Non-linear Sensitivity Analysis for Interpretable Structure Discovery

Theo Bourdais

PhD Student, Computing + Mathematical Sciences

SIAM Annual meeting

July 9th 2026

Caltech

Collaborators and collaborations

Caltech


Houman Owhadi


Ricardo Baptista


Zongren Zou

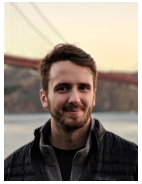

Pau Batlle Franch


Xianjin Yang

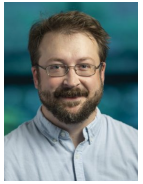
JPL


Nicolas Rouquette

Sandia National Lab


Alexey Voronin


Elise Walker


Nathaniel Trask

Bourdais, T., Batlle, P., Yang, X., Baptista, R., Rouquette, N., & Owhadi, H. (2024).
Codiscovering graphical structure and functional relationships within data: A gaussian process framework for connecting the dots.
Proceedings of the National Academy of Sciences

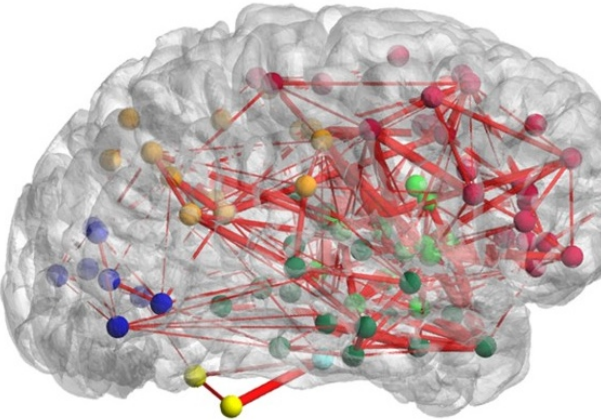
Zou, Z., Bourdais, T., Baptista, R., & Owhadi, H. (2026).
Graphical conditional generative modeling for digital twin modeling.
arXiv preprint arXiv:2606.16219.

Voronin, A., Walker, E., Bourdais, T., Trask, N., Owhadi, H. (2026).
DG2DAG: Refining prior skeletons under acyclicity for predictive surrogate construction
To appear

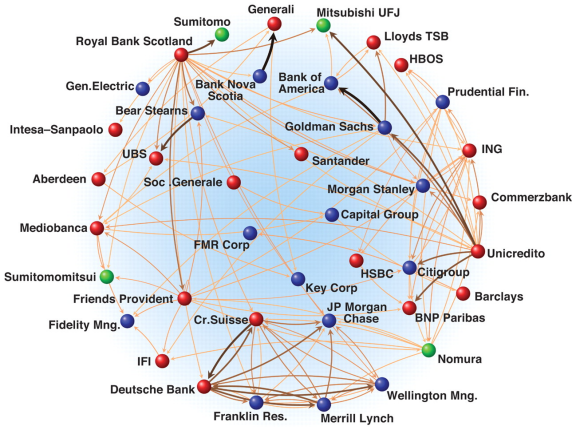
Caltech

Motivation

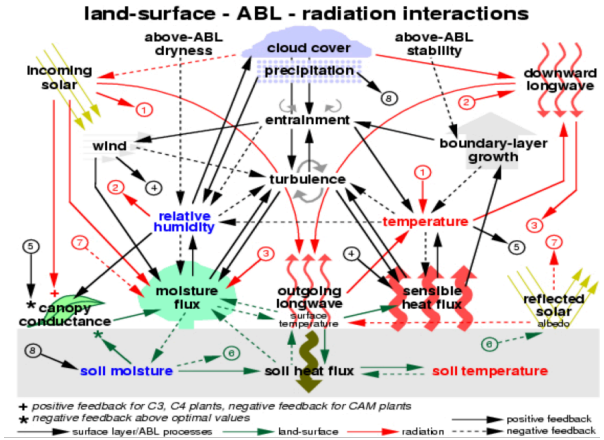
Complex systems are understood through graphs



Brain activation network



Systemic banking risks



Climate systems

Caltech

Graph discovery

Recover graphs using data

We will:

- Find the functional dependencies between the variables (the edges)
- Make the graph as sparse as possible while accurate

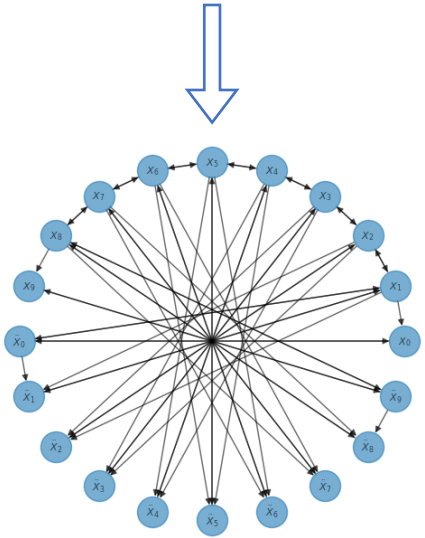
Why sparse?

