



Active Symbolic Discovery of Ordinary Differential Equations via Phase Portrait Sketching

Nan Jiang^{*}, Md Nasim^{*}, Yexiang Xue^{*}

^{*} Purdue University, ^{*} Cornell University

Email: jiang631@purdue.edu

Preliminaries

Ordinary Differential Equations (ODEs) describe the evolution of dynamical systems in continuous time. The temporal evolution of the system is

$$\frac{dx_i}{dt} = f_i(\mathbf{x}(t), \mathbf{c}), \quad \text{for } i = 1, \dots, n,$$

- Vector $\mathbf{x}(t) = (x_1(t), \dots, x_n(t)) \in \mathbb{R}^n$ is the time-dependent state variables;
- Function f_i can be a linear or nonlinear of variables $\mathbf{x}$ and coefficients $\mathbf{c}$;
- Time derivatives of the state variables dx_i/dt

The trajectory $\tau := (\mathbf{x}_0, \mathbf{x}(t_1), \dots, \mathbf{x}(t_k))$ is the sequence of state variables observed at discrete time $(t_1, \dots, t_k)$, possibly with noise, starting at $\mathbf{x}_0$.

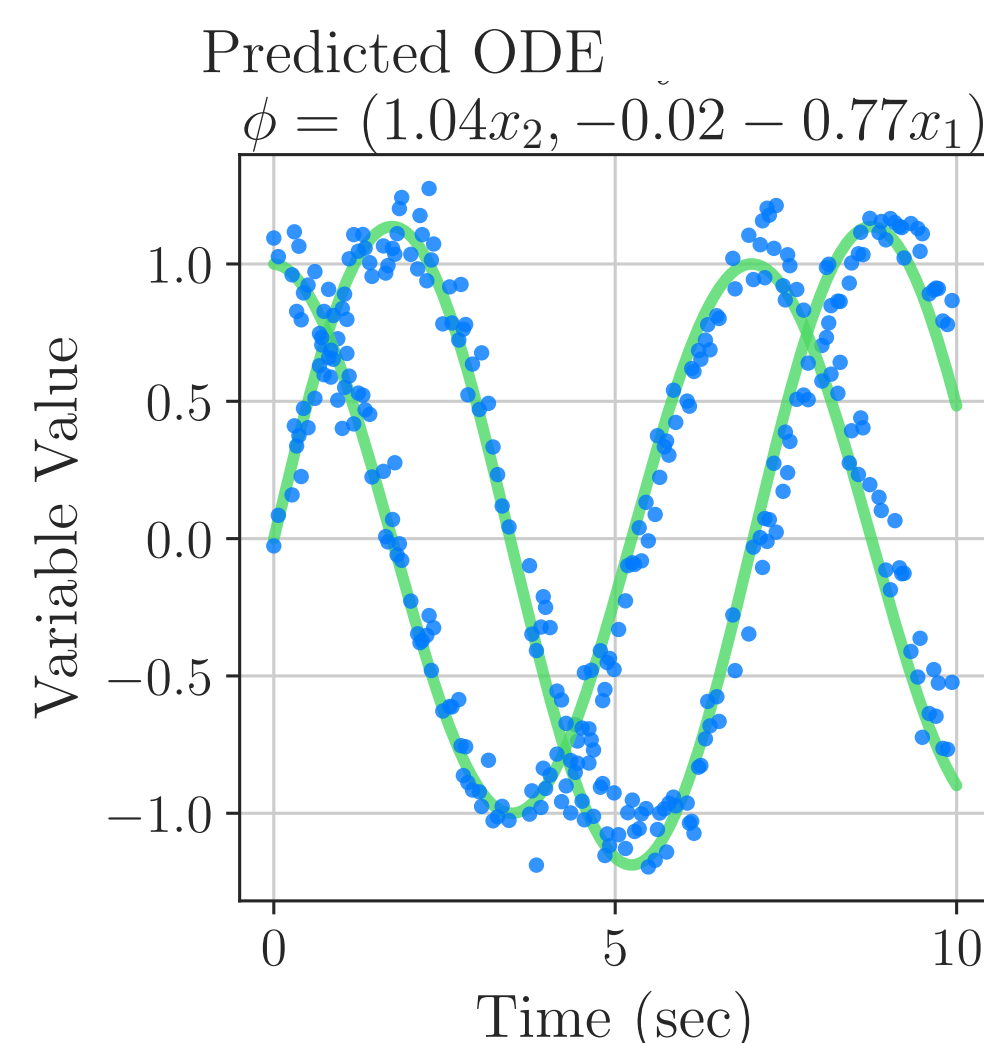
