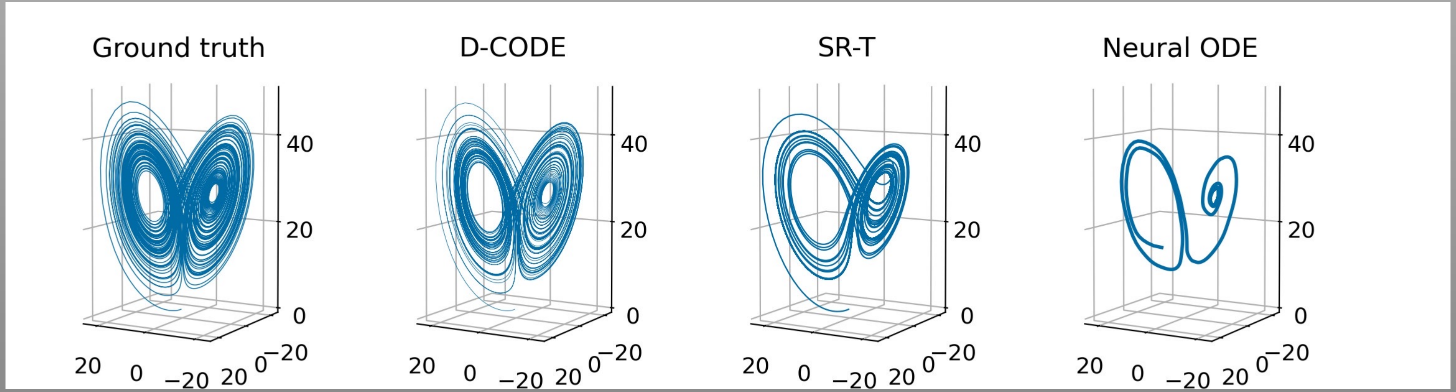
D-CODE: DISCOVERING CLOSED-FORM ODES FROM OBSERVED TRAJECTORIES



Paper Authors: Zhaozhi Qian, Krzysztof Kacprzyk , Mihaela van der Schaar

Image reproduced under fair use from https://openreview.net/pdf?id=wENMvlsxNN

Introduction

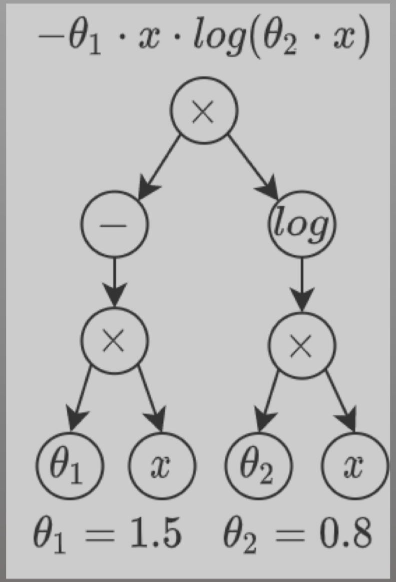
- For centuries, experts designed closed-form ODEs for natural systems
- Can this be automated?
- Symbolic regression produces a closed-form prediction $a = f(b)$
  - label-feature pairs (a_i, b_i) as training examples
- For ODE models, the time derivative or the “label” $\dot{x}(t)$ is *not* observed
- Existing approaches:
  - low measurement noise
  - frequent sampling.
- Discovery of Closed-form ODE framework (D-CODE)
  - advances symbolic regression
  - new objective function using variational formulation of ODEs
    - avoids the unobserved time derivative.

Related Work: Symbolic Regression

Method	Data	Allowed f^*	Est.	$\dot{x}$ Free	$\boldsymbol{x}(0)$ Free	Objective
Symbolic Reg [1]	$a, \boldsymbol{b}$	Closed-form	None	-	-	$\lvert\lvert a - f(\boldsymbol{b})\rvert\rvert_2$

Image reproduced under fair use from https://openreview.net/pdf?id=wENMvIsxNN

- A closed-form prediction $a = f(\boldsymbol{b})$ using supervised learning.
- Key challenge is *optimization*
  - searching for optimal *f* is thought to be NP-hard
    - combinatorial in the functional form
    - continuous in the constants
- Most existing work focuses on optimization.
  - Genetic programming
    - represents *f* as a *tree,* a set of heuristics to directly prune search space
  - reinforcement learning
  - pre-trained neural networks
  - Meijer G-functions
- All of the above use prediction RMSE as a loss function
  - But, the label (time derivative) is not observed for ODEs.