- Generalization + analysis tool

Note: we recover explicit equations, i.e.

$x_i = f(x_{\_i})$

x_1	$\ddot{x}_1$	. . .	$\ddot{x}_{10}$
0.45	0.66	. . .	-0.23
⋮	⋮	⋱	⋮
-0.78	-0.12	. . .	0.89



Caltech

Existing work on graph discovery

- Causal discovery methods (discussed later)
- Sparse regression (e.g. Sindy)

Caltech

Pruning algorithm outline

We use a simple greedy algorithm to get a sparse and accurate graph.

- Start from fully connected graph
- For each node:
  - While fidelity is high:
    - Approximate node as a function of ancestors
    - Find least important ancestor variable
    - Remove it

Goodness-of-fit metric

Gaussian process regression

Importance metric

We recover $x_i = f_i(\text{ancestors}(i))$, where $\text{ancestors}(i) \subset x_{-i}$,

This reduces to sparse regression $y = f(x)$

Caltech

Linear case

Straightforward linear metrics do not generalize

In the Ordinary Least Squares (OLS) case:

$$y = x^T\beta + \epsilon,\ \epsilon \sim \mathcal{N}(0,\sigma^2)$$

With n data points (X, Y), mean $\bar{y}$ and estimate $\hat{y}_i = x_i^T\hat{\beta}$, quality of fit is measured using:

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 1 - \frac{\|Y - X\beta\|^2}{n\text{Var}[Y]}$$

Thanks to orthogonality of noise and signal.

This breaks down in non-linear case

Caltech

GP regression and noise-to-signal ratio

Gaussian processes have a natural goodness-of-fit metric

To recover f s.t. $y = f(x_1, \ldots, x_n)$, with $Y = (y^k)_{k=1}^N$ and $X = (x_1^k, \ldots x_n^k)_{k=1}^N$, then

$$\hat{f} = \arg\min_{\phi \in \mathcal{H}_k} \|Y - \phi(X)\|^2 + \gamma \|\phi\|^2$$

- γ is the regularization
- $\mathcal{H}_k$ is defined by the kernel k:
  - k linear → Linear regression
  - k quadratic → Quadratic regression
  - …

Caltech

GP regression and noise-to-signal ratio

Gaussian processes have a natural goodness-of-fit metric

To recover f s.t. $y = f(x_1, \ldots, x_n)$, with $Y = (y^k)_{k=1}^N$ and $X = (x_1^k, \ldots x_n^k)_{k=1}^N$, then

$$\hat{f} = \arg\min_{\phi \in \mathcal{H}_k} \|Y - \phi(X)\|^2 + \gamma \|\phi\|^2$$

Noise | Signal

Noise has two sources:

- True noise ($Y = \hat{f}(X) + \epsilon$)
- Model misspecification

Goodness of fit: $\dfrac{\text{Noise}}{\text{Noise} + \text{Signal}}$

Caltech

Activation

Quadratic kernel

$$k(x, y) = (1 + \langle x; y \rangle)^2 = (1 + x_2 y_2)^2 + 2x_2 y_2 \langle x_{-2}; y_{-2} \rangle + k_{-2}$$

GPs have natural contribution scores

If $y = \hat{f}(x_1, \ldots, x_5)$ with a kernel k s.t.

$$k = k_2 + k_{-2}$$

depends on x_2 | does not depend on x_2

Then we can compute $\hat{f}_2$ and $\hat{f}_{-2}$ s.t.

