AAAI 2023 Spring Symposium Series 29/03/2023

Data-Driven and Knowledge-Based Causal Network Discovery for Identifying Differential Equations



Mitsuhiro Odaka[1,2,3,4,*] Morgan Magnin[3,2] Katsumi Inoue[2,1,3]

[1] The Graduate University for Advanced Studies, SOKENDAI, Japan
[2] National Institute of Informatics, Japan
[3] École Centrale de Nantes, France
[4] Japan Society for the Promotion of Science (JSPS) Research Fellowships for Young Scientists

*odaka@nii.ac.jp

Table of contents

Table of contents

Motivation: Uncovering dynamics

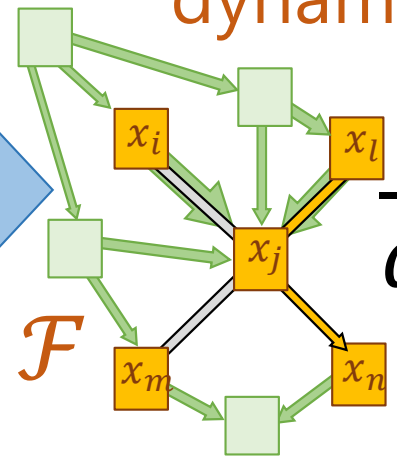
= Causal network discovery + Identifying governing equations

Equations governing the dynamics of a system

State variables

$X = (x_1, x_2, \ldots, x_n)$

Uncovering dynamics



$\frac{d}{dt}X = \mathcal{F}(X)$

Identifying networks and $\mathcal{F}$

↓

More profound understanding of the systems or the world

- ❑ Advocacy of the system control strategy
- ❑ Extrapolation by numerical simulation
- ❑ Comparison of different interventions

Data-Driven and Knowledge-Based (DD-KB) Integrated Approach

Learning and inference of causal networks and differential equations governing the dynamics of a system from **observed data** & **background knowledge**

Problem Settings

Inputs

- ❑ Observations $\boldsymbol{O}$

  Continuous multivariate time series data $X(t) \in \mathbb{R}^d$

  X: Observed variables t: Time index (arbitrary unit) $\in [0, \mathbb{R}^+\}$

- ❑ Background knowledge $\boldsymbol{B}$

Outputs

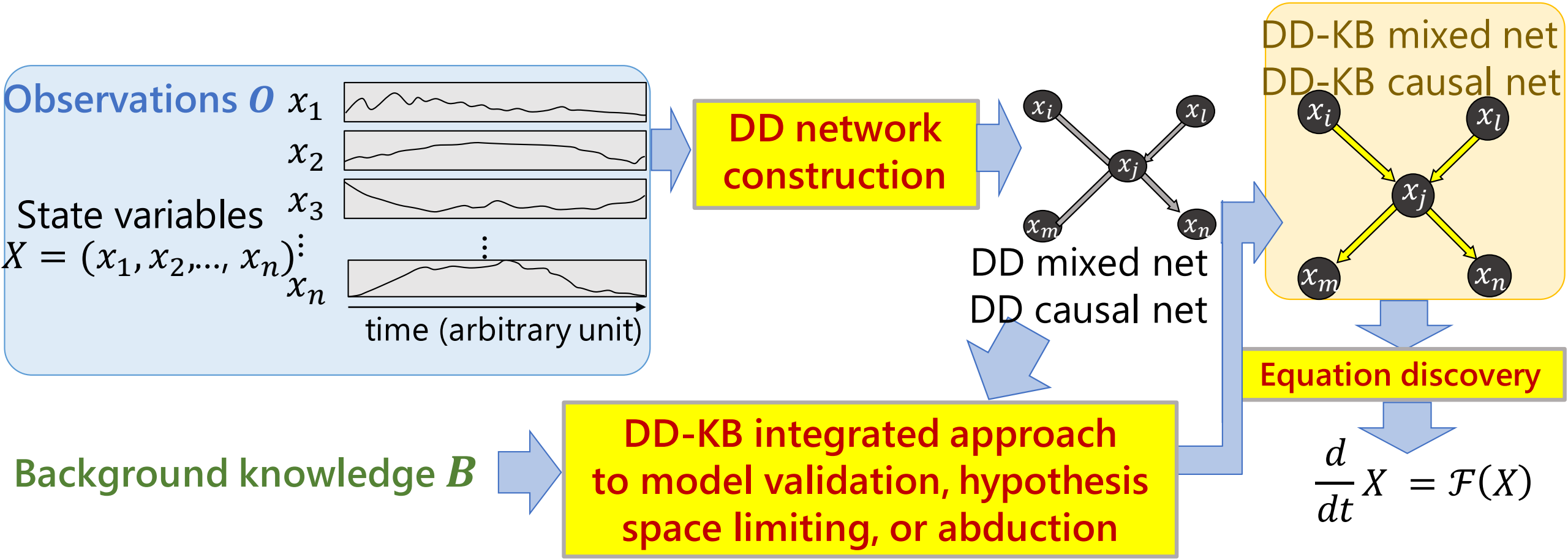
- ❑ Continuous deterministic dynamical system $\langle X, \dot{X}, \mathcal{F}, t \rangle$

  $\mathcal{F}$: Dynamic constraints that define equations governing a dynamics of the system

- ❑ DD-KB network $\mathcal{M}$

  (Directed/undirected mixed network, Causal network, etc.)