Related Work: Data-driven Approaches I

Method	Data	Allowed f^*	Est.	$\dot{x}$ Free	$\boldsymbol{x}(0)$ Free	Objective
Symbolic Reg [1]	$a, \boldsymbol{b}$	Closed-form	None	-	-	$\|\|a - f(\boldsymbol{b})\|\|_2$
2-step Sparse [2]	$\boldsymbol{y}(t)$	$\sum \theta_k h_k(\boldsymbol{x})$	$\widehat{\dot{x}}$	$\times$	$\checkmark$	$\sum_t \|\|\widehat{\dot{x}}(t) - f(\boldsymbol{y}(t))\|\|_2$

- **2-step sparse regression** on the estimated time derivative $\dot{x}_j$
- Assumes linear form of ODE: $\dot{x}_j(t) = \sum_{k=1}^{K} \theta_k h_k(\boldsymbol{x}(t))$
  - $\theta_k \in R$ are unknown constants
  - $h_k : R^D \rightarrow R$ are pre-specified functions
  - Uses L1 regularization to produce fewer terms
- Problems:
  - linear form is too restrictive
  - choice functions require expert inputs

Related Work: Data-driven Approaches II

Method	Data	Allowed f^*	Est.	$\dot{x}$ Free	$\boldsymbol{x}(0)$ Free	Objective
Symbolic Reg [1]	$a, \boldsymbol{b}$	Closed-form	None	-	-	$\|\|a - f(\boldsymbol{b})\|\|_2$
2-step Sparse [2]	$\boldsymbol{y}(t)$	$\sum \theta_k h_k(\boldsymbol{x})$	$\widehat{\dot{x}}$	$\times$	✓	$\sum_t \|\|\widehat{\dot{x}}(t) - f(\boldsymbol{y}(t))\|\|_2$
2-step Symbolic [3]	$\boldsymbol{y}(t)$	Closed-form	$\widehat{\dot{x}}$	$\times$	✓	$\sum_t \|\|\widehat{\dot{x}}(t) - f(\boldsymbol{y}(t))\|\|_2$

- 2-step symbolic regression
- Uses *estimated* time derivatives $\dot{x}_j$ as the label
- Employs *any* optimization to search for the optimal function
- Removes linear assumption on the functional form
- Challenges:
  - true label (derivative) may not be accurate
    - Measurement
    - discretization error
  - symbolic regression sensitive to inaccurate labels as the search space is huge

Related Work: Data-driven Approaches III

Method	Data	Allowed f^*	Est.	$\dot{x}$ Free	$\boldsymbol{x}(0)$ Free	Objective
Symbolic Reg [1]	$a, \boldsymbol{b}$	Closed-form	None	-	-	$\|\|a - f(\boldsymbol{b})\|\|_2$
2-step Sparse [2]	$\boldsymbol{y}(t)$	$\sum \theta_k h_k(\boldsymbol{x})$	$\widehat{\dot{x}}$	$\times$	$\checkmark$	$\sum_t \|\|\widehat{\dot{x}}(t) - f(\boldsymbol{y}(t))\|\|_2$
2-step Symbolic [3]	$\boldsymbol{y}(t)$	Closed-form	$\widehat{\dot{x}}$	$\times$	$\checkmark$	$\sum_t \|\|\widehat{\dot{x}}(t) - f(\boldsymbol{y}(t))\|\|_2$
ODE Approx [4]	$\boldsymbol{y}(t)$	Neural nets	$\widehat{\boldsymbol{x}}(0)$	$\checkmark$	$\times$	$\sum_t \|\|\boldsymbol{y}(t) - \widehat{\boldsymbol{x}}(t)\|\|_2$

- **Function approximator approach**
  - learns the true ODE with a function approximator g
    - a neural network (NN)
    - a Gaussian process (GP)
  - estimates the unknown initial condition x(0)
  - predicts the entire trajectory x by solving the approximated ODE g
  - trained by minimizing the error between predicted and the observed trajectories
- Challenges:
  - No closed-form expression to describe the dynamics
  - Performance depends on prediction horizon used for training.

D-CODE Algorithm- I

- Pre-processing
  - Estimate $\boldsymbol{x}_i : [0, T] \to R^J$, $i \leq N$ from noisy and discretely-sampled data D
  - Denoising and interpolating well-studied in statistics and signal processing
    - Gaussian process
    - spline regression
  - D-CODE agnostic to exact choice of smoothing algorithm

D-CODE Algorithm- II

- Pre-processing
  - Estimate $\boldsymbol{x}_i : [0, T] \to R^J$, $i \le N$ from noisy and discretely-sampled data D
- Consider $J \in N^+$, $T \in R^+$, continuous functions $x : [0,T] \to R^J$, $f : R^J \to R$, and $g \in C^1[0,T]$.