$$\hat{f} = \hat{f}_2 + \hat{f}_{-2}$$

depends on x_2 | does not depend on x_2

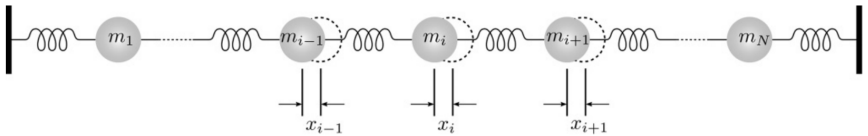
Since $\|\hat{f}\|_k^2 = \|\hat{f}_2\|_{k_2}^2 + \|\hat{f}_{-2}\|_{k_{-2}}^2$, we define the contribution of x_2 as

$$a_2 = \frac{\|\hat{f}_2\|_{k_2}^2}{\|\hat{f}\|_k^2}$$

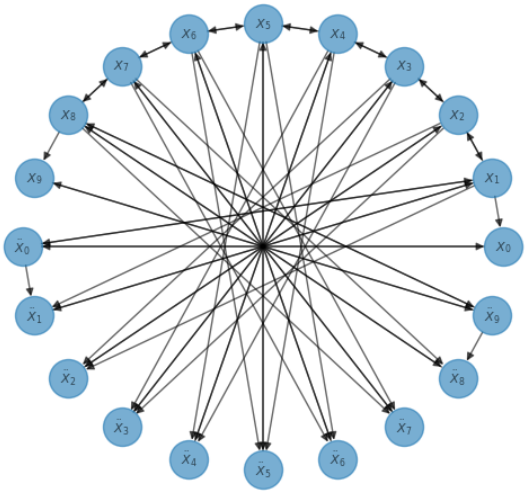
Caltech

Example: Fermi-Pasta-Ulam-Tsingou

A toy example with sparse known graph and uninformative prior



$$\ddot{x}_i = \frac{c^2}{h^2}(x_{i+1} + x_{i-1} - 2x_i)(1 + (x_{i+1} - x_{i-1})^2)$$

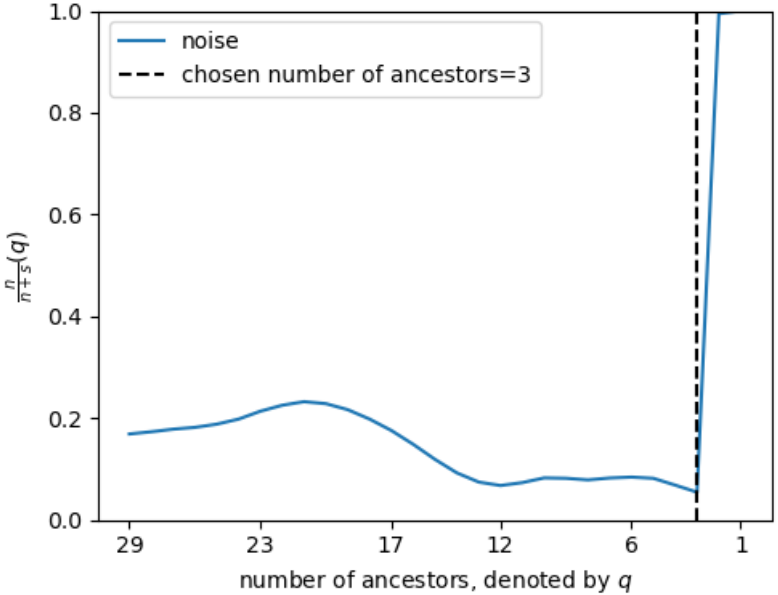


We observe 1000 snapshots of x_i, $\dot{x}_i$, $\ddot{x}_i$, and get perfect recovery

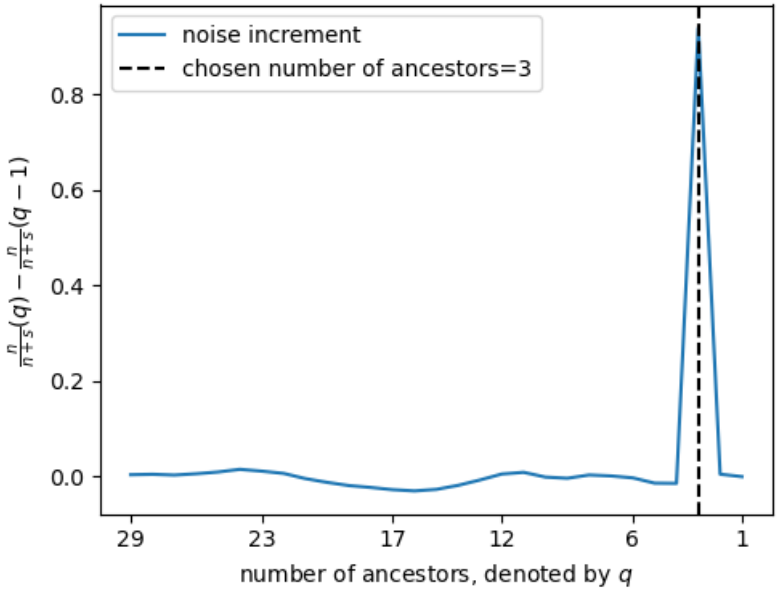
Caltech

Typical evolution of noise-to-signal ratio ($\ddot{x}_7$)

