Optimization and Machine Learning

Lecture 7: MINLP models for symbolic expressions

Sanjeeb Dash
IBM Research
(co-lecturer Parikshit Ram)

JPOC 13 Summer School, June 26-28, 2023
University of Clermont-Ferrand

Lecture 7 Outline

- Symbolic regression
- MINLP models
- Combining reasoning and regression
- Applications to real scientific data
- Polynomial optimization
- Numerical Experiments

Derivable scientific discovery

Goal: Discover "meaningful" scientific models from experimental data

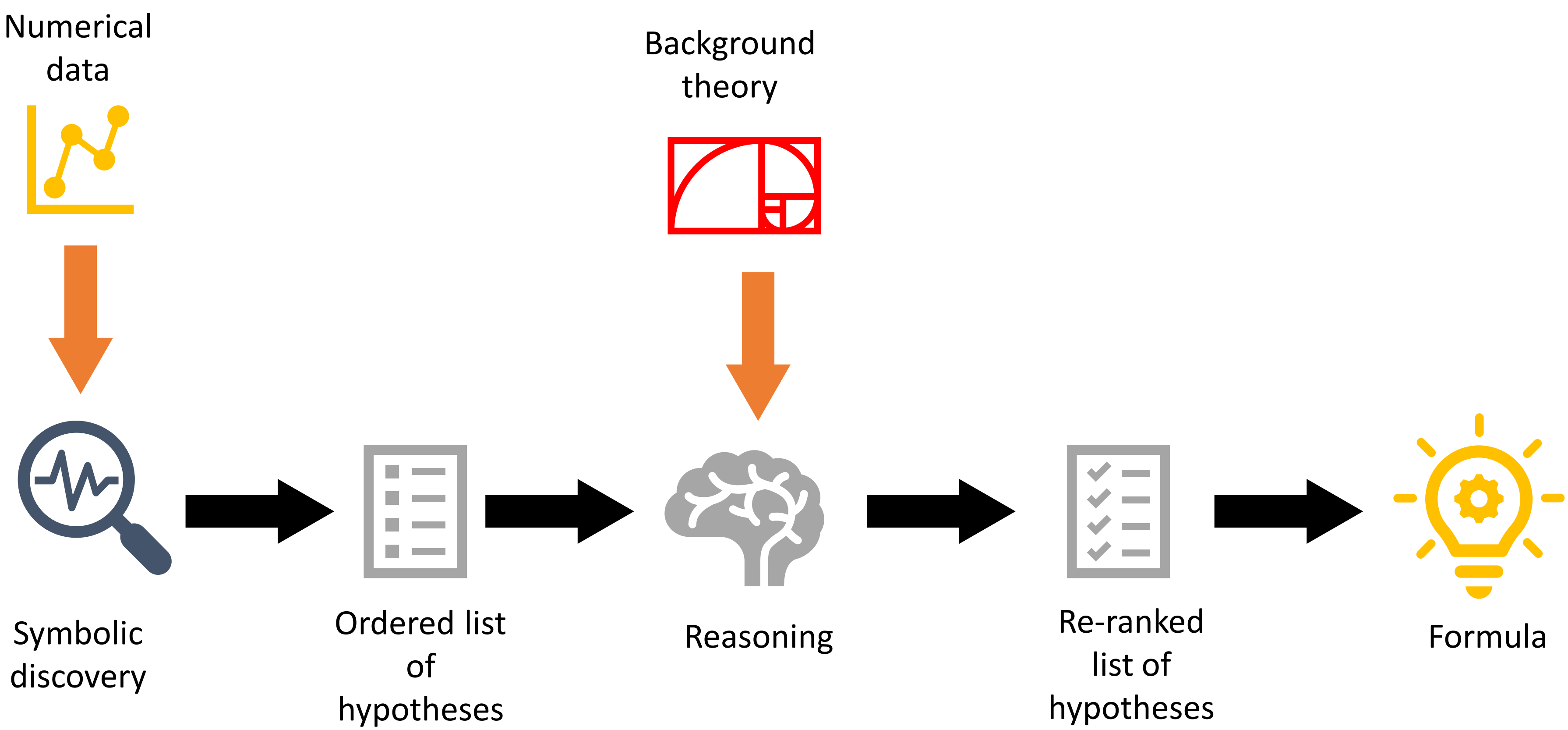
NNs:
- good for discovery of patterns and relations in data
- drawback: "black-box" models

Standard regression:
- the functional form is given, discovery = parameter fitting