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big) g(t) dt + \int_0^T x_j(t) \dot{g}(t) dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

- Search for the function f_j, $\forall j \le J$ that is consistent with $\boldsymbol{x}_i$. Formally,

$$\widehat{f}_j = \arg\min_f \sum_{i=1}^{N} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_i, g_s)^2$$

- Details:
  - Search over f is symbolic,
  - C_j is obtained numerically,
  - g_s has multiple choices, e.g., $g_s(t) = \sqrt{2/T} \cdot \sin(s\pi t/T)$.

Theoretical Results - I

Lemma (Fundamental lemma of calculus of variations)**.** *Let* h *be a continuous function on a closed interval* $[0, T]$*.* h *is equal to 0 everywhere if and only if* $\int_0^T h(t)g(t)\,dt = 0$ *for all* $g \in \mathcal{C}^1[0, T]$ *such that* $g(0) = g(T) = 0$*.*

Example: $g_s(t) = \sqrt{2/T} \cdot \sin(s\pi t/T)$

Theoretical Results- II

ODE Problem

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)),\ \forall j = 1, \ldots, J,\ \forall t \in [0, T] \tag{1}$$

Proposition 1. *(Hackbusch, 2017) Consider* $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, *a continuously differentiable function* $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, *and continuous functions* $f_j : \mathbb{R}^J \to \mathbb{R}$ *for* $j = 1, \ldots, J$. *Then* $\boldsymbol{x}$ *is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

Recall the definition of C_j:

$$C_j(f, \boldsymbol{x}, g) := \int_0^T f\big(\boldsymbol{x}(t)\big)g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt; \quad \forall j \in \{1, 2, \ldots, J\}$$

Theoretical Results - III

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\dot{x}_j(t) = f_j(\boldsymbol{x}(t)) \iff f_j(\boldsymbol{x}(t)) - \dot{x}_j(t) = 0$$

$$f_j(\boldsymbol{x}(t)) - \dot{x}_j(t) = 0\ \forall t \in [0, T]$$

$$\Updownarrow$$

Fundamental Lemma of Calculus of Variations

$$\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t)) g(t) dt = 0\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Theoretical Results - IV

Proposition 1. *(Hackbusch, 2017) Consider* $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, *a continuously differentiable function* $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, *and continuous functions* $f_j : \mathbb{R}^J \to \mathbb{R}$ *for* $j = 1, \ldots, J$. *Then* $\boldsymbol{x}$ *is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t))g(t)dt = \int_0^T f_j(\boldsymbol{x}(t))g(t)dt - \int_0^T \dot{x}_j(t)g(t)dt$$

Linearity

Theoretical Results - V

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \ldots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\begin{aligned}\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t))g(t)dt &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt - \int_0^T \dot{x}_j(t)g(t)dt \\ &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt - x_j(T)g(T) + x_j(0)g(0)\end{aligned}$$

Linearity

Integration by parts

Theoretical Results - VI

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \dots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \dots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\begin{aligned}\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t))g(t)dt &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt - \int_0^T \dot{x}_j(t)g(t)dt \\ &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt - x_j(T)g(T) + x_j(0)g(0) \\ &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt\end{aligned}$$

Linearity

Integration by parts

Since, g(0)=g(T)=0

Theoretical Results - VII

Proposition 1. *(Hackbusch, 2017) Consider* $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, *a continuously differentiable function* $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$, *and continuous functions* $f_j : \mathbb{R}^J \to \mathbb{R}$ *for* $j = 1, \ldots, J$. *Then* $\boldsymbol{x}$ *is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \ldots, J\},\ \forall g \in \mathcal{C}^1[0, T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t))g(t)dt = \int_0^T f_j(\boldsymbol{x}(t))g(t)dt - \int_0^T \dot{x}_j(t)g(t)dt$$

Linearity

$$= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt - x_j(T)g(T) + x_j(0)g(0)$$

Integration by parts

$$= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt$$

Since, g(0)=g(T)=0

$$= C_j(f_j, \boldsymbol{x}, g)$$

By definition of C_j

Theoretical Results - VIII

Proposition 1. *(Hackbusch, 2017) Consider $J \in \mathbb{N}^+$, $T \in \mathbb{R}^+$, a continuously differentiable function $\boldsymbol{x} : [0,T] \to \mathbb{R}^J$, and continuous functions $f_j : \mathbb{R}^J \to \mathbb{R}$ for $j = 1, \dots, J$. Then $\boldsymbol{x}$ is the solution to the system of ODEs in Equation 1 if and only if*

$$C_j(f_j, \boldsymbol{x}, g) = 0,\ \forall j \in \{1, \dots, J\},\ \forall g \in \mathcal{C}^1[0,T],\ g(0) = g(T) = 0$$