Symbolic Discovery of ODEs Given

- A dataset $D = \{\tau_1, \dots, \tau_N\}$, where each τ_i is a ground-truth trajectory
- A set of mathematical operators $\{+, -, \times, \div, \sin, \log, \dots\}$

Objective: predict the best symbolic form of ODE that attain minimum loss on the data:

$$\arg \min_{\phi \in \Pi} \sum_{\tau \in D} \sum_{i=1}^k L(\mathbf{x}(t_i), \hat{\mathbf{x}}(t_i)), \quad \text{where } \hat{\mathbf{x}}(t_i) = \mathbf{x}_0 + \int_0^{t_i} \phi(\mathbf{x}(t), \mathbf{c}) dt$$

- Predicted trajectory $\mathbf{x}_0, \hat{\mathbf{x}}(t_1), \dots, \hat{\mathbf{x}}(t_k)$;
- Ground-truth trajectory $\mathbf{x}_0, \mathbf{x}(t_1), \dots, \mathbf{x}(t_k)$;
- Loss function L can be normalized mean squared error.



The GAP: model using pre-collected dataset are limited

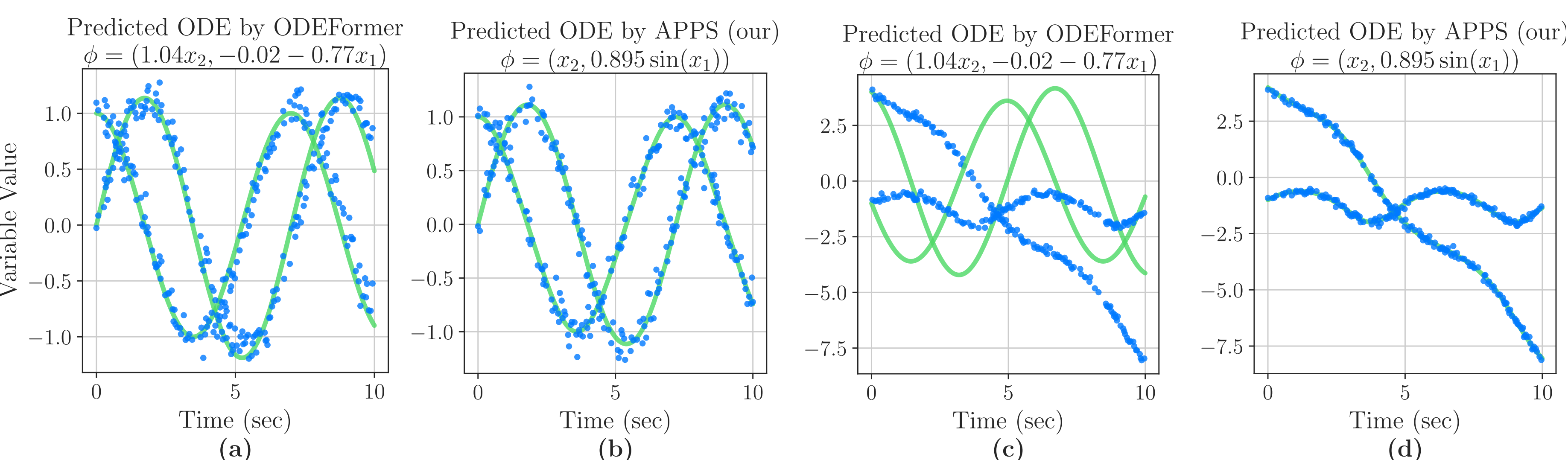


Figure 1. The quality of predicted ODE from passively-learned baseline is heavily influenced by the collected training data while our APPS is not.