Data-Driven and Knowledge-Based (DD-KB) Integrated Approach



Aim 1: To apply DD-KB integrated approach to gene network inference for buidling new COVID-19 pathways

Aim 2: To develop method for learning causal network in continuous domain

Table of contents

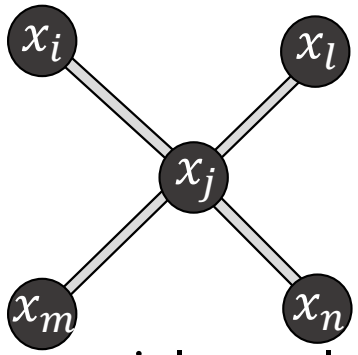
Data-driven & knowledge-based (DD-KB) integrated approach to gene network inference for constructing COVID-19 pathways

[Odaka et al. Preprint at *Research Square.* 2022]
[Odaka et al. *CAMDA.* 2022]

Observations
$\boldsymbol{O}$

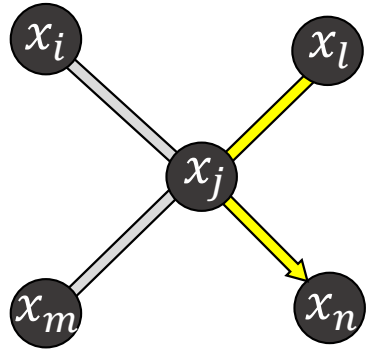
Gene expression profiles of bronchial lavage fluid samples from COVID-19 severe patients (n=10) & negative control (n=2) [Grant et al. *Nature.* 2021]
(cells 15,481 × genes 21,819)

Gene network inference



2nd-order partial correlation network (undirected)

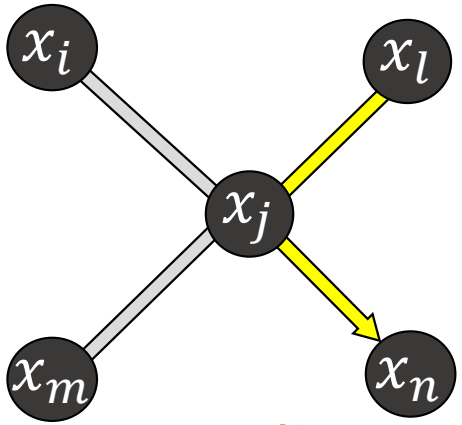
Unification & Model validation



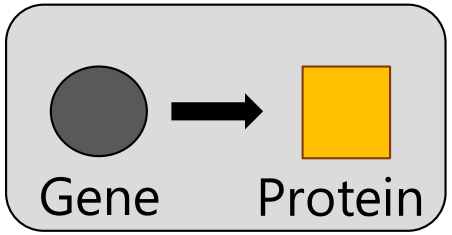
DD-KB directed / undirected mixed net

Background knowledge $\boldsymbol{B}$

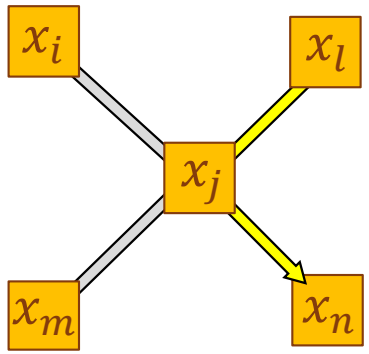
Pathway Commons ver 12 (Reactome, Panther, HumanCyc, BIND, MSigDB), BioGRID ver 4.4, STRING ver 11.5