Proof Sketch

$$\begin{aligned}
\int_0^T (f_j(\boldsymbol{x}(t)) - \dot{x}_j(t))g(t)dt &= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt - \int_0^T \dot{x}_j(t)g(t)dt \\
&= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt - x_j(T)g(T) + x_j(0)g(0) \\
&= \int_0^T f_j(\boldsymbol{x}(t))g(t)dt + \int_0^T x_j(t)\dot{g}(t)dt \\
&= C_j(f_j, \boldsymbol{x}, g)
\end{aligned}$$

Linearity

Integration by parts

Since, g(0)=g(T)=0

C_j = 0 by Proposition 1

$$\boxed{\dot{x}_j(t) = f_j(\boldsymbol{x}(t))}$$

Theoretical Results - IX

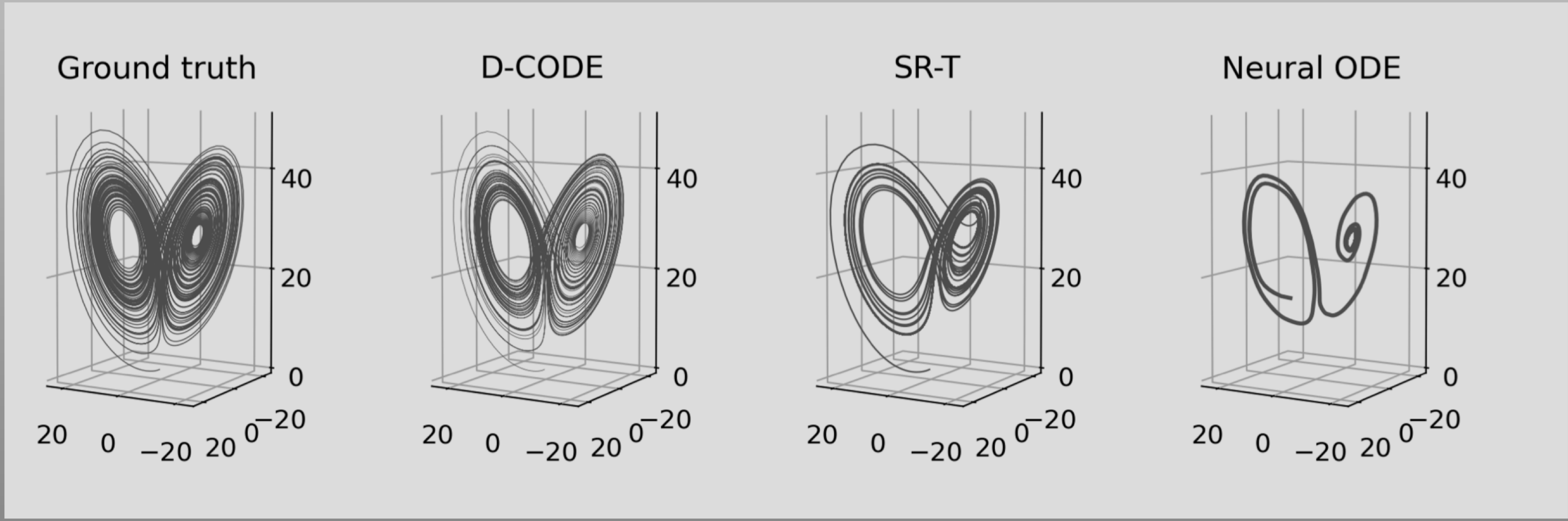
Theorem 1. *Consider* $J \in \mathbb{N}^+$, $j \in \{1, \ldots, J\}$, $T \in \mathbb{R}^+$. *Let* $f^* : \mathbb{R}^J \to \mathbb{R}$ *be a continuous function, and let* $\boldsymbol{x} : [0, T] \to \mathbb{R}^J$ *be a continuously differentiable function satisfying* $\dot{x}_j(t) = f^*(\boldsymbol{x}(t))$. *Consider a sequence of functions* $(\widehat{\boldsymbol{x}}_k)$, *where* $\widehat{\boldsymbol{x}}_k : [0, T] \to \mathbb{R}^J$ *is a continuously differentiable function. If* $(\widehat{\boldsymbol{x}}_k)$ *converges to* $\boldsymbol{x}$ *in* L^2 *norm. Then for any Lipschitz continuous function* f

$$\lim_{S\to\infty} \lim_{k\to\infty} \sum_{s=1}^{S} C_j(f, \widehat{\boldsymbol{x}}_k, g_s)^2 = d_{\boldsymbol{x}}(f, f^*)^2, \tag{7}$$

where $\{g_1, g_2, \ldots\}$ *is a Hilbert (orthonormal) basis for* $L^2[0, T]$ *such that* $\forall i$, $g_i(0) = g_i(T) = 0$ *and* $g_i \in \mathcal{C}^1[0, T]$.

