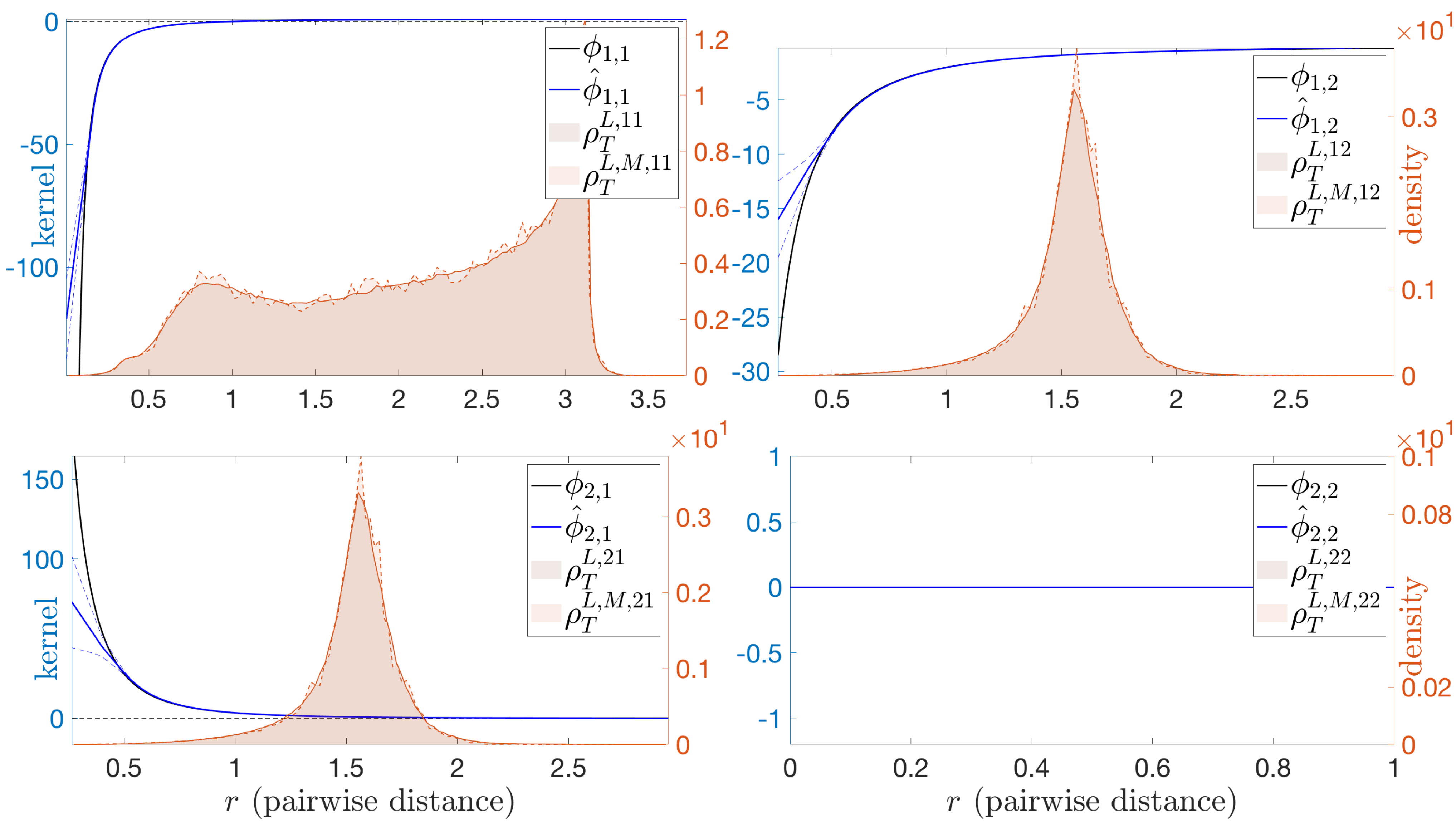
where $\mathcal{N}(\mathcal{H}, \eta)$ is the η-covering number of $\mathcal{H}$ under the ∞-norm.