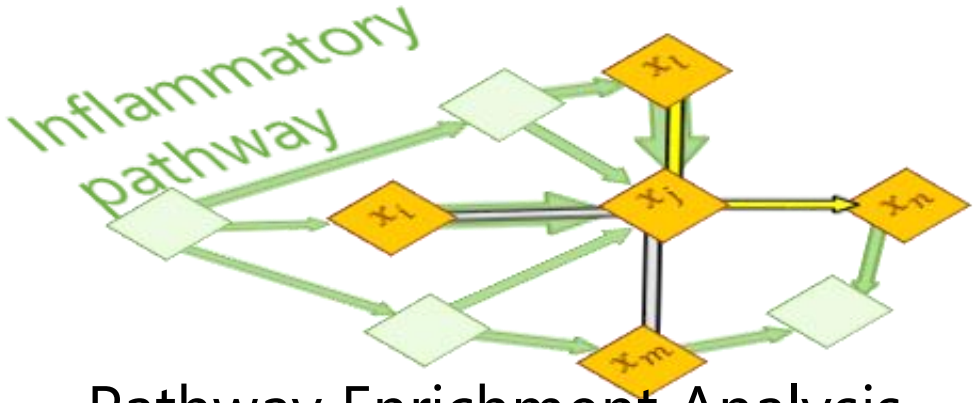


DD-KB directed / undirected mixed net



Gene-Protein Conversion



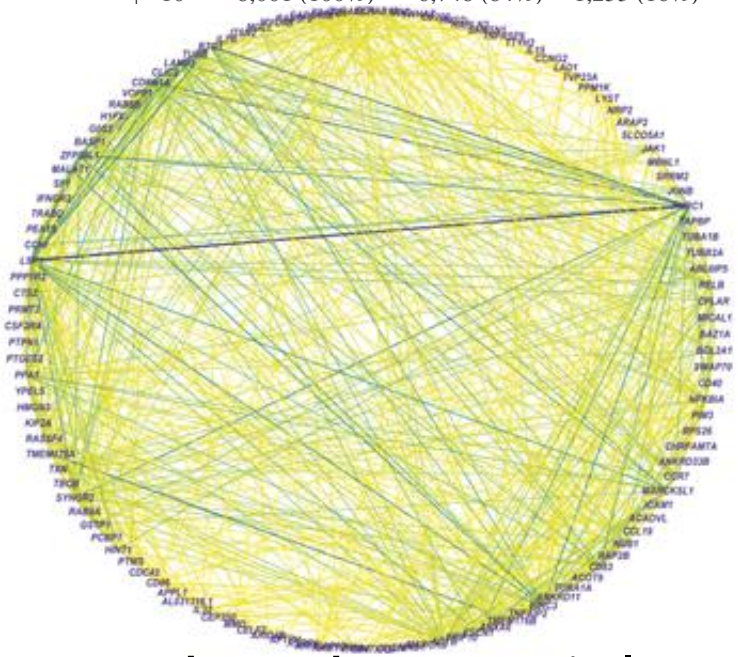


Pathway Enrichment Analysis
Mapping onto function-known KEGG pathways
e.g., Activation of genes of inflammatory pathway w/ significant difference

DD-KB integrated approach to gene network inference for constructing COVID-19 pathways

Spurious correlation removal

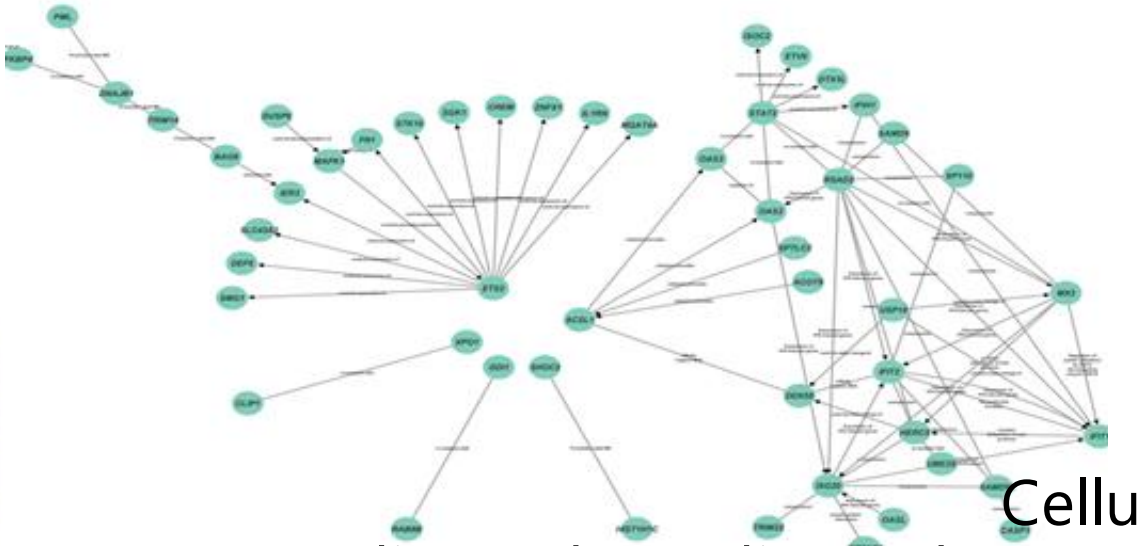
	Day	Full	Removed	Final
ACTB	1	5,671 (100%)	4,757 (84%)	914 (16%)
	5	6,328 (100%)	5,301 (84%)	1,027 (16%)
	10	5,050 (100%)	4,257 (84%)	793 (16%)
ICAM1	1	7,503 (100%)	6,309 (84%)	1,194 (16%)
	5	20,706 (100%)	18,914 (91%)	1,792 (9%)
	10	8,001 (100%)	6,748 (84%)	1,253 (16%)
C15orf48	1	9,453 (100%)	7,995 (85%)	1,458 (15%)
	5	13,530 (100%)	12,049 (89%)	1,481 (11%)
	10	8,001 (100%)	6,748 (84%)	1,253 (16%)