Symbolic regression:
- the functional form is not given but is instead composed from the data
- models are more "interpretable" and require less data

Cornelio, Dash, Josephson, Goncalves, Austel, Clarkson, Megiddo, Horesh, Combining data and theory for derivable scientific discovery with AI-Descartes. Nature Comm. 2023

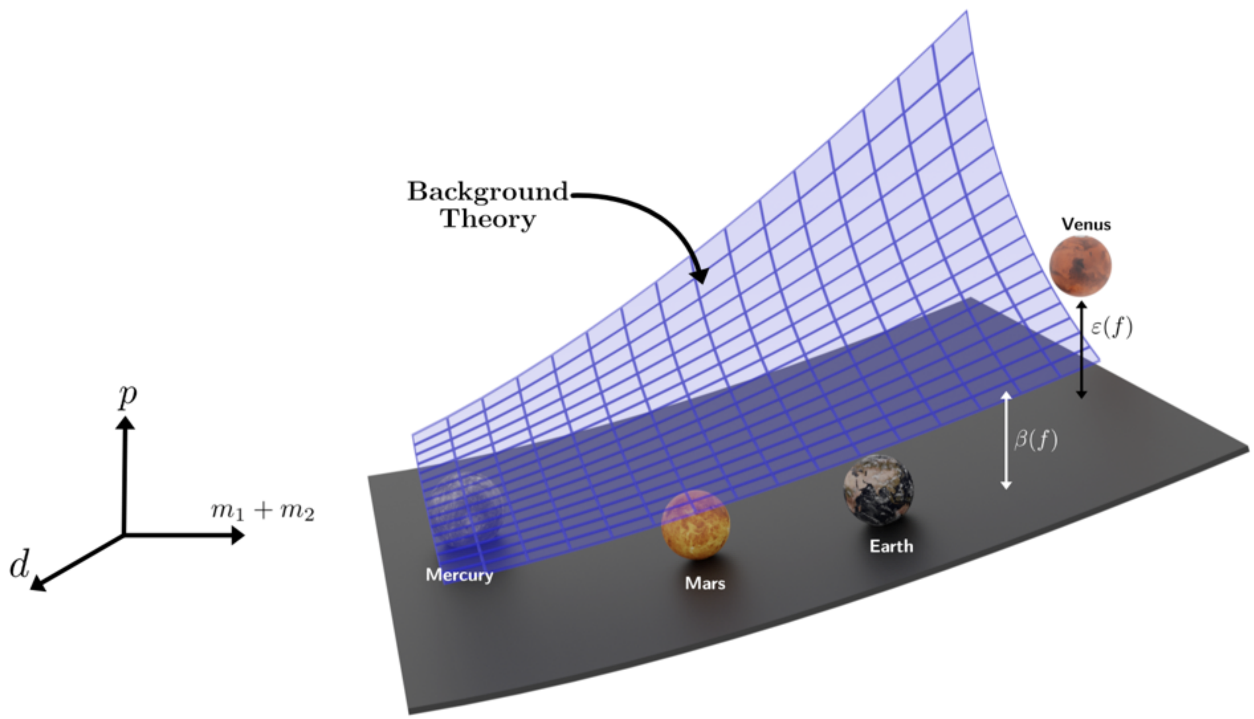
System overview



Given observational data, the goal is to discover an interpretable, mathematical model in a symbolic form that is consistent with background theory

Main Idea

Unify symbolic regression with formal reasoning
- Provide a proof of derivability of a formula produced from data OR
- Calculate how close a formula is to a derivable formula.



Results

Label	Formula	AI-Descartes	AI Feynman	PySR	BMS
I.6.20a	$e^{-\theta^2/2}/\sqrt{2\pi}$	X	X	X	X
I.6.20	$e^{-\frac{\theta^2}{2\sigma^2}}/\sqrt{2\pi\sigma^2}$	X	X	X	X
I.6.20b	$e^{-\frac{(\theta-\theta_1)^2}{2\sigma^2}}/\sqrt{2\pi\sigma^2}$	X	X	X	X
I.8.14	$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$	X	X	X	X
I.9.18	$\frac{Gm_1m_2}{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$	X	X	X	X
I.10.7	$\frac{m_0}{\sqrt{1-v^2/c^2}}$	✓	X	X	X
I.11.19	$x_1y_1+x_2y_2+x_3y_3$	X	X	X	X
I.12.1	μN_n	✓	$✓^2$	✓	✓
I.12.2	$q_1q_2/(4\pi\varepsilon r^2)$	$✓^1$	X	$✓^1$	$✓^1$
I.12.4	$q_1/(4\pi\varepsilon r^2)$	$✓^1$	$✓^1$	$✓^1$	$✓^1$
I.12.5	q_2E_f	$✓^1$	$✓^2$	✓	✓
I.13.4	$\frac{1}{2}m(v^2+u^2+w^2)$	X	X	X	X

