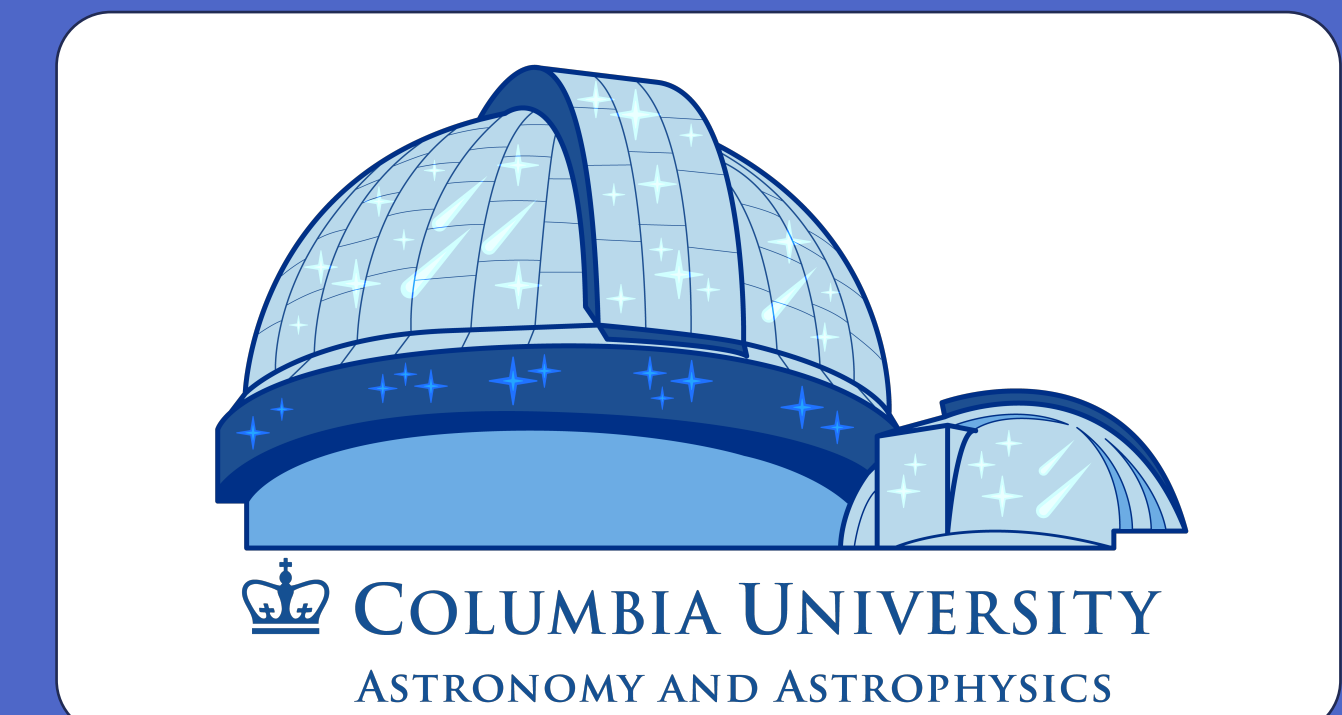


DegenDetector: Symbolic Recovery of Parameter Degeneracies in Bayesian Posteriors