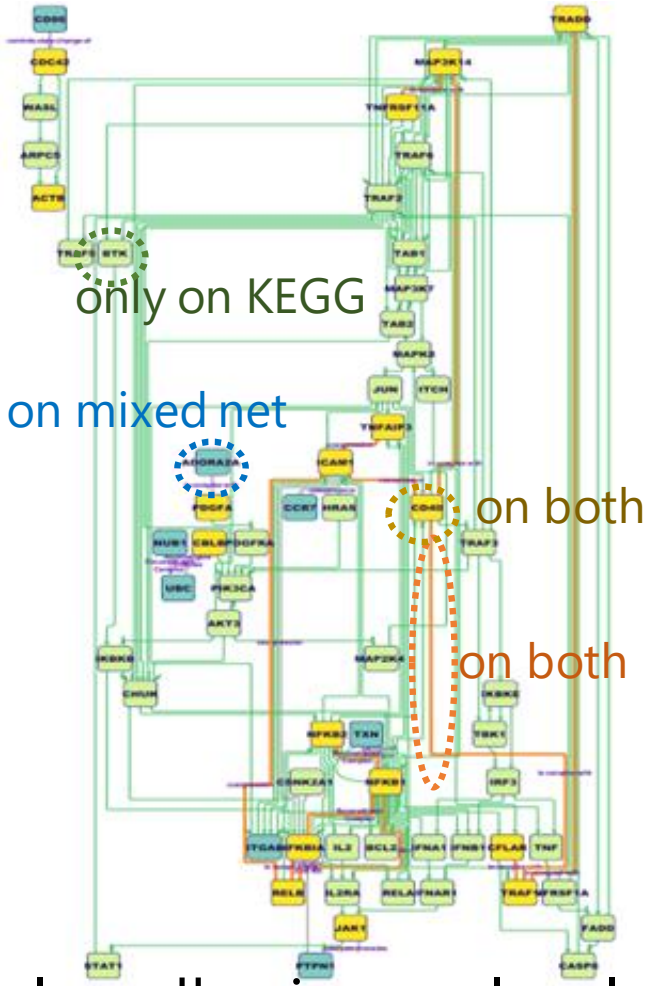
2nd-order partial correlation network

Model validation by knowledge

Source	Target	Weight	Relation	Knowledge bases
REPIN1	SNN	0.366753074	controls-expression-of	MSigDB
CCL22	CCR6	0.189063535	in-complex-with	PathwayCommons_Get_API
DUSP5	PTPN1	0.428685	neighbor-of	Panther
JAK1	PTPN1	0.564563581	catalysis-precedes	Panther
NFKBIA	PTPN1	0.580935037	Proximity Label-MS	BioGRID
LAMP3	RAB9A	0.487968997	Proximity Label-MS	BioGRID



DD-KB directed / undirected mixed net



Cellular adhesion molecule *ICAM1* pathways inferred from COVID-19 omics data

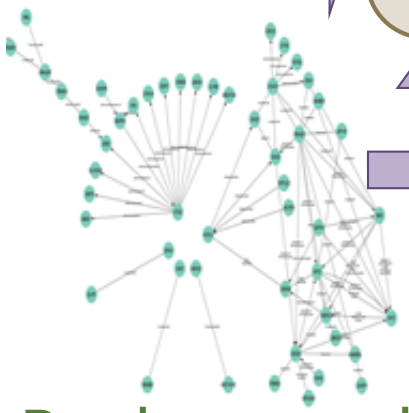
(1) Pathways missing from the current C19DMap
(2) Molecules supporting SARS-CoV-2 cell-to-cell transimission hypothesis [Odaka & Inoue *Heliyon* 2021]

Table of contents

Learning differential equations via causal network discovery from multivariate time series

Observations $\boldsymbol{O}$

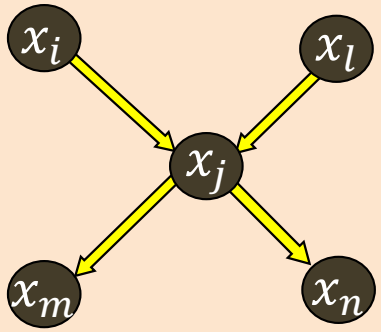
Causal network discovery
from multivariate time series by adversarial learning
[Odaka et al. *JSAI*. 2023]



Input as candidate skeleton

Background knowledge $\boldsymbol{B}$

- ❑ Networks purely from knowledge bases
- ❑ DD-KB directed / undirected mixed net



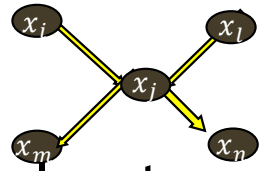
- The model is represented by **structural causal model**.
- **Granger causality** [Granger *Econometrica* 1969] defines edge. (Causal effects between time-series sequences)

DD-KB causal net (DAG)

Imputing as inductive bias to limit hypothesis space of equation discovery

Equation discovery
e.g., Sparse Identification of Nonlinear Dynamics (SINDy)
[Brunton et al. *PNAS*. 2016]

$$\frac{d}{dt}X = \mathcal{F}(X)$$



ODEs, Causal network

Model verification
- ❑ Sensitivity analysis, Stability analysis
- ❑ Reachability analysis, model checking
- ❑ Assessment w/ model selection criteria (AIC, BIC, MDL, etc.)

Verified ODEs, Causal network

General pros/cons in deep learning and analytical technique

	Deep learning	Analytical technique
Robustness to noise	●	▲
Scalability	●	▲
Interpretability	▲	●

To cover the pros and cons of each other, we combine **deep learning** and **analytical technique** in equation discovery.