We want to chose the best finite-dimensional $\mathcal{H}$ to minimize the error of the estimator. There are two competing issues. We want $\mathcal{H}$ to be large so that the bias term above is small. But we also want $\mathcal{H}$ to be small so that the covering number $\mathcal{N}(\mathcal{H}, \frac{\epsilon}{48_0 R^2})$ to be small. A balanced choice leads to the main theorem.

Examples: multi-type agents

We may extend to first order agent systems with multiple types of agents, with different interaction kernels for each directed pair of interactions.

$$\dot{\boldsymbol{x}}_i(t) = \sum_{i'=1}^{N} \frac{\kappa_{k_{i'}}}{N_{k_{i'}}} \phi_{k_i k_{i'}}(r_{ii'}(t)) \boldsymbol{r}_{ii'}(t)$$

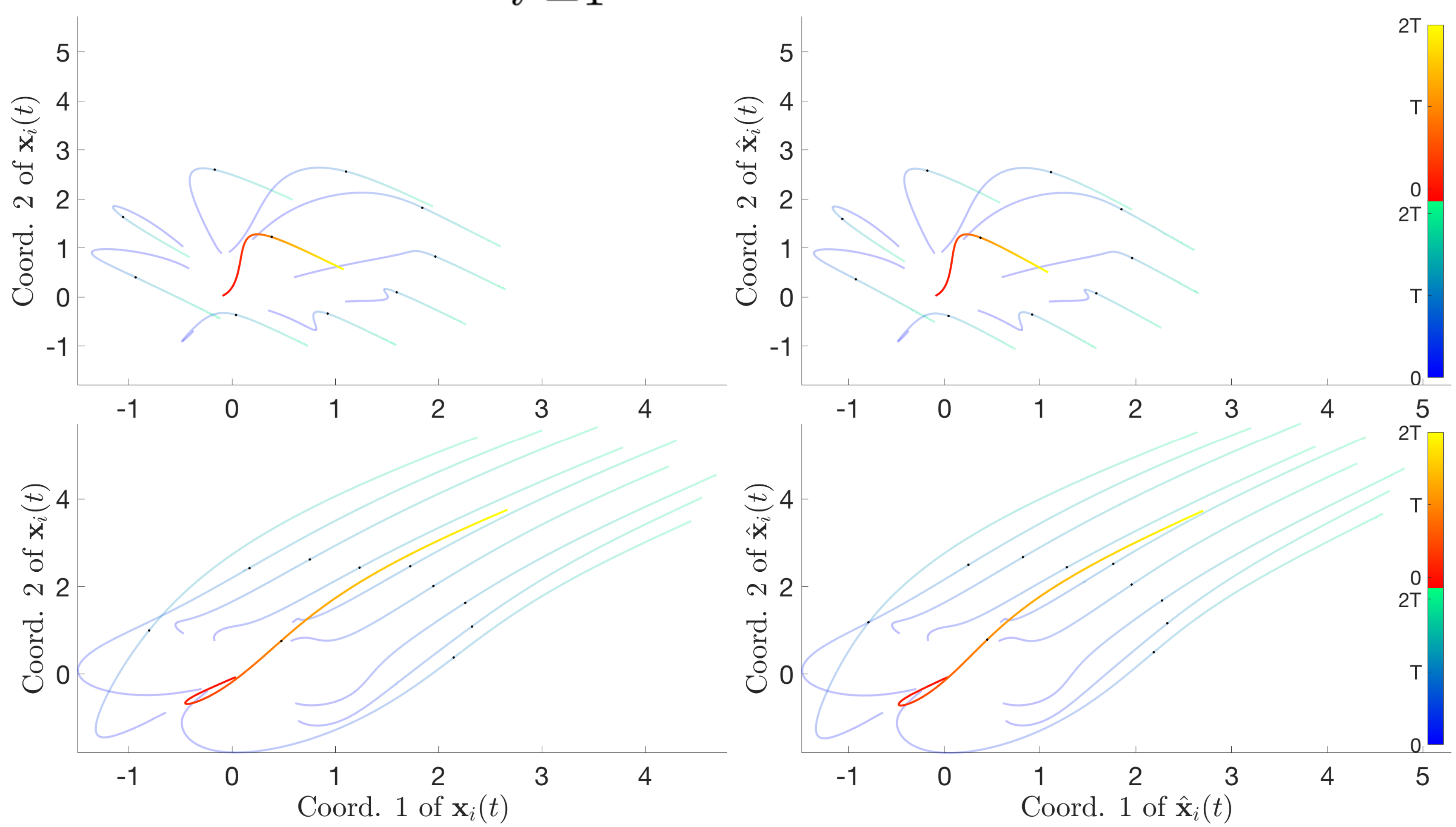