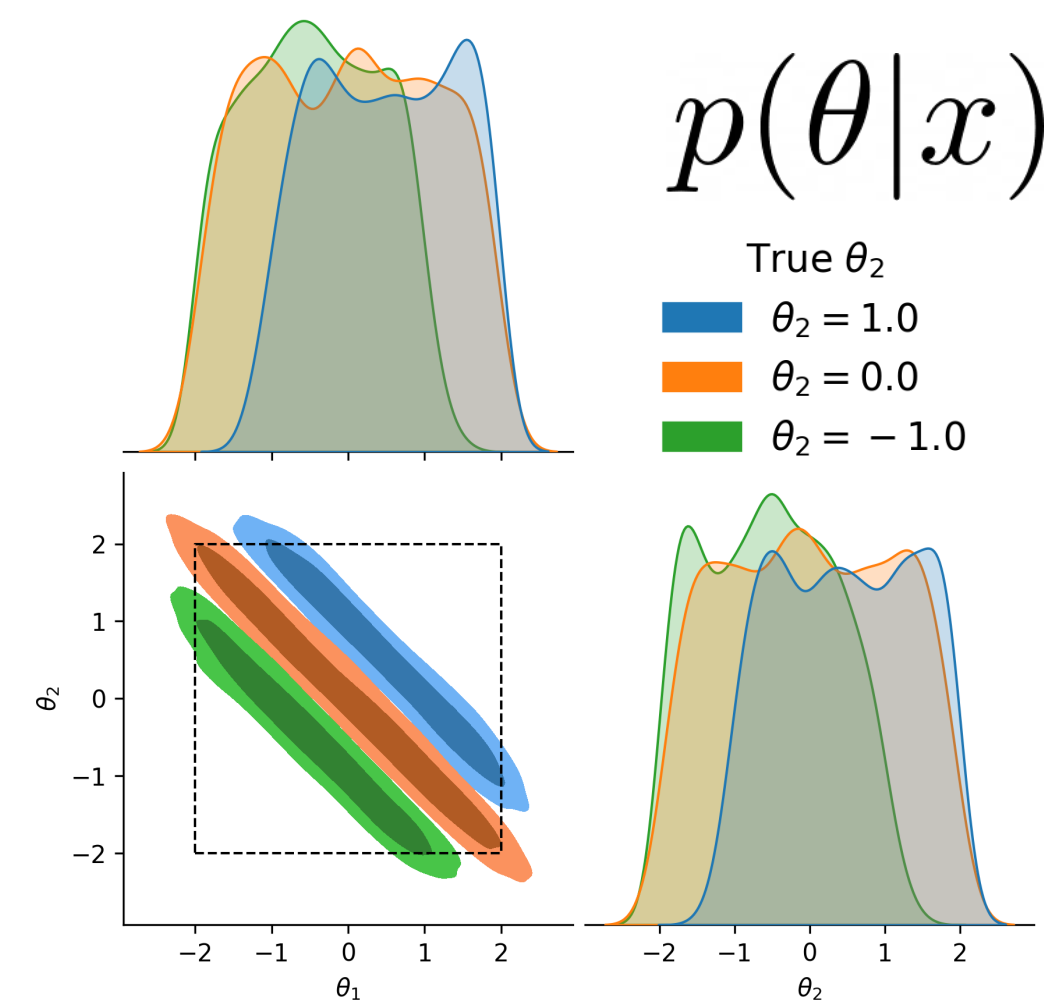
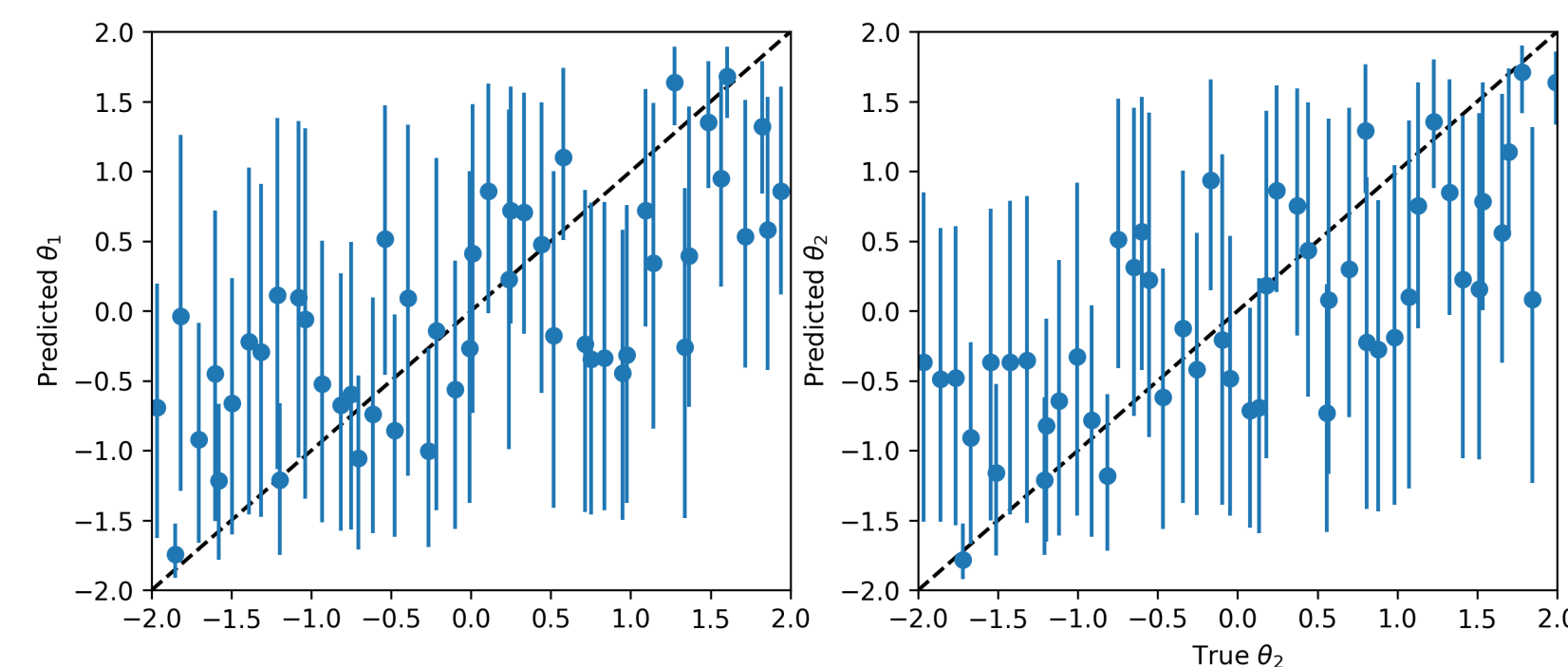
Chaipat Tirapongprasert, Matthew Ho
Columbia Department of Astronomy, Columbia Astrophysics Lab