Problem settings

Given: Observed data

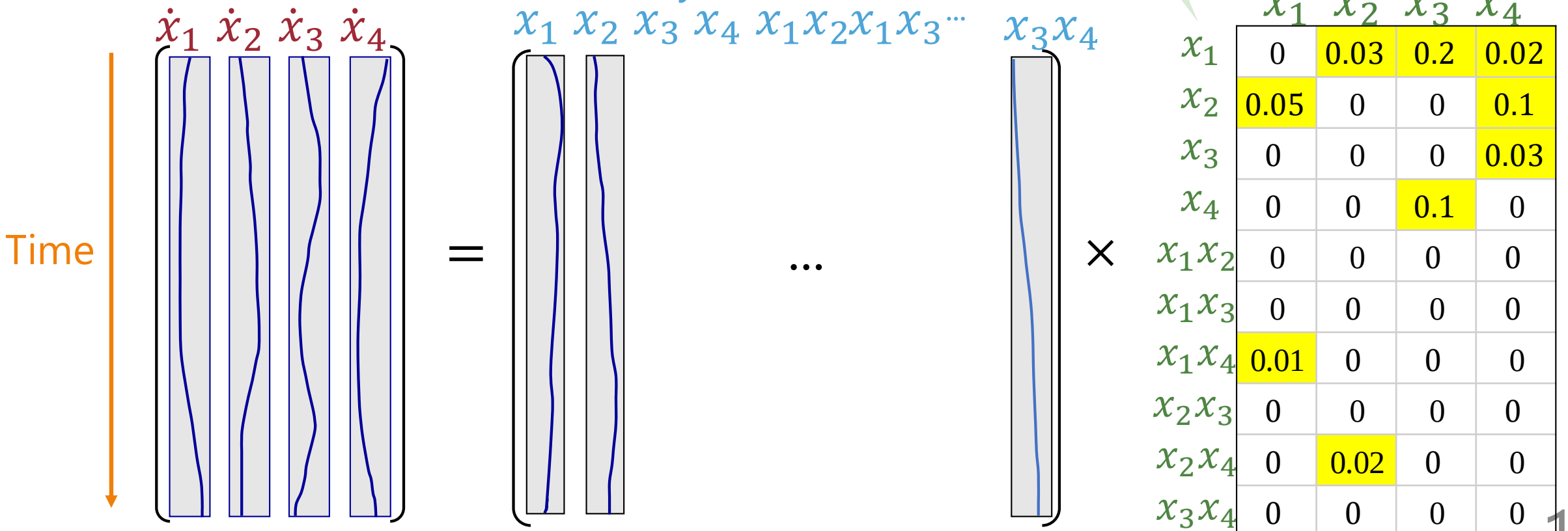
$$\frac{d}{dt}\boxed{X} = \Theta(\boxed{X})\,\boxed{\Xi}$$

Selecting variables amenable for the dynamics representation

Time series of endogenous vars' derivatives

Time series of endogenous vars & synthetic vars

Find: Coefficient vector



	x_1	x_2	x_3	x_4
x_1	0	0.03	0.2	0.02
x_2	0.05	0	0	0.1
x_3	0	0	0	0.03
x_4	0	0	0.1	0
x_1x_2	0	0	0	0
x_1x_3	0	0	0	0
x_1x_4	0.01	0	0	0
x_2x_3	0	0	0	0
x_2x_4	0	0.02	0	0
x_3x_4	0	0	0	0

Existing equation discovery techniques have not often explicitly discovered causality. → Overfitting by dull terms