Example 1st order Prey-Predator system. The interaction kernels and ρ_L^T's.

Examples: multi-type agents

We may extend to first order agent systems with multiple types of agents, with different interaction kernels for each directed pair of interactions.

$$\dot{\boldsymbol{x}}_i(t) = \sum_{i'=1}^{N} \frac{\kappa_{k_{i'}}}{N_{k_{i'}}} \phi_{k_i k_{i'}}(r_{ii'}(t)) \boldsymbol{r}_{ii'}(t)$$

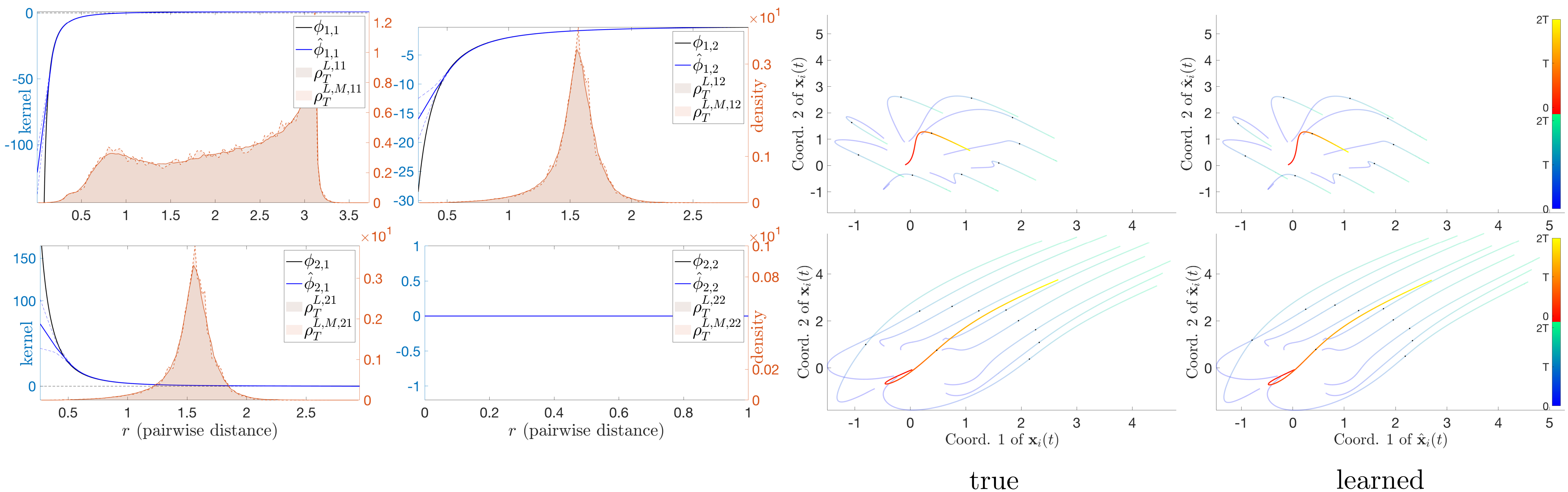


Example 1st order Prey-Predator system: trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

Examples: multi-type agents

We may extend to first order agent systems with multiple types of agents, with different interaction kernels for each directed pair of interactions.