	AI-Descartes	AI Feynman	PySR	BMS
Number of (✓, $✓^1$, $✓^2$, $✓^3$, X)	(13, 32, 4, 0, 32)	(0, 25, 8, 0, 48)	(16, 21, 2, 0, 41)	(10, 17, 11, 1, 42)
Total $✓^*$	49/81	33/81	40/81	39/81
Accuracy	**60.49%**	40.74%	49.38%	48.15%

Supplementary Table 13. Results on 81/100 problems from the Feynman Database for Symbolic Regression (problems not containing trigonometric functions). The accuracy of the best method is marked with bold font.

Regression

Symbolic Regression: Given $\mathbf{X}^1, \ldots, \mathbf{X}^k \in \mathbb{R}^n$ and $Y^1, \ldots, Y^k \in \mathbb{R}$, find a function $f(x)$ composed of list of input operators (e.g., $\{+, -, \times, \div\}$) and arbitrary constants such that $Y^i \approx f(\mathbf{X}^i)$.

Linear Regression: $f(x)$ is a linear function $c_1x_1 + c_2x_2 + \cdots + c_nx_n$

Early work
- Connor,Taylor('77), Langley ('81)

Genetic Programing
- Koza('92), Schmidt, Lipson ('09,'10) - Eureqa

Mixed-integer nonlinear programming
- Cozad ('14), Horesh, Liberti, Avron ('16), Cozad, Sahinidis ('18)
- Austel, Dash, Gunluk, Horesh, Liberti, Nannicini, Schieber ('17)

Other methods for physics problems:
- Udrescu, Tegmark ('19,'20) – AI Feynman

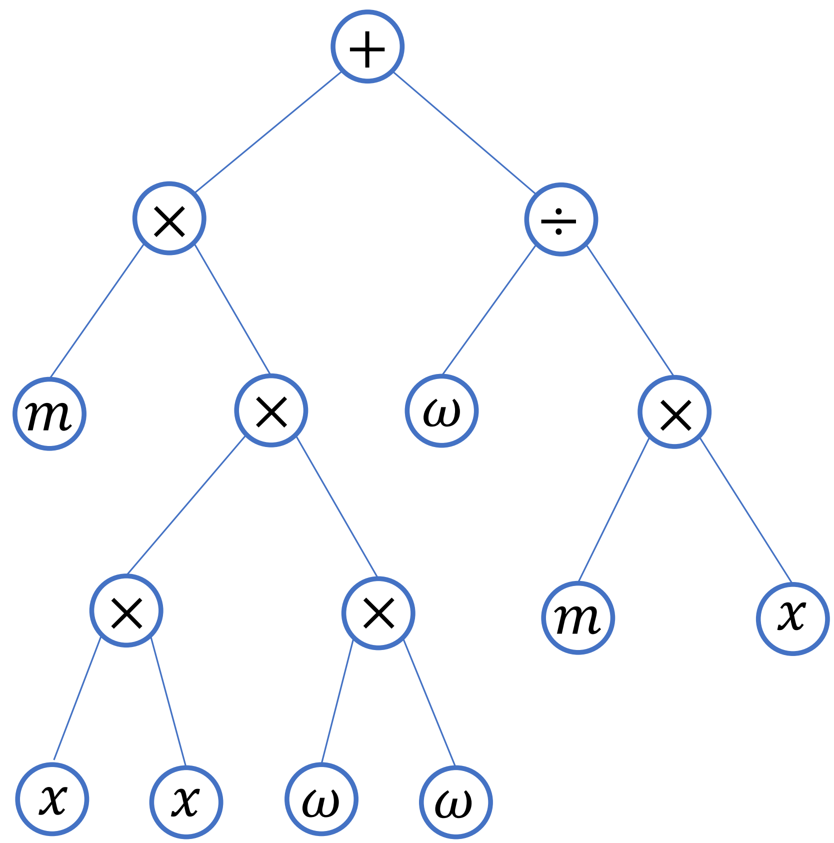
Expression tree

$$f(m,x,\omega) = mx^2\omega^2 + \frac{\omega}{mx}$$

Nodes are labeled by: binary and unary operators (such as $+, -, \times, \log$), variables, and constants