Problem settings

Given: Observed data

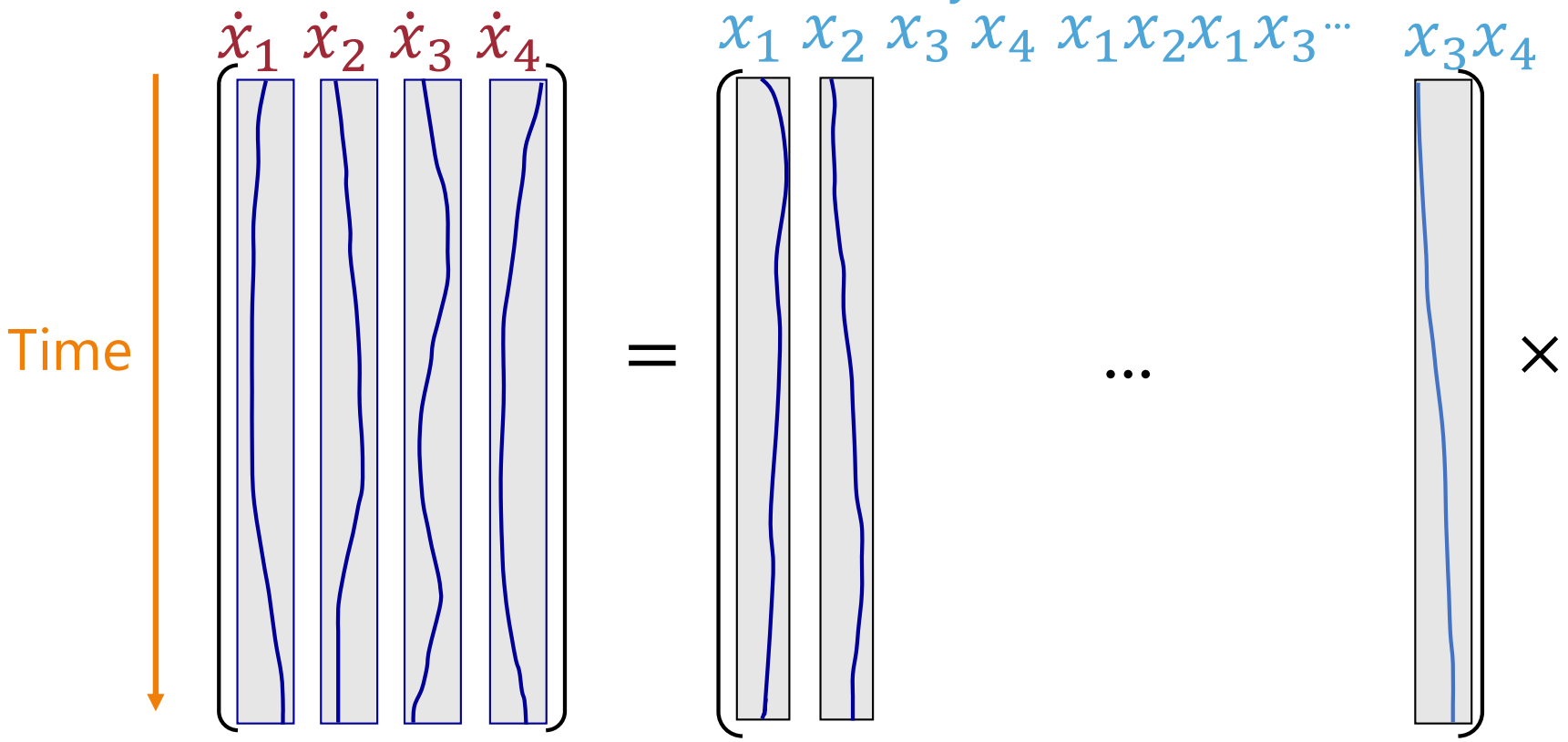
$$\frac{d}{dt}X = \Theta(X)\,\Xi$$

Selecting variables amenable for the dynamics representation

Time series of endogenous vars' dervatives

Time series of endogenous vars & synthetic vars

Find: Coefficient vector



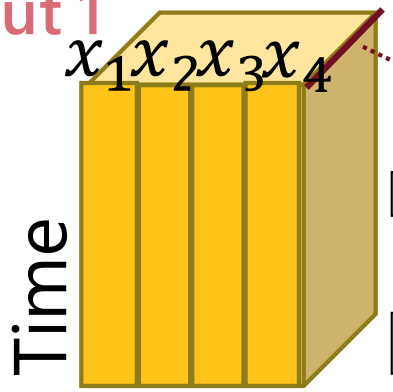
	x_1	x_2	x_3	x_4
x_1	0	0.03	0.2	0.02
x_2	0.05	0	0	0.1
x_3	0	0	0	0.03
x_4	0	0	0.1	0
x_1x_2	0	0	0	0
x_1x_3	0	0	0	0
x_1x_4	0.01	0	0	0
x_2x_3	0	0	0	0
x_2x_4	0	0.02	0	0
x_3x_4	0	0	0	0

Our method detects data-driven causality to filter out non-causal elements from parameter estimation.

Parsimonious Equation Learning with Causality (PELC)

inspired by linear structural equations with GAN
[Kalainathan et al. *J. Mach. Learn. Res.* 2022]

Input 1



Multivariate time series

Time delay

Discriminator

Distance function

Generator

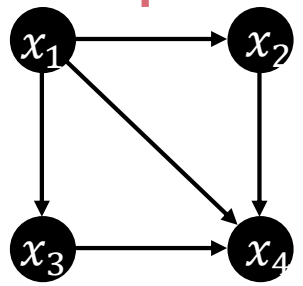
Generated multivariate time series

Causal structure tensor

Causal structure tensor minimizing distance function

Discritization

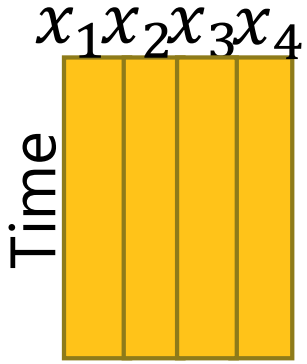
Output 1



Temporal causal net

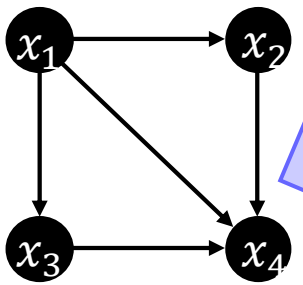
Input 1

Multivariate time series

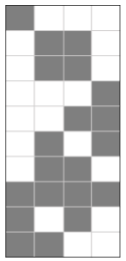


Input 2 (=output 1)