When pruning true ancestors, noise goes from low to high and we observe the largest increment



Evolution of noise-to-signal ratio



Increments $\frac{n}{n+s}(q) - \frac{n}{n+s}(q-1)$ for q the number of ancestors

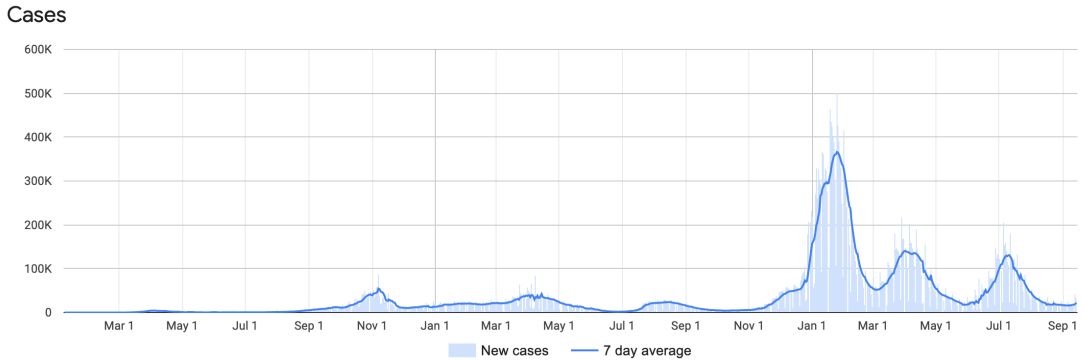
Caltech

Example: COVID dataset in France

Real world dataset

Daily values of 31 variables during 500 days:

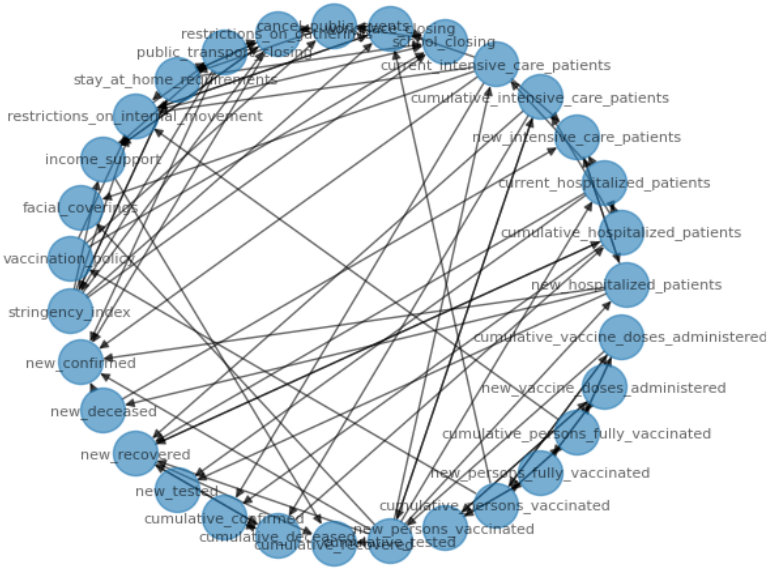
- Epidemiology dataset
- Hospital dataset
- Vaccine dataset
- Policy dataset