$$\dot{\boldsymbol{x}}_i(t) = \sum_{i'=1}^{N} \frac{\kappa_{k_{i'}}}{N_{k_{i'}}} \phi_{k_i k_{i'}}(r_{ii'}(t)) \boldsymbol{r}_{ii'}(t)$$

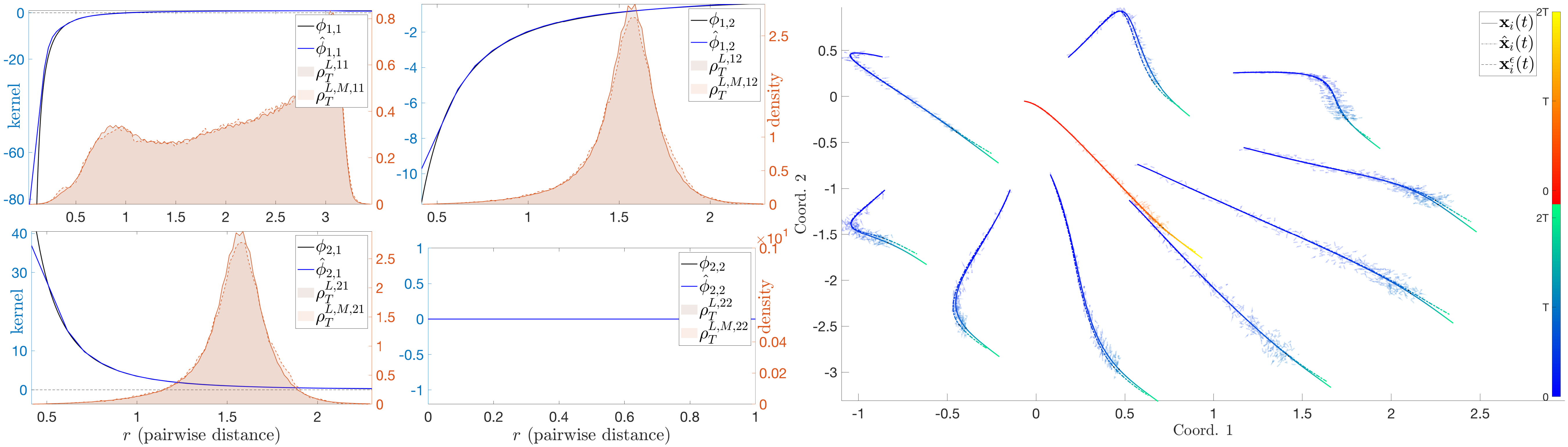


Example 1st order Prey-Predator system. Left: the interaction kernels and ρ_L^T's. Right: trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

Examples: multi-type agents + noise

We may extend to first order agent systems with multiple types of agents, with different interaction kernels for each directed pair of interactions.

$$\dot{\boldsymbol{x}}_i(t) = \sum_{i'=1}^{N} \frac{\kappa_{k_{i'}}}{N_{k_{i'}}} \phi_{k_i k_{i'}}(r_{ii'}(t)) \boldsymbol{r}_{ii'}(t)$$

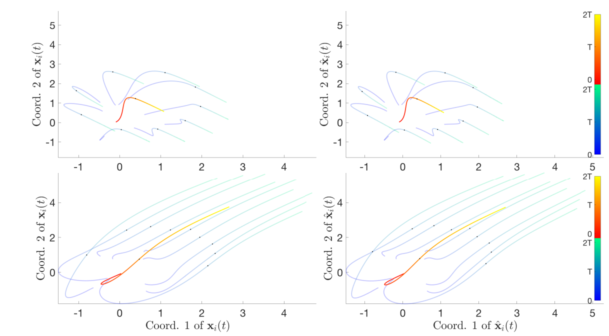


Example 1st order Prey-Predator system + noise: multiplicative noise $\sim \frac{1}{10}\mathrm{Unif}[-\frac{1}{2}, \frac{1}{2}]$ is added to observed positions and velocities.