Temporal causal net

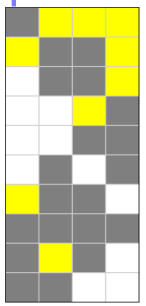


SINDy



Limiting hypothesis space

Coeff vector Ξ



Variable selection, Transforming into ODEs

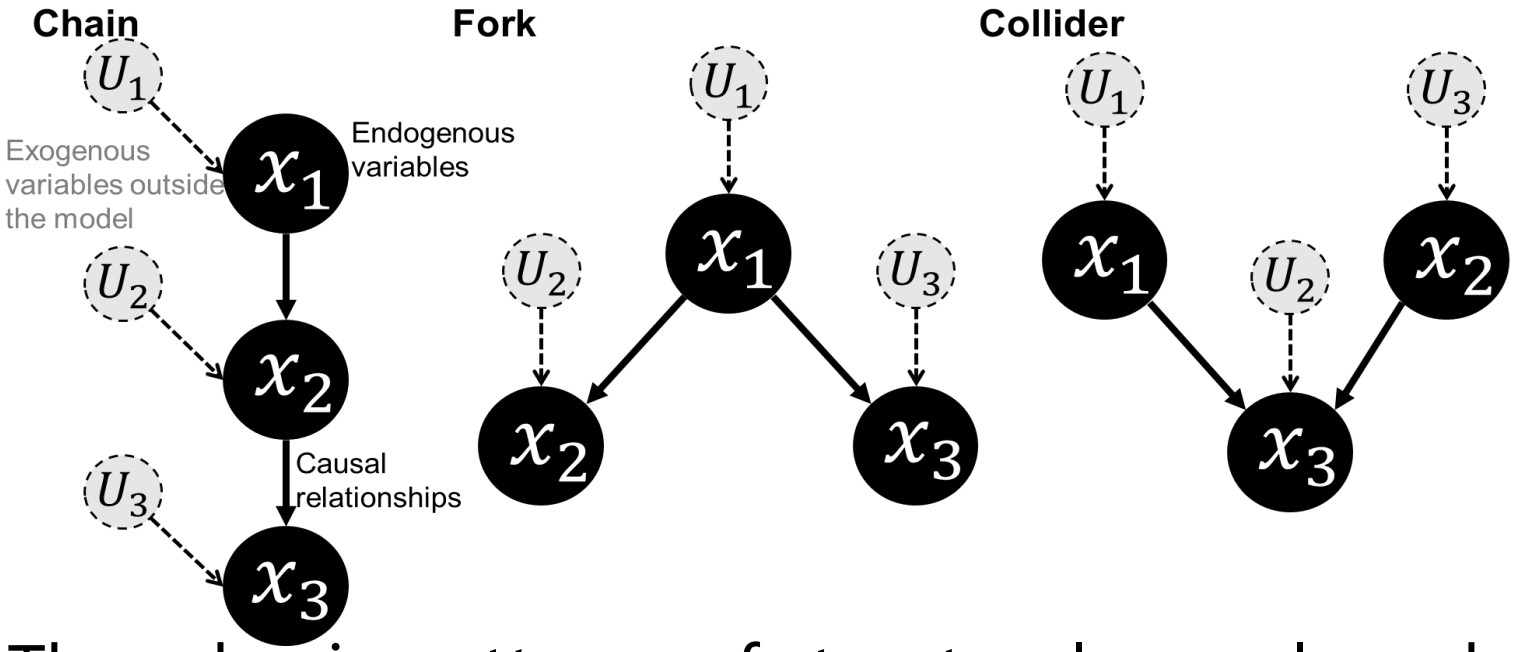
Final output

ODEs $\frac{d}{dt}X = \mathcal{F}(X)$

Experimental settings

Comparative techniques \

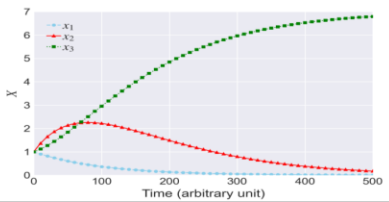
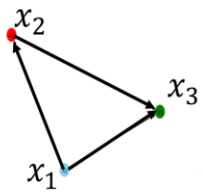
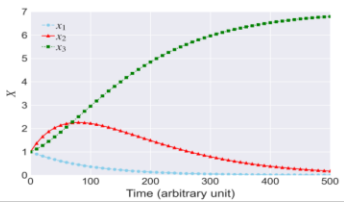
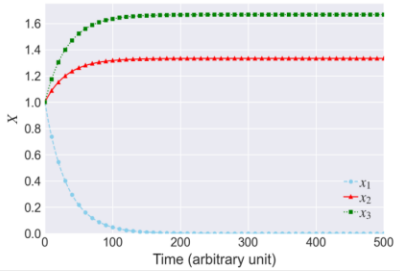
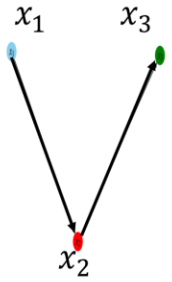
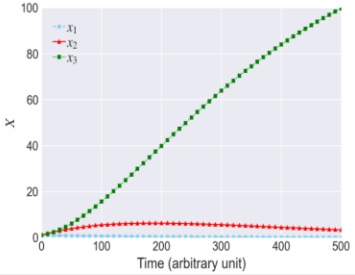
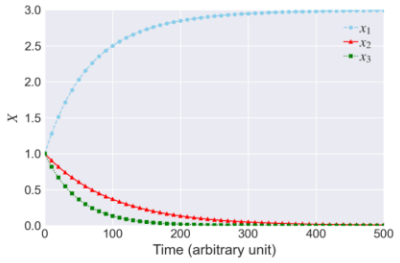
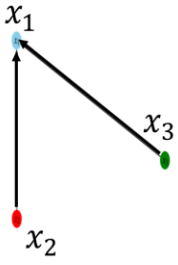
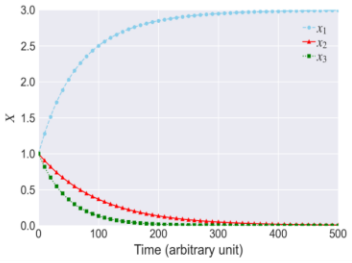
- SINDy (Sparse regression)
- VAR-LiNGAM (Linear structural equation with time delay)
  (Vector Auto-Regression Linear Non-Gaussian Acyclic Model)
  [Hyvärinen+*J. Mach. Learn. Res.*'10] Bootstrap sampling size: 100
- PELC (Proposed method)



Three basic patterns of structural causal models

\ are applied to time-series data generated by three ODEs.