Edges link these entities in a way that is consistent with a prescribed grammar

Full expression tree



MINLP formulation

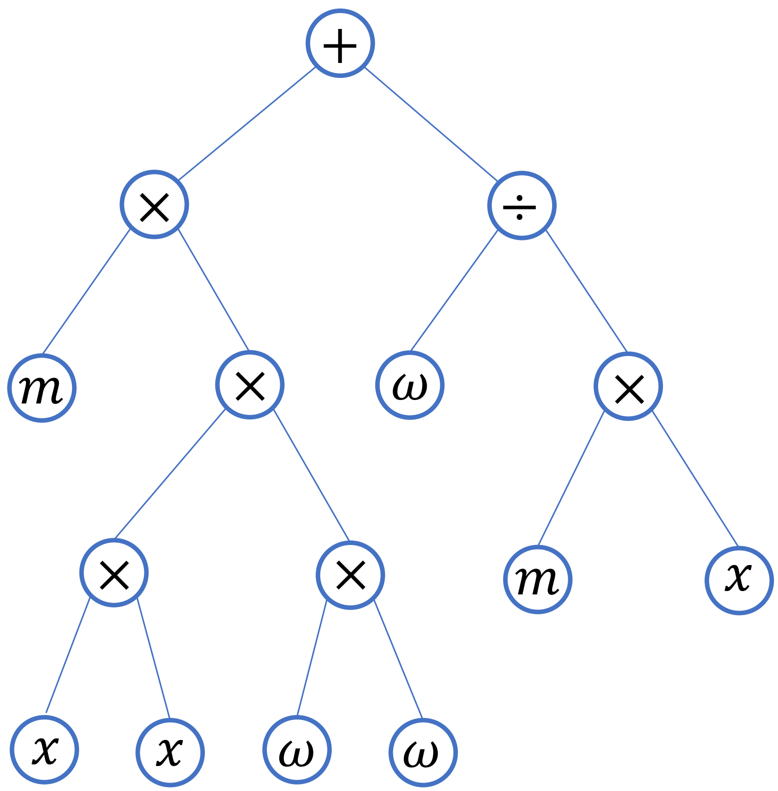
Binary variables choose locations of operators in non-leaf nodes of the expression tree and locations of variables and constants in leaf nodes

Continuous variables used for constant values and to calculate the value of the generated function and to compute error.

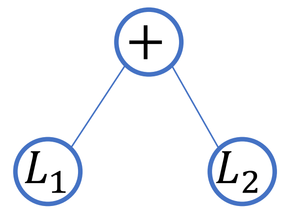
L-monomial tree representation

L-monomial = $hx_1^{a_1}x_2^{a_2}\cdots x_n^{a_n}$ where the powers can be positive and negative integers, and h is a constant

Full expression tree



L-monomial tree



$$L_1 = mx^2\omega^2$$

$$L_2 = \frac{\omega}{mx}$$

New MINLP formulation

We enumerate L-monomial expression trees, prune potentially redundant ones (e.g., $L_1/L_2 = L_3$) and solve an MINLP for each tree (using BARON)

The MINLP has variables p for independent feature powers, z for position of constants (whether it is 1 or a different number for an L-monomial), and h for constant values

$$
\begin{aligned}
\min \quad & \sum_{i\in I} (Y^{(i)} - f_{\mathbf{h},\mathbf{p},\mathbf{z},T}(\mathbf{X}^{(i)}))^2 \\
\text{s.t.} \quad & -\delta \le p_i \le \delta \quad \text{for } i = 1,\ldots,mn \\
& -\Omega z_i + (1 - z_i) \le h_i \le \Omega z_i + (1 - z_i) \quad \text{for } i = 1,\ldots,m \\
& \sum_{i=1}^{m} z_i \le k \\
& \mathbf{z} \in \{0,1\}^m, \quad \mathbf{p} \in \mathbb{Z}^{mn}
\end{aligned}
$$