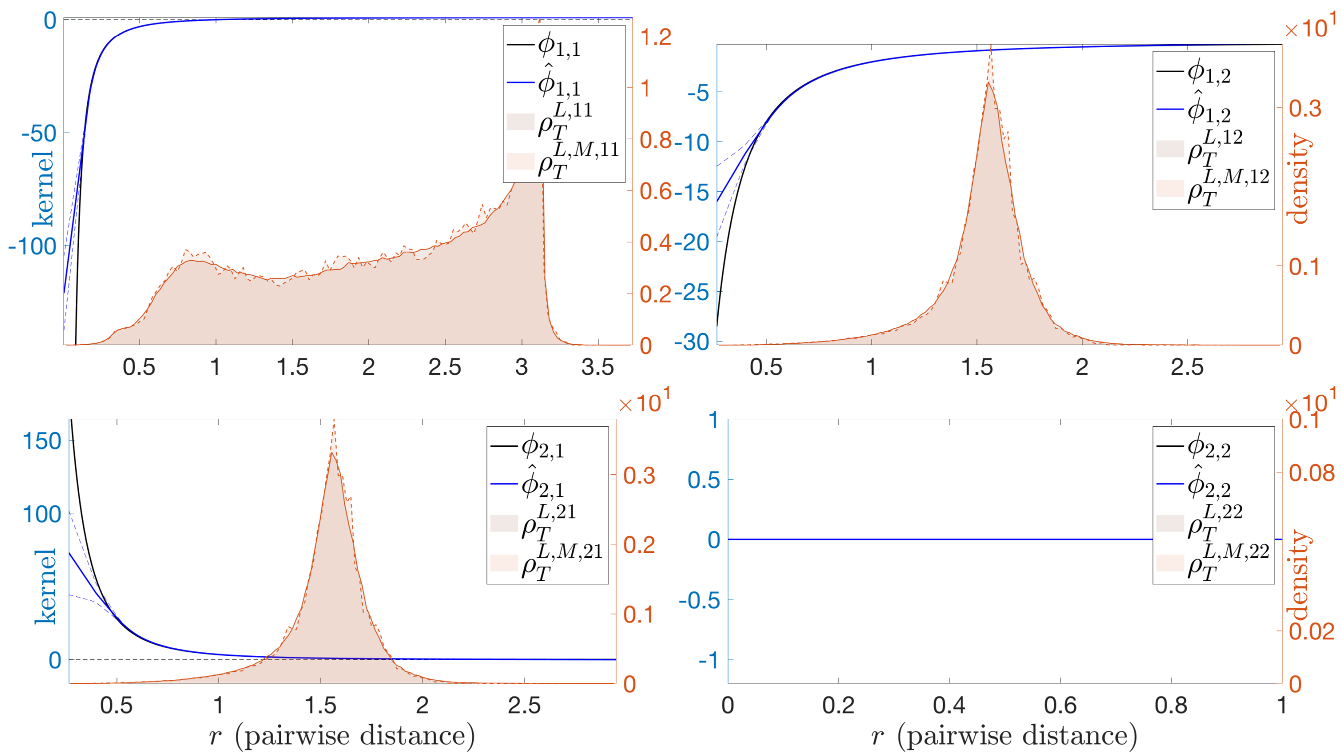
Examples: multi-type agents + scaling N

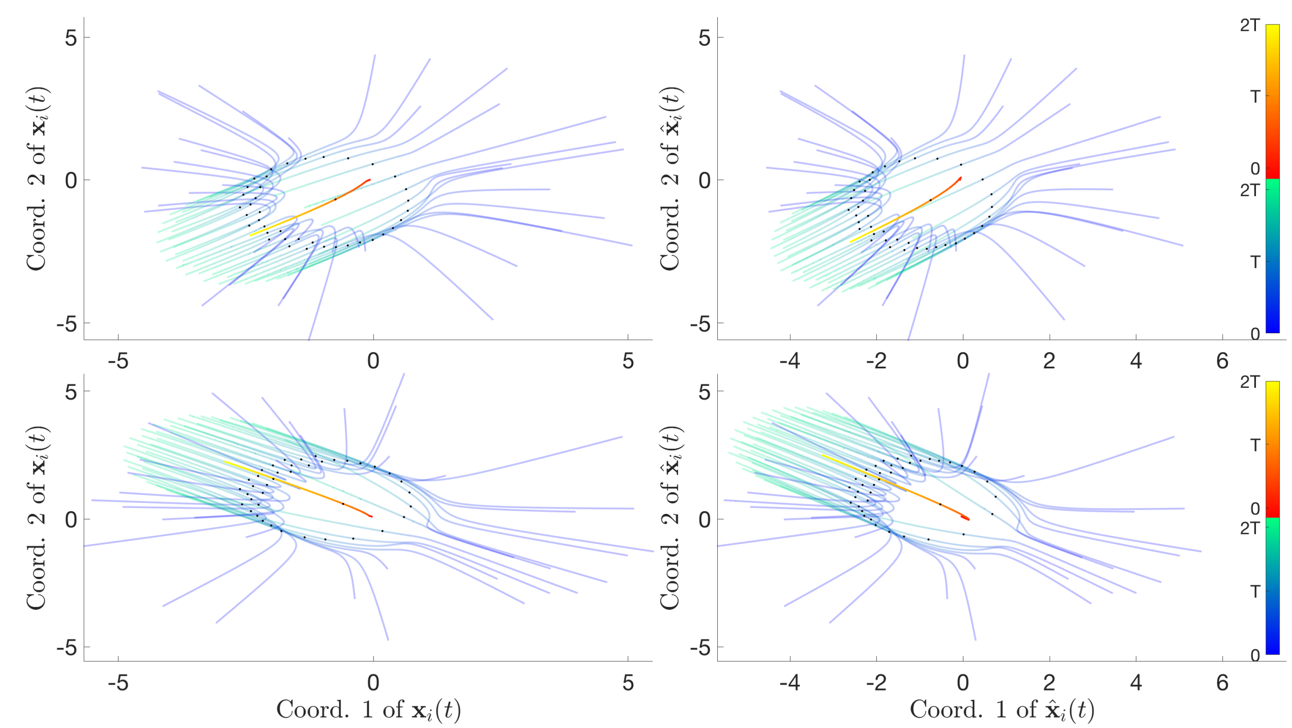
We may extend to first order agent systems with multiple types of agents, with different interaction kernels for each directed pair of interactions.

$$\dot{\boldsymbol{x}}_i(t) = \sum_{i'=1}^{N} \frac{\kappa_{k_{i'}}}{N_{k_{i'}}} \phi_{k_i k_{i'}}(r_{ii'}(t))\boldsymbol{r}_{ii'}(t)$$