Distance $d_{\boldsymbol{x}}(f, f^*) := ||f \circ \boldsymbol{x} - f^* \circ \boldsymbol{x}||_2 = ||(f - f^*) \circ \boldsymbol{x}||_2$

Results- I



$$\frac{\mathrm{d}x}{\mathrm{d}t} = \sigma(y - x),$$
$$\frac{\mathrm{d}y}{\mathrm{d}t} = x(\rho - z) - y,$$
$$\frac{\mathrm{d}z}{\mathrm{d}t} = xy - \beta z.$$

Image reproduced under fair use from https://openreview.net/pdf?id=wENMvIsxNN

Results- II

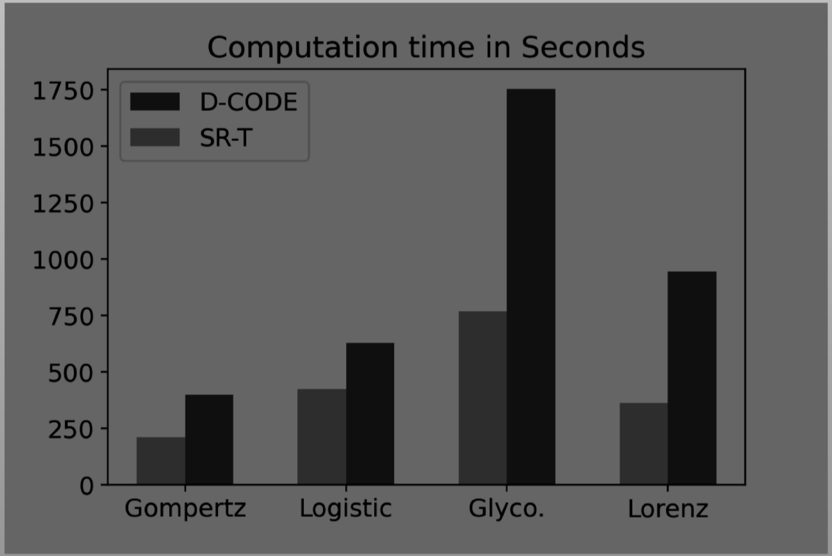


Image reproduced under fair use from https://openreview.net/pdf?id=wENMvIsxNN

Equation	Method	Success Prob. $\sigma_R = 0.01$	0.1	0.2	RMSE $\hat{\theta}$ (10^{-2}) $\sigma_R = 0.01$	0.1	0.2
$\dot{x}_1(t) = \theta_1 - \theta_2 x_1(t) - x_1(t)x_2(t)^2$	SR-T	0.45 (.05)	0.27 (.05)	0.00 (.00)	1.19 (.15)	2.37 (.38)	NA
	SR-S	0.44 (.06)	0.11 (.04)	0.00 (.00)	1.67 (.24)	2.07 (.47)	NA
	SR-G	0.44 (.05)	0.05 (.02)	0.00 (.00)	1.87 (.22)	2.18 (.46)	NA
	D-CODE	**0.58 (.05)**	**0.51 (.05)**	**0.26 (.05)**	**1.01 (.12)**	**1.55 (.28)**	**2.00 (.27)**
$\dot{x}_2(t) = -x_2(t) + \theta_3 x_1(t) + x_1(t)x_2(t)^2$	SR-T	0.99 (.03)	0.88 (.03)	0.00 (.00)	0.14 (.03)	0.43 (.03)	NA
	SR-S	0.95 (.02)	0.04 (.02)	0.00 (.00)	0.10 (.01)	0.41 (.18)	NA
	SR-G	0.99 (.01)	**0.94 (.02)**	0.25 (.04)	0.22 (.02)	1.10 (.07)	2.85 (.02)
	D-CODE	**1.00 (.00)**	**0.91 (.03)**	**0.65 (.05)**	**0.07 (.01)**	**0.40 (.07)**	**0.61 (.05)**

Future Work

- Unobserved variables
  - The ODE may have variables that are not directly observable.
  - Examples?
- Complex equations
  - High-dimensionality of variables
  - Multiple mathematical operations
  - May be challenging even for experts with domain knowledge.
- Extreme measurement settings
  - Too much noise
  - Too large step size