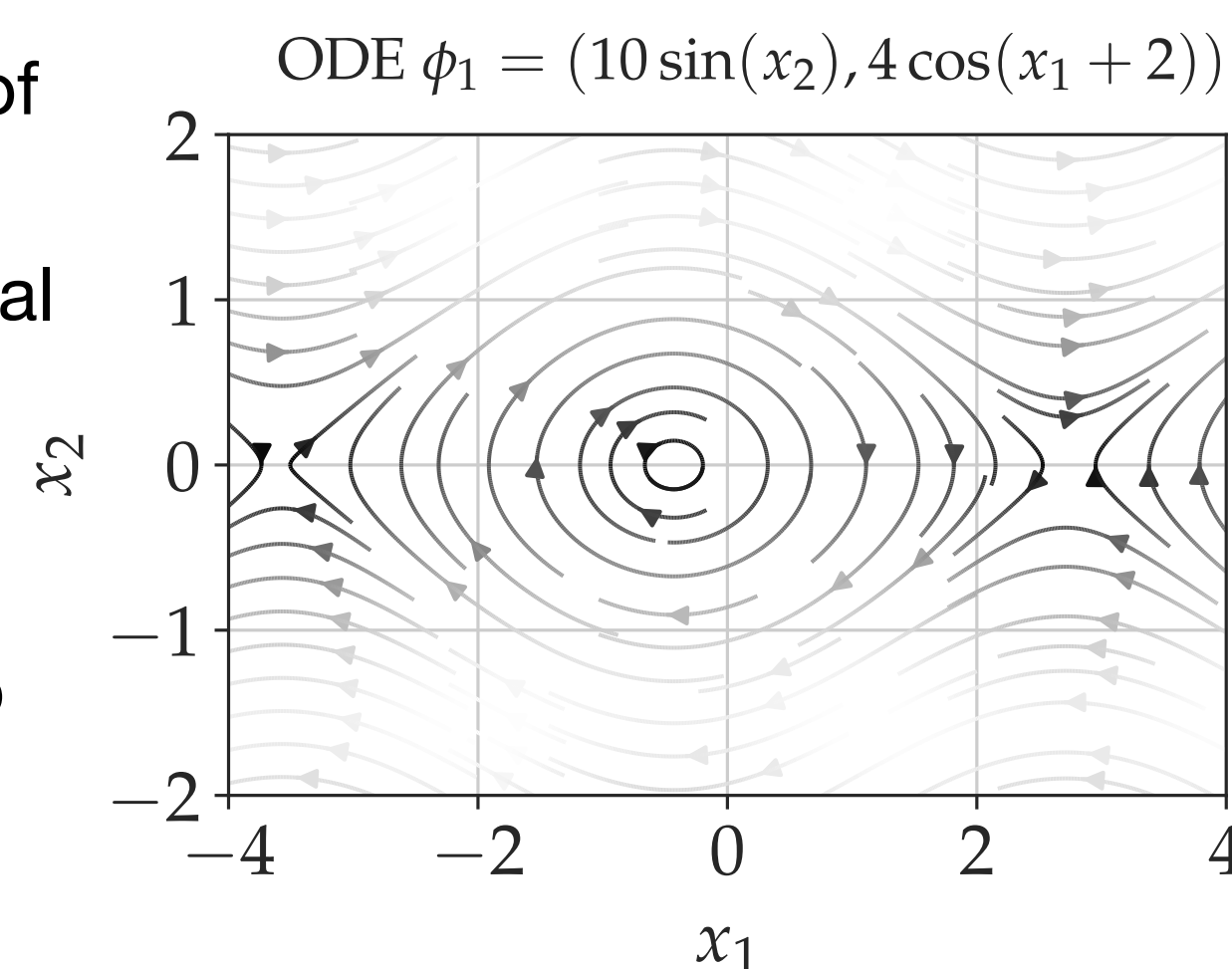
— Predicted x_1 — Predicted x_2
• True x_1 • True x_2

Motivation: How to actively query informative data for evaluating ODEs?