Results

Label	Formula	AI-Descartes	AI Feynman	PySR	BMS
I.6.20a	$e^{-\theta^2/2}/\sqrt{2\pi}$	X	X	X	X
I.6.20	$e^{-\frac{\theta^2}{2\sigma^2}}/\sqrt{2\pi\sigma^2}$	X	X	X	X
I.6.20b	$e^{-\frac{(\theta-\theta_1)^2}{2\sigma^2}}/\sqrt{2\pi\sigma^2}$	X	X	X	X
I.8.14	$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2}$	X	X	X	X
I.9.18	$\frac{Gm_1m_2}{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$	X	X	X	X
I.10.7	$\frac{m_0}{\sqrt{1-v^2/c^2}}$	✓	X	X	X
I.11.19	$x_1y_1+x_2y_2+x_3y_3$	X	X	X	X
I.12.1	μN_n	✓	$✓^2$	✓	✓
I.12.2	$q_1q_2/(4\pi\varepsilon r^2)$	$✓^1$	X	$✓^1$	$✓^1$
I.12.4	$q_1/(4\pi\varepsilon r^2)$	$✓^1$	$✓^1$	$✓^1$	$✓^1$
I.12.5	q_2E_f	$✓^1$	$✓^2$	✓	✓
I.13.4	$\frac{1}{2}m(v^2+u^2+w^2)$	X	X	X	X

	AI-Descartes	AI Feynman	PySR	BMS
Number of (✓, $✓^1$, $✓^2$, $✓^3$, X)	(13, 32, 4, 0, 32)	(0, 25, 8, 0, 48)	(16, 21, 2, 0, 41)	(10, 17, 11, 1, 42)
Total $✓^*$	49/81	33/81	40/81	39/81
Accuracy	**60.49%**	40.74%	49.38%	48.15%

Supplementary Table 13. Results on 81/100 problems from the Feynman Database for Symbolic Regression (problems not containing trigonometric functions). The accuracy of the best method is marked with bold font.

Reasoning

1 - Constraints

Check if candidate formulas satisfy constraints, eg
- Monotonicity
- Conditions at the limit
- Nonnegativity

2 - Derivability

Derive a formula from *axioms* defining a background theory (use KeYmaera X)

3 - Reasoning measures

$$\beta^r_\infty = \max_{1 \leq i \leq m} \left\{ \frac{|f(\mathbf{X}^i) - f_\mathcal{B}(\mathbf{X}^i)|}{|f_\mathcal{B}(\mathbf{X}^i)|} \right\}$$

= Relative error between f (induced from data) and a derivable formula deducible from the axioms $f_\mathcal{B}$

Pointwise reasoning error: S = datapoints
Generalization reasoning error: S contains datapoints

Kepler's third law of planetary motion

$$p = \sqrt{\frac{4\pi^2 d^3}{G(m_1 + m_2)}}$$

Background Theory

K1. center of mass definition

K2. distance between bodies

K3. gravitational force

K4. centrifugal force

K5. force balance

K6. period definition

K7. non-negativity constraints

K1. $m_1 * d_1 = m_2 * d_2$

K2. $d = d_1 + d_2$

K3. $F_g = \frac{G m_1 m_2}{d^2}$

K4. $F_c = m_2 d_2 w^2$

K5. $F_g = F_c$

K6. $p = \frac{2\pi}{w}$

K7. $m_1 > 0,\ m_2 > 0,\ p > 0,\ d_1 > 0,\ d_2 > 0$.