Preliminary results (ongoing) and discussion

	Ground Truth		SINDy	VAR-LiNGAM	PELC		
	ODEs	Original dynamics	ODEs	ODEs	Causal network	ODEs	Reproduced dynamics
Chain	$\dot{x_1} = -0.01\,x_1$ $\dot{x_2} = 0.05\,x_1 - 0.01\,x_2$ $\dot{x_3} = 0.01\,x_2$		$\dot{x_1} = -0.003x_2$ $\dot{x_2} = 0.014x_1 - 0.003x_2 + 0.004x_3$ $\dot{x_3} = 0.003x_2$ RT: 0.00899	$\dot{x_1} = -0.01\,x_1$ $\dot{x_2} = 0.05\,x_1 - 0.01\,x_2$ $\dot{x_3} = 0.01\,x_2$ RT: 4.279		$\dot{x_1} = -0.01\,x_1$ $\dot{x_2} = 0.05\,x_1 - 0.01\,x_2$ $\dot{x_3} = 0.01\,x_2$ RT: 160.392	
Fork	$\dot{x_1} = -0.03\,x_1$ $\dot{x_2} = 0.01\,x_1$ $\dot{x_3} = 0.02\,x_1$		$\dot{x_1} = -0.008x_1$ $\dot{x_2} = 0.000$ $\dot{x_3} = 0.005x_1$ RT: 0.0107	$\dot{x_1} = -0.05\,x_1 + 0.12\,x_2 - 0.09\,x_3$ $\dot{x_2} = 0.13\,x_1 - 0.59x_2 + 0.48x_3$ $\dot{x_3} = -0.08\,x_1 + 0.48x_2 - 0.38x_3$ RT: 4.365		$\dot{x_1} = -0.005x_1$ $\dot{x_2} = 0.08x_1 - 0.005x_2$ $\dot{x_3} = 0.04\,x2$ RT: 159.674	
Collider	$\dot{x_1} = 0.01\,x_2 + 0.02\,x_3$ $\dot{x_2} = -0.01\,x_2$ $\dot{x_3} = -0.02\,x_3$		$\dot{x_1} = -0.024x_2x_3$ $\dot{x_2} = 0.000$ $\dot{x_3} = 0.000$ RT: 0.0137	$\dot{x_1} = 0.01\,x_2 + 0.02\,x_3$ $\dot{x_2} = -0.01\,x_2$ $\dot{x_3} = -0.02\,x_3$ RT: 4.537		$\dot{x_1} = 0.01\,x_2 + 0.02\,x_3$ $\dot{x_2} = -0.01\,x_2$ $\dot{x_3} = -0.02\,x_3$ RT: 160.071	

RT: Run Time (s)

- SINDy did not reproduce the original dynamics in a pilot study (perhaps due to our failure of parameter setting …)
- VAR-LiNGAM and PELC reproduced ODEs of chain and collider SCMs.
- PELC reproduced dynamics with a fewer variables than VAR-LiNGAM.

Summary of Parsimonious Equation Learning with Causality (PELC)

- PELC learns *causal structure tensor*, which represents linear structural equations with time delay.
- Causal network topology is incorporated into hypothesis space of equation discovery as inductive bias.
- Limiting hypothesis space by replacing non-causal elements of coefficient vectors with zero

Open questions

- Is it plausible to associate **causality** with **differential equations**?
- If so, what is the best definition of causality?
  (Granger causality, transfer entropy, conditional probability, etc.)

Table of contents

Related work on causal discovery

- Statistical approach (based on Strctural Equation Modeling):
  LiNGAM (Linear Non-Gaussian Acyclic Model) [Shimizu+J. Mach. Learn. Res.'06]
- Bayesian approach: Structure learning of Bayesian networks
- ML approach: DAG-GNN [Yu et al. *ICML* 2019]
  Causal discovery w/ reinforcement learning [Zhu et al. *ICLR* 2020]

Related work on equation discovery

- Constraining hypothesis space with context-free grammar [Todorovski & Džeroski. *ICML.* 1997]
- Constraint system identification with human-computer communication [Stolle & Bradley. 2007]
- Process model induction from data + knowledge on observed behavior [Bridewell et al. *Mach.Learn.*2008]
- Genetic learning of free-form natural laws [Schmidt & Lipson. *Science.* 2009]
- Physics-informed learning from small data [Raissi et al. *J. Comput. Phys.* 2019]
- Learning symmetry and separability in data (AI-Feynman) [Udrescu & Tegmark. *Sci. Adv.* 2020]

Their relationships

- Transforming ODEs into deterministic SCMs via equilibrium equations under intervention
  [Mooij et al. *UAI* 2013]
- Constructing SCMs by reproducing asymptotic behavior of ODEs under intervention
  [Rubenstein et al. *UAI* 2018]
- 1-to-1 correspondence between ODEs and SCMs has not been adequately developed.

Concluding remarks

❑ Summary of contributions

- ➢ Data-driven & knowledge-based (DD-KB) integrated approach to uncovering dynamics
- ➢ Case study: Gene network inference for constructing COVID-19 pathways
- ➢ PELC: Causal network discovery w/ GAN + Mapping discovered causal structure onto hypothesis space in equation discovery

❑ Limitation

- ➢ Parameter sensitivity
- ➢ Generalization

❑ Future work

- ➢ Sensitivity analysis
- ➢ Learning from partial or small data
- ➢ Neuro-symbolic AI: Realizing causal & equation discovery in the same hypothesis space
- ➢ High-dimensional synthetic data experiments, Real-world dataset applications