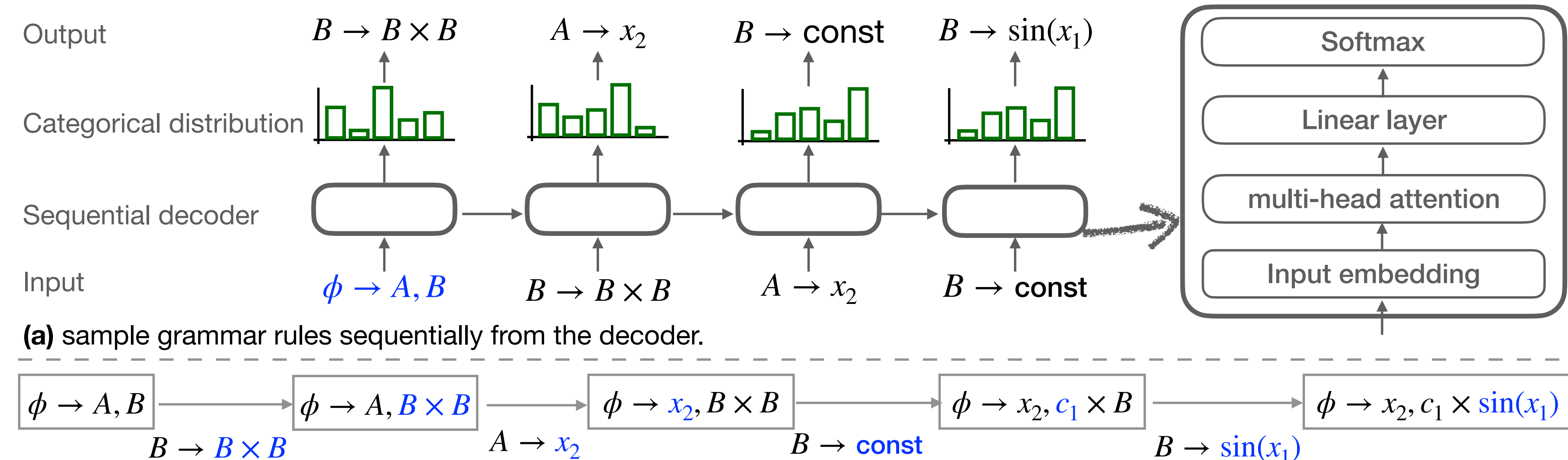
Phase Portrait is a qualitative analysis tool for studying the behavior of dynamical systems.

- A curve is a short trajectory of the system over time from a given initial condition.
- The arrow on the curve indicates the direction of change.

Butterfly Effect states small variations in initial conditions can lead to vastly different outcomes.