Kepler’s third law of planetary motion

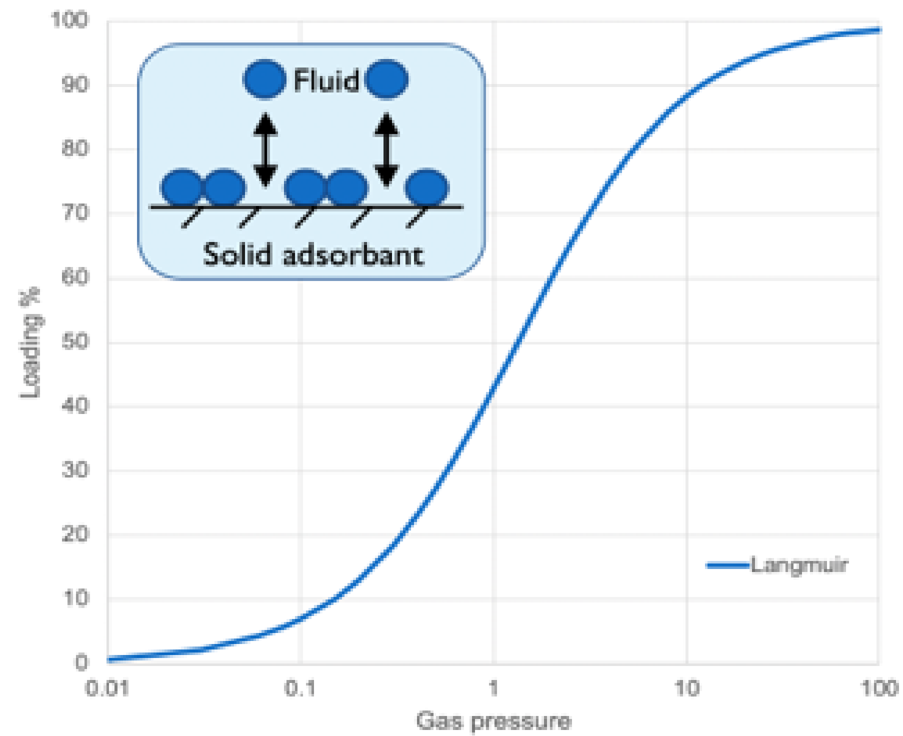
1	2	3	4	5	6	7	8	9	10
Dataset	Candidate formula $p =$	numerical error ε_2^r	ε_∞^r	point. reas. err. β_2^r	β_∞^r	gen. reas. error $\beta_{\infty,S}^r$	dependencies m_1	m_2	d
solar	$\sqrt{0.1319 \cdot d^3}$	.01291	.006412	.0146	.0052	.0052	0	0	1
	$\sqrt{0.1316 * (d^3 + d)}$	1.9348	1.7498	1.9385	1.7533	1.7559	0	0	0
	$(0.03765 d^3 + d^2)/(2+d)$	.3102	.2766	.3095	.2758	.2758	0	0	0
exoplanet	$\sqrt{0.1319 d^3 / m_1}$	.08446	.08192	.02310	.0052	.0052	0	0	1
	$\sqrt{m_1^2 m_2^3 / d + 0.1319 d^3 / m_1}$	.1988	.1636	.1320	.1097	> 550	0	0	0
	$\sqrt{(1 - .7362 m_1) d^3 / 2}$	1.2246	.4697	1.2418	.4686	.4686	0	0	1
binary stars	$1/(d^2 m_1^2) + 1/(d m_2^2) - m_1^3 m_2^2 + \sqrt{.4787 d^3 / m_2 + d^2 m_2^2}$	.002291	.001467	.0059	.0050	timeout	0	0	0
	$(\sqrt{d^3} + m_1^3 m_2 / \sqrt{d}) / \sqrt{m_1 + m_2}$	.003221	.003071	.0038	.0031	timeout	0	0	0
	$\sqrt{d^3 / (0.9967 m_1 + m_2)}$	.005815	.005337	.0014	.0008	.0020	1	1	1

Langmuir’s adsorption equation

This describes the amount of adsorbtion of gas molecules on a solid surface (“loading”) as a function of the pressure of the gas.

$$\frac{q}{q_{max}} = \frac{K_a \cdot p}{1 + K_a \cdot p}$$

- p = gas pressure
- q = loading on surface
- q_{max} = maximum loading
- K_a = adsorption strength



Langmuir's adsorption equation

Background theory

L1. Site balance: $S_0 = S + S_a$

L2. Adsorption rate model: $r_{ads} = k_{ads} \cdot p \cdot S$

L3. Desorption rate model: $r_{des} = k_{des} \cdot S_a$

L4. Equilibrium assumption: $r_{ads} = r_{des}$

L5. Mass balance on q: $q = S_a$.