Check out our code!



The Problem

When data constrains a combination of parameters rather than individuals, posterior mass concentrates on manifolds — a parameter degeneracy. These are notoriously hard to detect from marginals alone."



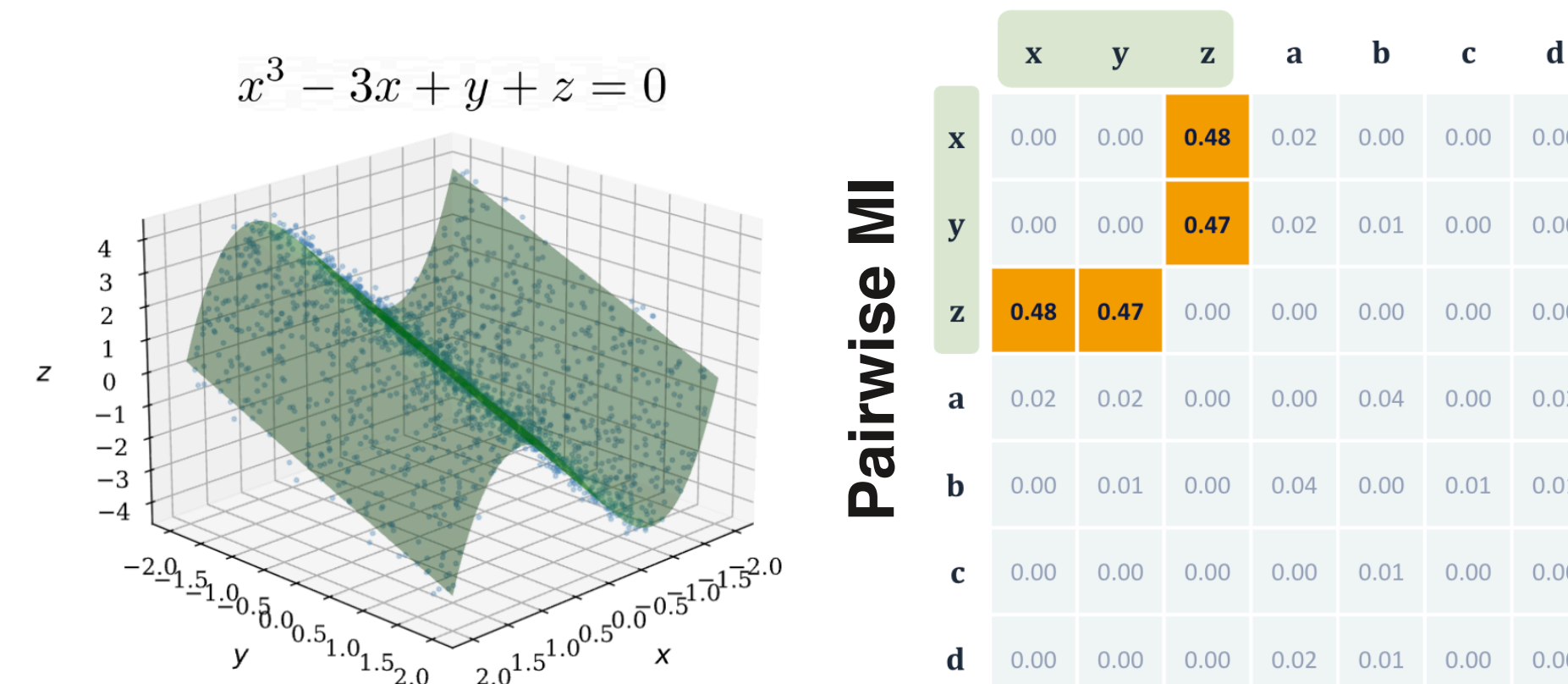
Especially in high-dimensional data, it can be difficult to detect the degeneracy behavior purely from the marginal distributions.