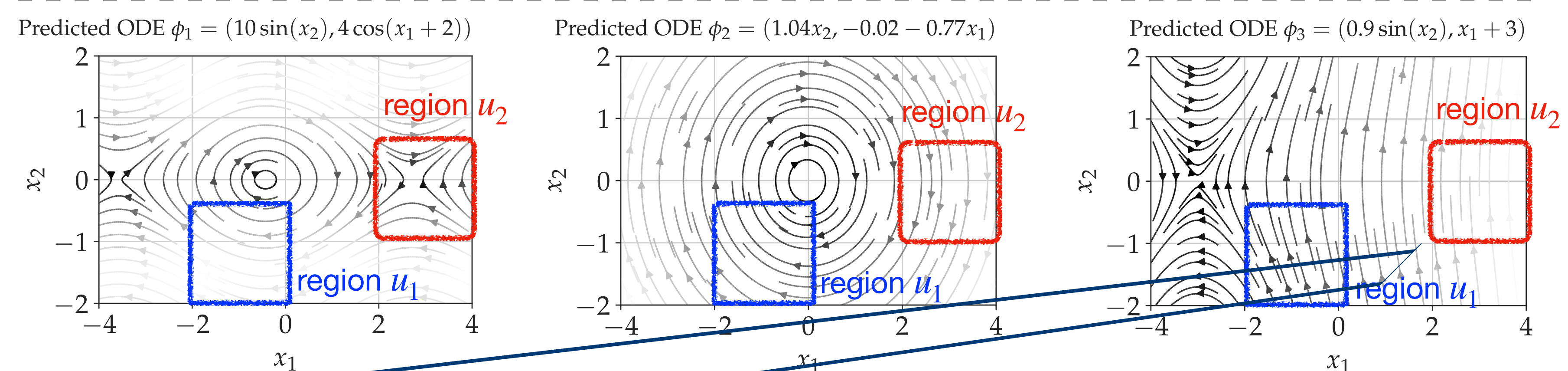


Methodology: Active Symbolic Discovery of ODEs via Phase Portrait Sketching



(a) sample grammar rules sequentially from the decoder.

(b) convert a sequence of rules into an ODE by context-free grammar: $\phi = (x_2, c_1 \sin(x_1))$.



(c) region u_2 is selected to draw data other than region u_1 for its higher informativeness.

Experimental Analysis

	Strogatz dataset ($\sigma^2 = 0, \alpha = 0$)				ODEbase dataset ($\sigma^2 = 0, \alpha = 0$)			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 2$	$n = 3$	$n = 4$	$n = 5$
SPL	0.787	0.892	1.921	2.865	0.867	2.17	4.75	13.16
E2ETransformer	$6.47E-4$	1.620	<i>T.O.</i>	<i>T.O.</i>	0.757	<i>T.O.</i>	<i>T.O.</i>	<i>T.O.</i>
ProGED	0.129	0.666	2.68	3.856	0.317	2.134	<i>T.O.</i>	<i>T.O.</i>
SINDy	$1.90E-4$	0.217	1.539	4.810	0.521	2.112	8.334	52.12
ODEFormer	0.0303	0.9261	1.033	1.010	0.213	0.245	1.213	3.148
APPS (ours)	2.06E-6	0.2912	1.011	0.521	0.1318	0.1306	1.046	3.054

Table 1: On the *noiseless* datasets with regular time sequence ($\sigma^2 = 0, \alpha = 0$), Median NMSE is reported over the best-predicted expression found by all the algorithms.

	Noisy Strogatz datasets ($\sigma^2 = 0.01, \alpha = 0$)				Irregular Strogatz dataset ($\sigma^2 = 0, \alpha = 0.1$)			
	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$n = 1$	$n = 2$	$n = 3$	$n = 4$
SPL	0.938	1.019	2.915	3.068	0.127	0.526	3.196	4.193
SINDy	$6.4E-3$	4.152	2.498	5.21	$6.66E-4$	0.472	0.827	4.163
ProGED	0.121	0.658	3.673	3.856	0.134	0.769	2.766	4.181
ODEFormer	0.139	0.621	2.392	0.812	0.031	1.036	1.51	1.011
APPS (ours)	7.75E-4	0.369	1.381	0.657	1.06E-6	0.215	1.012	0.947

Table 2: On the Strogatz dataset, the Median NMSE is reported over the best-predicted expression found by all the algorithms under noisy or irregular time sequence settings.

Table 3: Our method shows a smaller ranking-based distance. Also APPS takes less memory consumption and less computational time.

	Ranking-based distance ($\downarrow$)	Running Time ($\downarrow$)	Peak Memory ($\downarrow$)
APPS (ours)	0.08	5.2 sec	3.76 MB
QbC	0.13	13.4 sec	51 MB
CoreSet	0.22	4.3 sec	2.74 GB