$\mathcal{K}$ - CONSTRAINTS

C1. $f(0) = 0$

C2. $(\forall p > 0)\ (f(p) > 0)$

C3. $(\forall p > 0)\ (f'(p) \geq 0)$

C4. $0 < \lim_{p \to 0} f'(p) < \infty$

C5. $0 < \lim_{p \to \infty} f(p) < \infty$

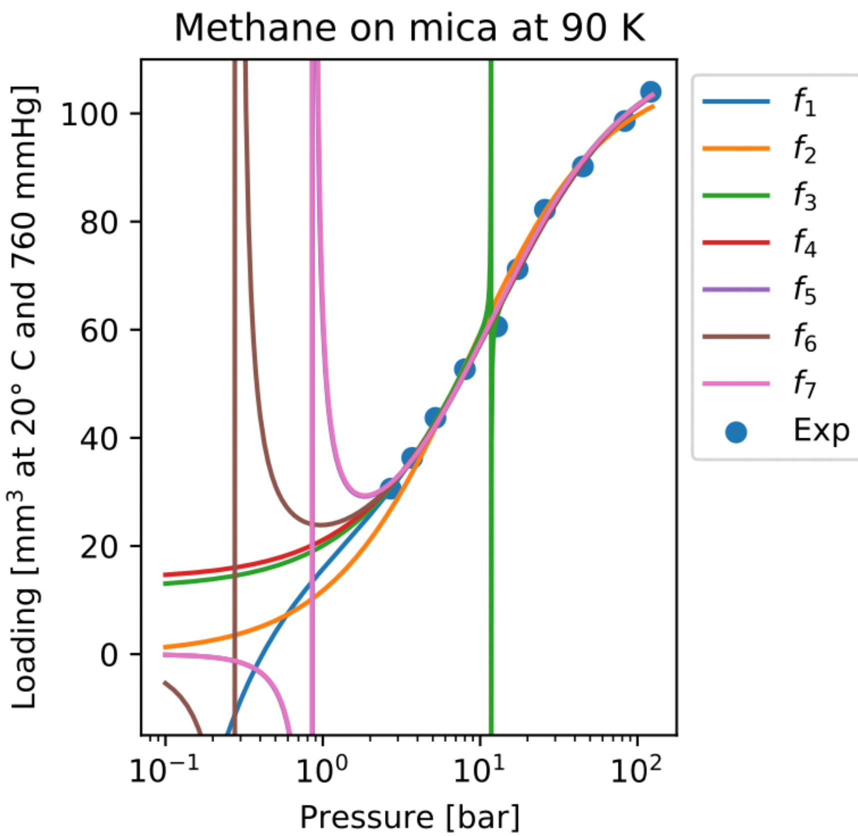
Work using reasoning to check for constraint satisfaction:
- Scott, Panju, Ganesh '21: LGML
- Ashok, Scott, Wetzel, Panju, Ganesh '21: LGGA

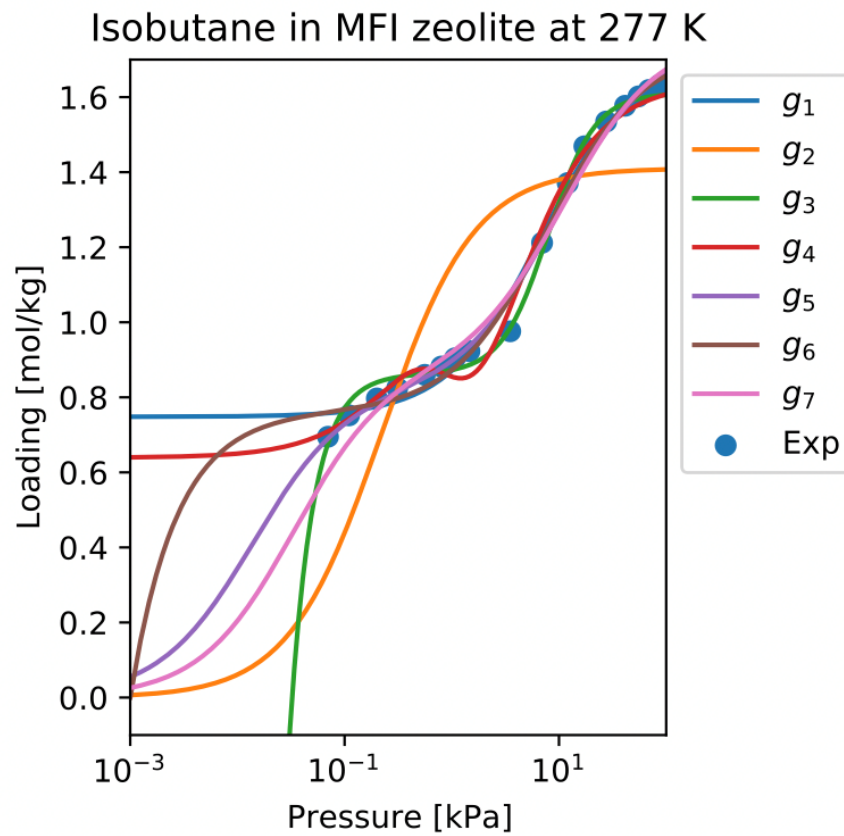
We allow background theory to contain variables not present in data.