Key Contribution

Given posterior samples, DegenDetector automatically detects and outputs closed-form symbolic expressions describing the degeneracy manifolds.

Methodology

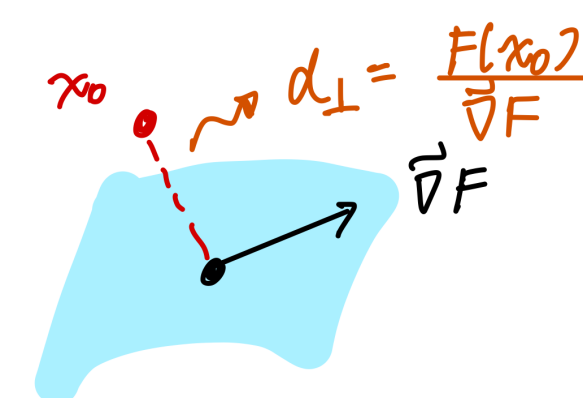
1) Mutual Information (MI) Screening
narrow down the parameter combinations



2) Greedy Symbolic Regression
assume separable and fit one component at a time

Problem: The space of formulas grows combinatorially as $\mathcal{O}(k^k)$.
Solution: Split into k independent one-dimensional formulas.

$$g_1(\theta_{j_1}) + g_2(\theta_{j_2}) + \dots + g_k(\theta_{j_k}) - c = 0$$



$$\mathcal{L}_\perp = \frac{1}{N} \sum_{i=1}^N \frac{[\sum_l g_l(\theta_l^{(i)}) - c]^2}{\sum_l [g'_l(\theta_l^{(i)})]^2 + \epsilon}$$

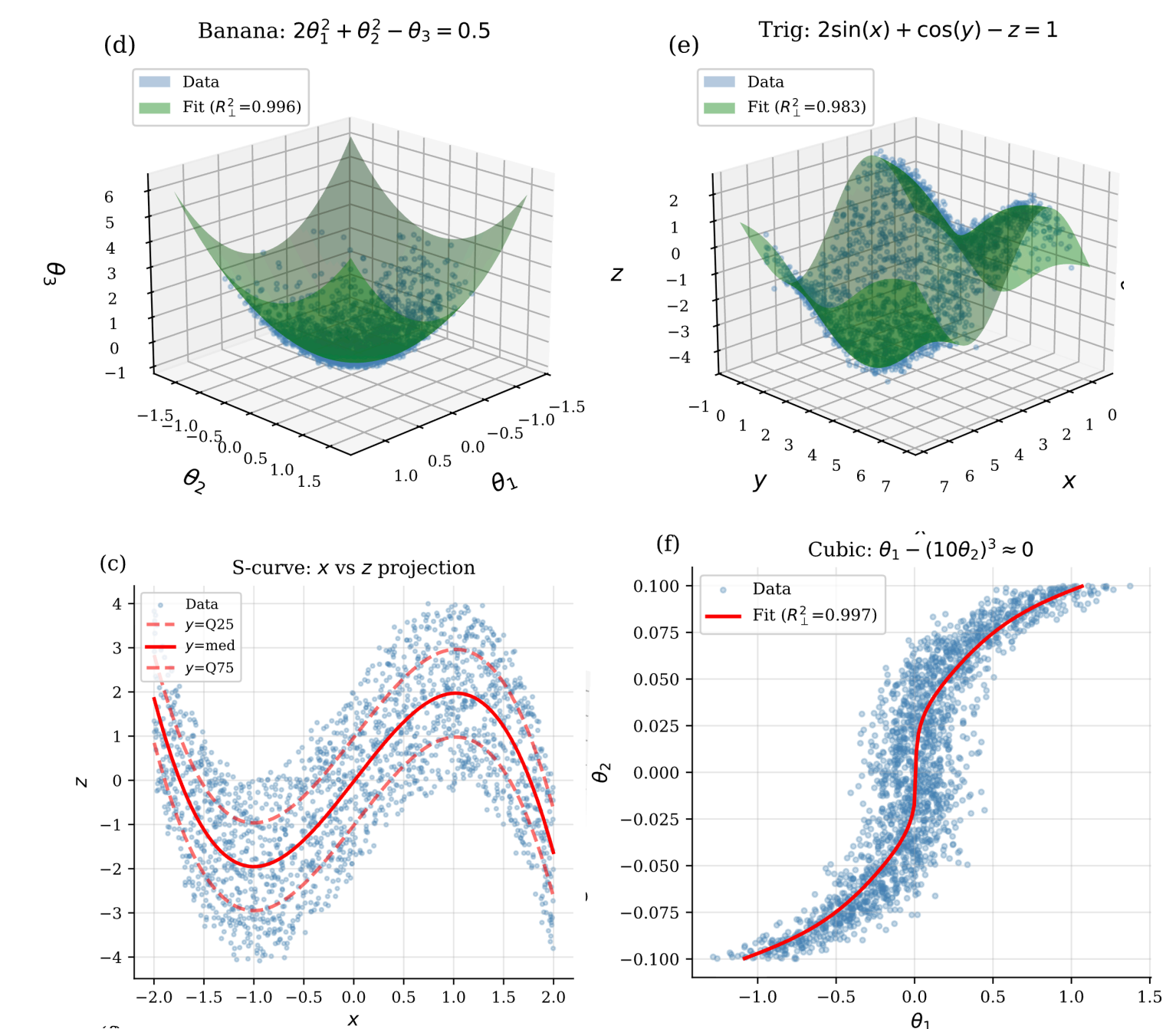
Fit univariate functions to minimize orthogonal R^2