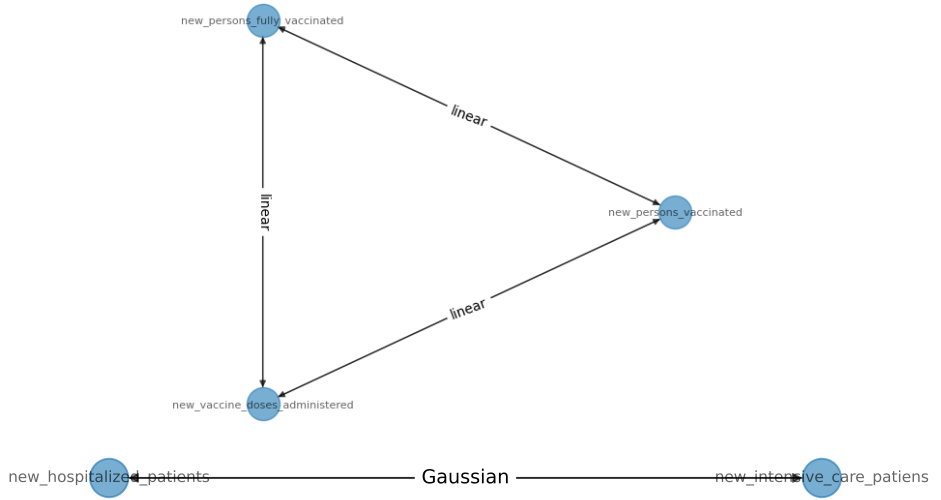
Caltech

Example: COVID dataset in France

Full graph



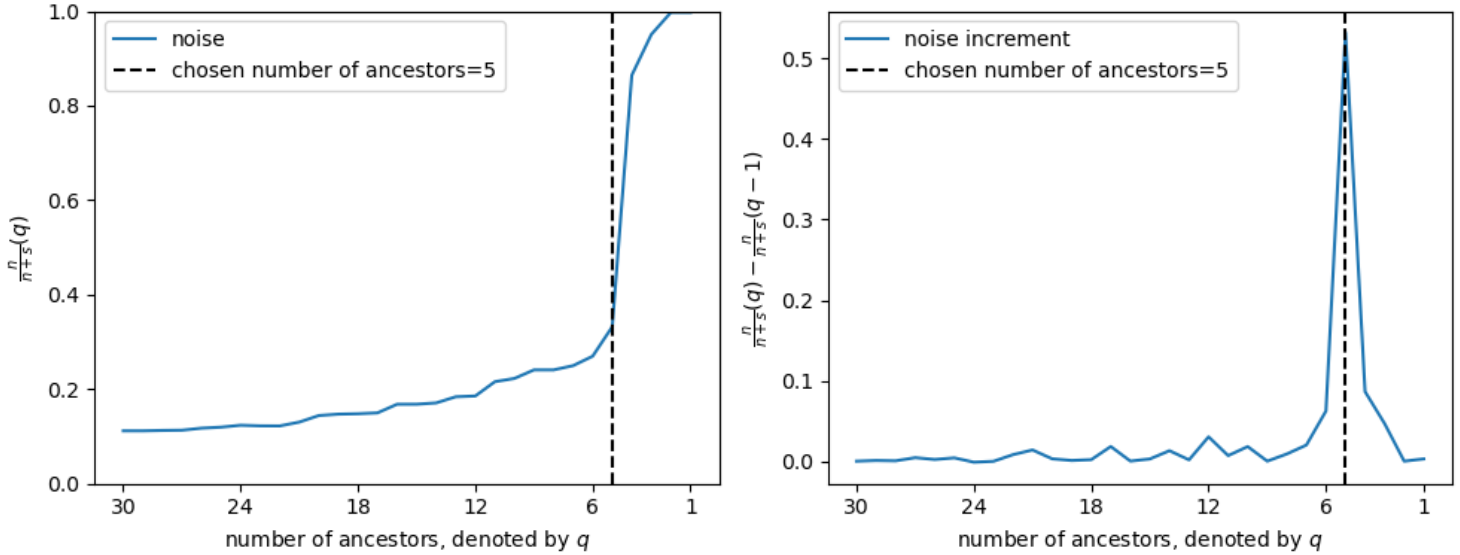
Subgraphs



Caltech

Example: COVID dataset in France

We observe the same behavior on real data as in toy example



Evolution of the noise-to-signal ratio when pruning ancestors for the cumulative number of hospitalized patients (quadratic kernel).

Caltech

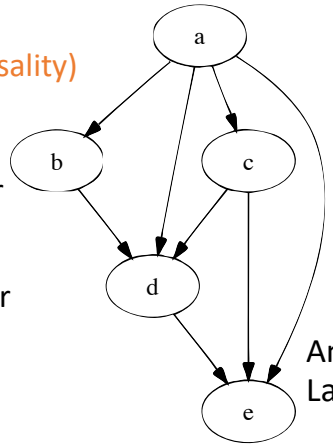
Connection to causal discovery

NP-Hard

Causal discovery (usually)

- Search for the best DAG (Directed Acyclic Graph)
- Sophisticated graph search algorithms based on: independence tests, scores…
- Stronger Gaussian priors
  - Often linear Gaussian (close to linear regression) (linear is the worst for causality)
  - Other prior include
    - $y = f(\sum_i \alpha_i x_i)$ with f non-linear
    - $y = \sum_i \alpha_i f_i(x_i)$ with f_i non-linear