$N \mapsto 4N$

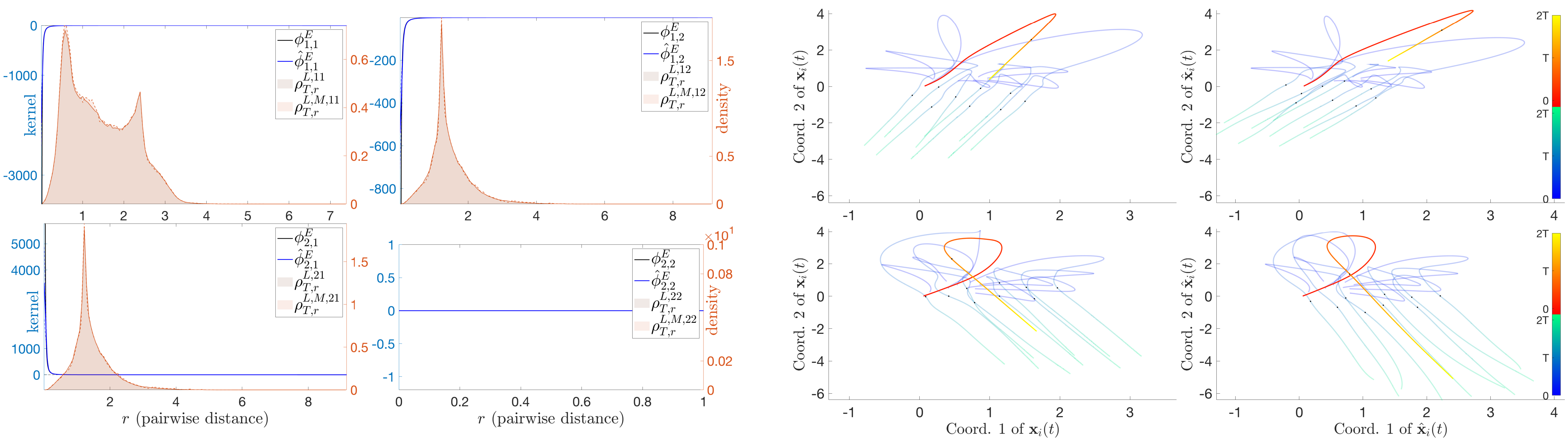




Example 1st order Prey-Predator system. Left: the interaction kernels and ρ_L^T's. Right: trajectories of the true system (left col.) and learned system (right col.) with an initial condition from training data (top) and a new one (bottom).

Examples: 2nd order systems

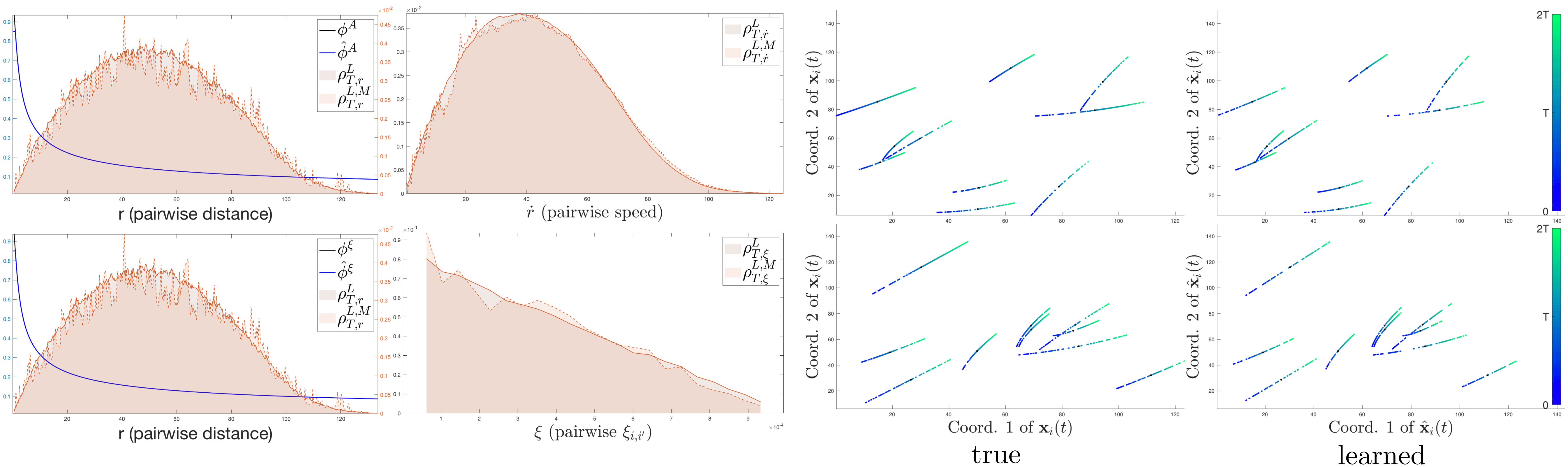
$$\begin{cases} m_i \ddot{\boldsymbol{x}}_i = F_i^{\boldsymbol{v}}(\dot{\boldsymbol{x}}_i, \xi_i) + \displaystyle\sum_{i'=1}^{N} \frac{\kappa^{\boldsymbol{v}}_{\mathcal{k}_{i'}}}{N_{\mathcal{k}_{i'}}} \left(\phi^E_{\mathcal{k}_i \mathcal{k}_{i'}}(r_{ii'})\boldsymbol{r}_{ii'} + \phi^A_{\mathcal{k}_i \mathcal{k}_{i'}}(r_{ii'})\dot{\boldsymbol{r}}_{ii'}\right)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



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).