3) Diagnostics
show top symbolic equations and diagnostic visualizations

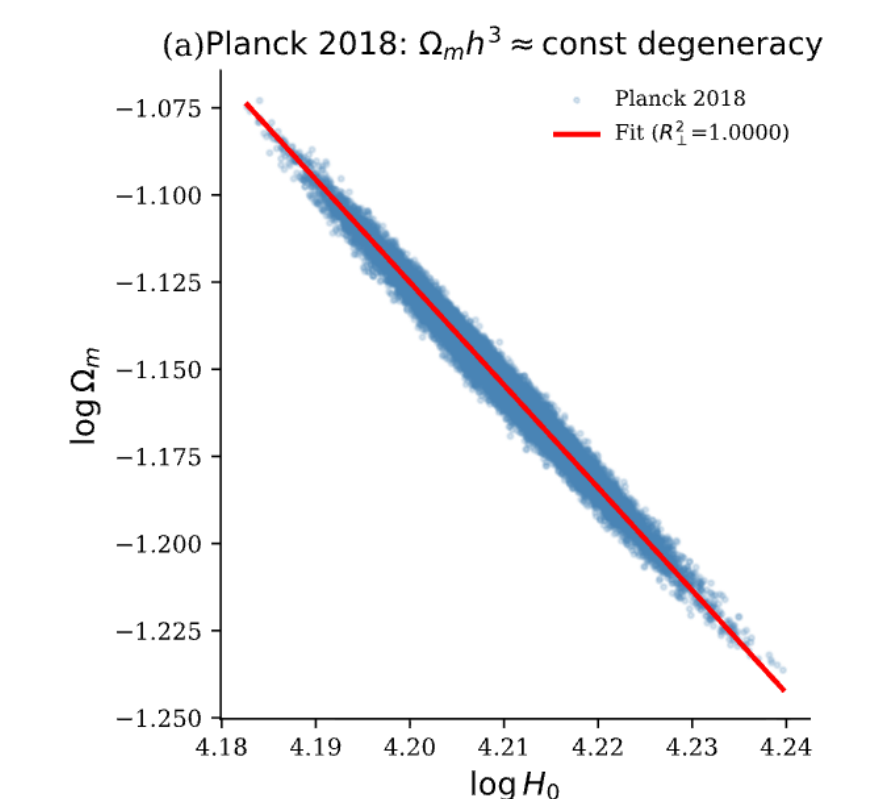
Fitted equations (ranked by MI score):
[1] Parameters: (x, y, z)
MI score: 0.9525
Candidates (5 total, ranked by functional form consensus then R^2_{ortho} ; top form: 4/5 agree, 2 distinct form(s)):
Top form: $(c_1*x - c_2)*((c_1*x - c_2)**2 - c_3) + c_4*y + c_5 + c_6*z - c_7 = c_8$
[1] $(0.86*x - 0.01)*((0.86*x - 0.01)**2 - 2.28) + 0.67*y + 0.018 + 0.67*z - 0.00 = 0.01$
 R^2_{ortho} : 0.99
Residual std: 0.14
Complexity: 11

Experiments & Results

Synthetic Benchmarks



LogDegen
Finding the CMB horizon-angle degeneracy



$$123.97 \log H_0 + 42.07 \log \Omega_m = \text{const}$$

Future Work

- Automatic reduction of symbolic expressions
- Extend to very high-dimensional inference and non-separable symbolic expressions
- Construct global reparameterizations
- Apply to simulation-based inference problems