$y = x_1 x_2$ is not expressible this way



Our method

- Search for the best DG (potentially cyclic)
- Simple, greedy pruning method
- Weaker Gaussian prior
  - Includes linear, quadratic etc…
  - Combinatorial Gaussian kernel:

$$k(x,y) = \prod_i \left(1 + e^{\|x_i - y_i\|^2}\right)$$

$$k(x,y) = \sum_{I \subset \{1,\ldots,k\}} e^{\|x_I - y_I\|^2}$$

Polynomial

An extension of our work is in use at Sandia National Labs

Caltech

Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of causal discovery methods based on graphical models. *Frontiers in genetics*, *10*, 524.

Digital twins use case

Digital twins require a reduced set of variables, which makes the system stochastic

Let a complex system q we want to predict and control, s.t.

$$\dot{q}(t) = f(q, \omega)$$

With ω representing noise, the environment etc. We are interested of finding a reduced set of variables of interest $q_{\mathcal{J}}$ and approximate update

$$q_{\mathcal{J}}(t + \Delta t) = \Phi(q_{\mathcal{J}}(t), \eta)$$

Where η represents all unknowns, which are interpreted as random. $\Phi(q_{\mathcal{J}}, \cdot\,)$ is the conditional distribution of interest

Caltech

From functional dependencies to distributional

Complex systems are approximated using distributional dependency

We have assumed and recovered noisy functional dependencies

$$y = f(x) + \epsilon,\ \epsilon \sim \mathcal{N}\left(0, \sigma^2\right)$$

Can we generalize to any dependency?

- Use flow matching (a stochastic interpolant) to find T s.t.

$$y \overset{d}{=} T(\epsilon, x),\ \epsilon \sim \mathcal{N}(0,1)$$

- Perform sensitivity analysis on T

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., & Le, M. (2022). Flow matching for generative modeling. *arXiv preprint*
Albergo, M. S., Boffi, N. M., & Vanden-Eijnden, E. (2023). Stochastic interpolants: A unifying framework for flows and diffusions. *arXiv preprint*

Caltech

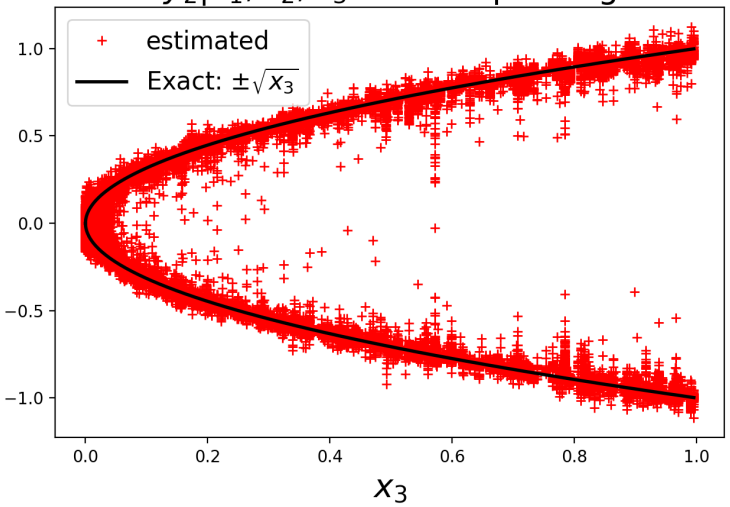
Algebraic equations

Stochastic modeling is useful when the relationships are not bijective

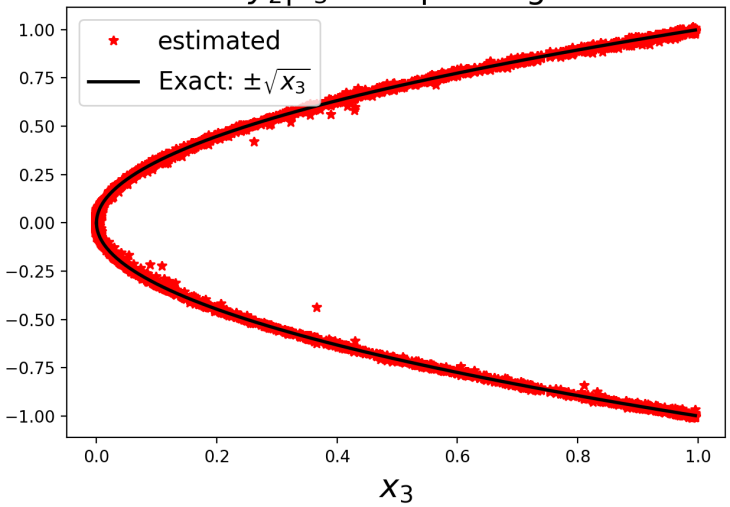
$$
\begin{aligned}
y_1 &= x_1\omega, \\
y_2^2 &= x_3, \\
y_3 &= \sqrt{1-\alpha(x_2)}\xi + \sqrt{\alpha(x_2)}\eta,
\end{aligned}
$$