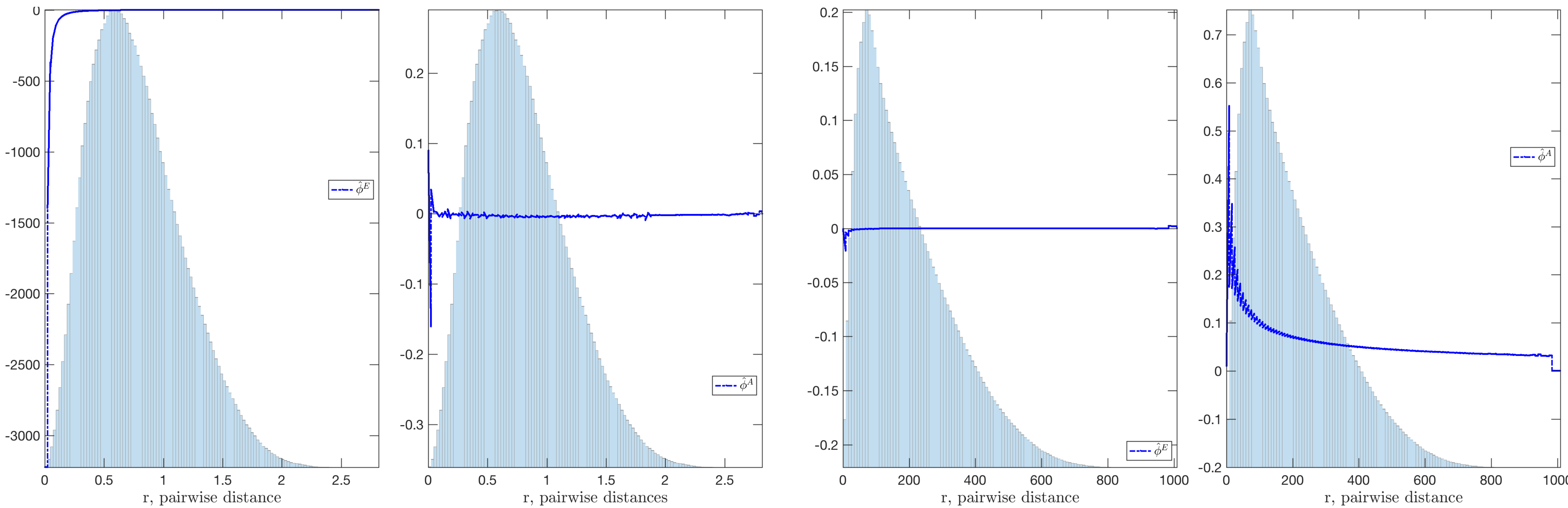
Example with environment (phototaxis)

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



Example 2nd order Phototaxis model, which includes an environment modeling light, interacting with the agents. 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).

Testing hypotheses for agent systems



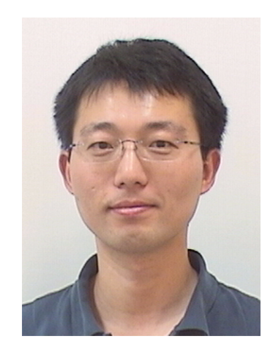
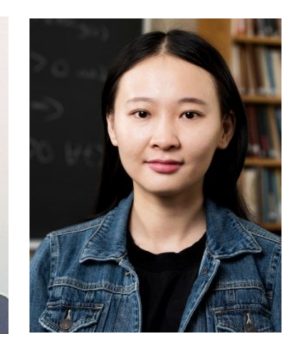
Example We want to test if a 2nd order system is driven by energy or alignment interactions. Left: we learn a general model (with both types of interaction) on a system with only energy interaction terms: we obtain $\hat{\phi}^A$ is $\approxeq 0$. Right: learning on a system with only alignment term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

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