where $\alpha(x) \in [0,1]$, $\omega, \xi \sim N(0,1)$, and $\eta \sim \text{Laplace}(0, 1/\sqrt{2})$

$y_2|x_1, x_2, x_3$ without pruning

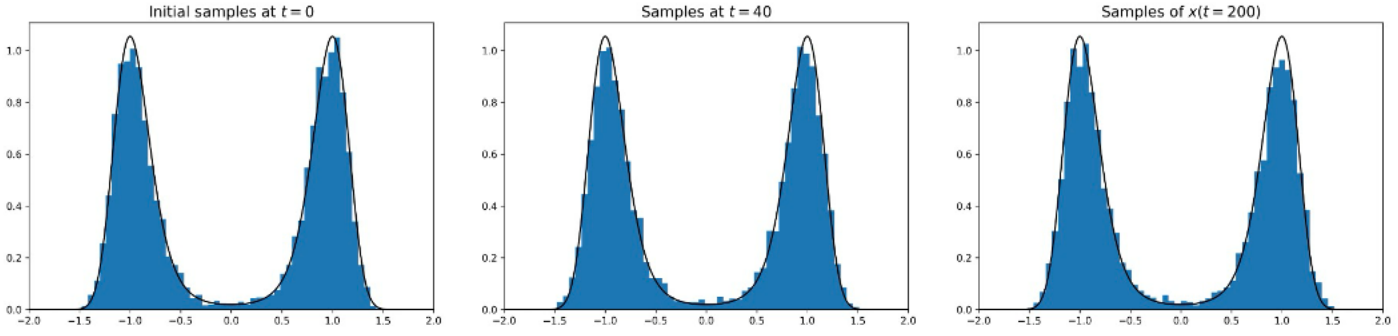


$y_2|x_3$ with pruning

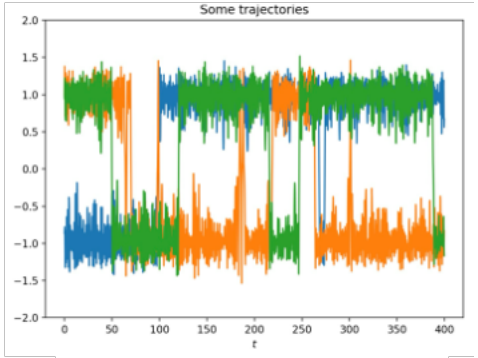


Double well potential

Stochastic modeling is natural here



(c) Generated samples of x at different time.

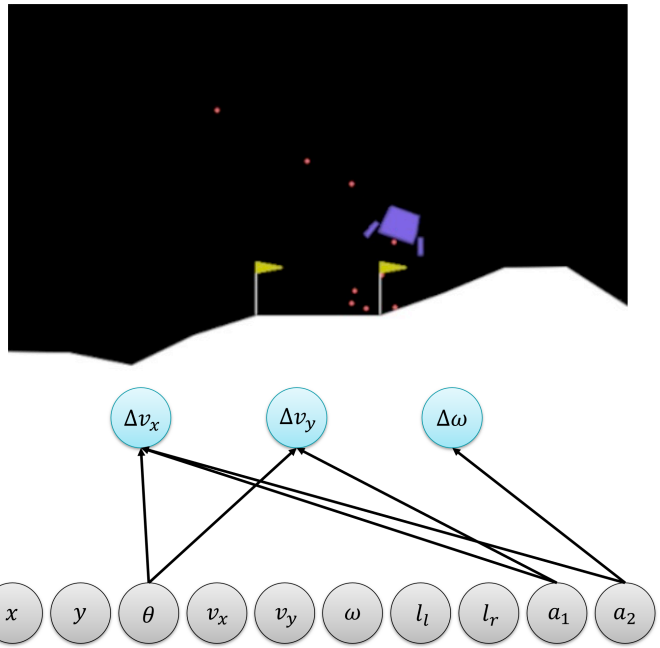


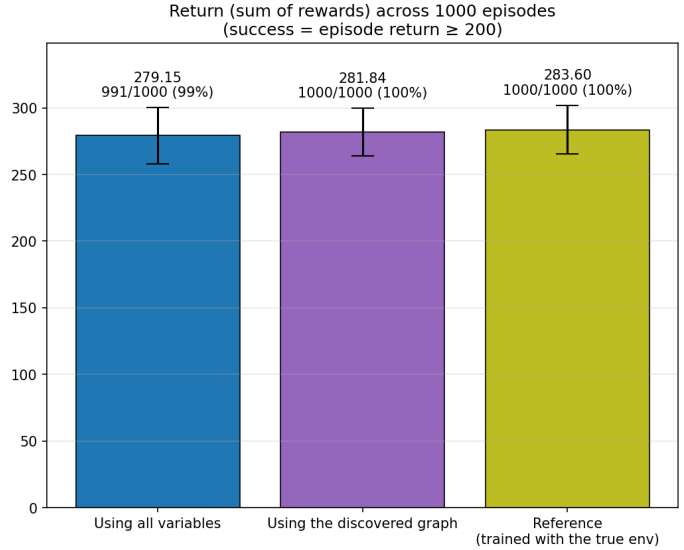
(d) Generated trajectories.

Caltech

Lunar lander

We replace the physics engine with a pruned graph, train a RL agent, and get the same performance

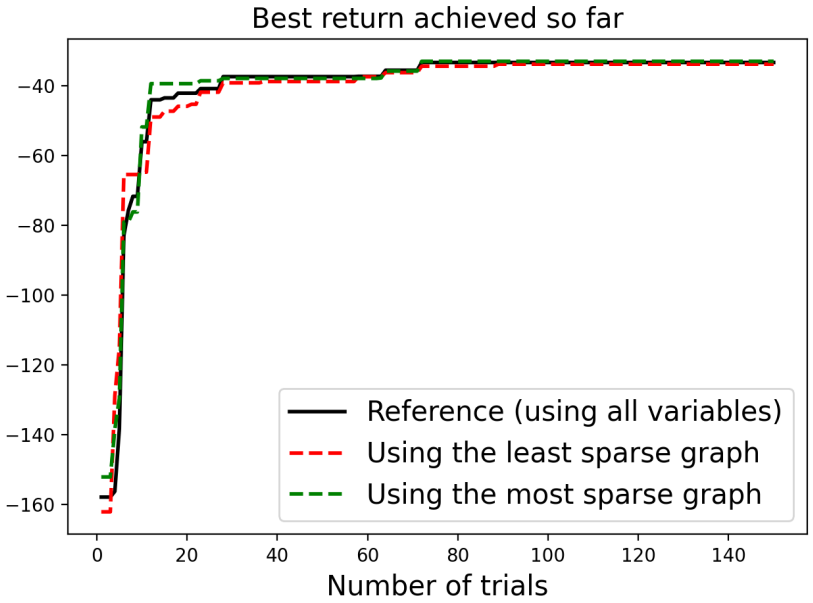


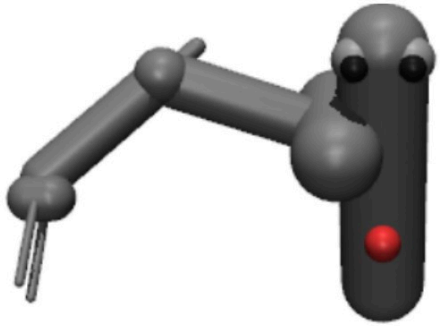


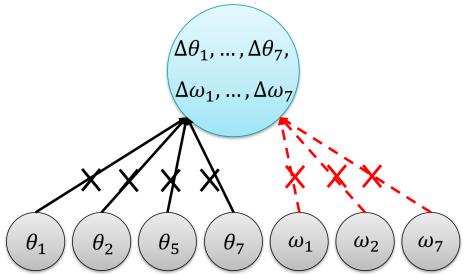
Caltech

7 degrees-of-freedom reacher

Removing variables doesn't affect downstream performance







Caltech

Conclusion

- We developed a Gaussian Process-based framework to recover dependencies between variables
- Recovers known equations in toy examples
- Yields plausible results for real datasets
- Allows for approximation and control of complex systems

Caltech