Results

Data	Condition	Candidate formula $q =$	Numerical Error ε_2^r	ε_∞^r	provability	$\mathscr{K}$ constr.
Langmuir [25, Table IX]	2 const.	$f_1: (p^2+2p-1)/(.00888p^2+.118p)$	.06312	.04865	timeout	2/5
		$f_2: p/(.00927p+.0759)$ *	.1799	.1258	Yes	5/5
	4 const.	$f_3: (p^2-10.5p-15.)/(.00892p^2-1.23)$	.04432	.02951	timeout	2/5
		$f_4: (8.86p+13.9)/(.0787p+1)$	.06578	.04654	No	4/5
		$f_5: p^2/(.00895p^2+.0934p-.0860)$	.07589	.04959	No	2/5
	4 const. extra-point	$f_6: (p^2+p)/(.00890p^2+.106p-.0311)$	.06833	.04705	timeout	2/5
		$f_7: (112p^2-p)/(p^2+10.4p-9.66)$	.07708	.05324	timeout	3/5
Sun et al. [26, Table 1]	2 const.	$g_1: (p+3)/(.584p+4.01)$	.1625	.1007	No	4/5
		$g_2: p/(.709p+.157)$	.9680	.5120	Yes	5/5
	4 const.	$g_3: (.0298p^2+1)/(.0185p^2+1.16)-.000905/p^2$	.1053	.05383	timeout	2/5
		$g_4: 1/(p^2+1)+(2.53p-1)/(1.54p+2.77)$	.1300	.07247	timeout	3/5
	4 constants extra-point	$g_5: (1.74p^2+7.61p)/(p^2+9.29p+0.129)$	.1119	.0996	timeout	5/5
		$g_6: (.226p^2+.762p-7.62*10^{-4})/(.131p^2+p)$	.1540	.09348	timeout	2/5
		$g_7: (4.78p^2+26.6p)/(2.71p^2+30.4p+1.)$	.1239	.1364	timeout	5/5

Langmuir results





f_2 and g_2 are derivable with KeyMaera. g_5, g_7 satisfy the constraints and are derivable from the *two-site theory*, but we cannot derive them.

Polynomial optimization

Let $p(x), q_1(x), q_2(x), \ldots, q_m(x)$ be polynomials.

$$p(x) = q_1(x)^2 + q_2(x)^2 + \cdots q_m(x)^2 \Rightarrow p(x) \geq 0$$

Hilberts thm:

$$p(x) \text{ quadratic, and } p(x) \geq 0 \Rightarrow p(x) = q_1(x)^2 + q_2(x)^2 + \cdots q_m(x)^2$$

Artin's thm:

$$p(x) \geq 0 \Rightarrow q_0(x)^2 p(x) = q_1(x)^2 + q_2(x)^2 + \cdots q_m(x)^2$$

Polynomial optimization

Putinar's Positivestellensatz: Consider the basic (semi)algebraic sets

$$\mathcal{G} := \{\boldsymbol{x} \in \mathbb{R}^n : \; g_1(\boldsymbol{x}) \geq 0, \ldots, g_m(\boldsymbol{x}) \geq 0\}$$
$$\mathcal{H} := \{\boldsymbol{x} \in \mathbb{R}^n : \; h_1(\boldsymbol{x}) = 0, \ldots h_n(\boldsymbol{x}) = 0\}$$

where g_i, h_j are polynomials, and $\mathcal{G}$ satisfies the Archimedean property. Then

$$f(\boldsymbol{x}) \geq 0 \text{ for all } \boldsymbol{x} \in \mathcal{G} \cap \mathcal{H}$$

if and only if

$$f(x) = \alpha_0(x) + \sum_{i=1}^{m} (\alpha_i(\boldsymbol{x}))^2 g_i(\boldsymbol{x}) + \sum_{j=1}^{n} \beta_j(\boldsymbol{x}) h_j(\boldsymbol{x}).$$

Example: Kepler

Kepler's third law
$$p = \sqrt{\frac{4\pi^2(d_1 + d_2)^3}{G(m_1 + m_2)}},$$

p - rotational period, d_1, d_2 distances to common center of mass, m_1, m_2 masses, $G = 6.6743 \times 10^{-11} m^3 kg^{-1} s^{-2}$ universal gravitational constant

Axioms:

$$\begin{aligned} d_1 m_1 - d_2 m_2 &= 0 \\ (d_1 + d_2)^2 F_g - G m_1 m_2 &= 0 \\ F_c - m_2 d_2 w^2 &= 0 \\ F_c - F_g &= 0 \\ wp &= 1 \end{aligned}$$

Incorrect formula: $p^2 m_1 - 0.1319(d_1 + d_2)^3 = 0$

Solution of Kepler

$$\min \sum_{i=1}^{n} |q(\boldsymbol{x}_i)|,$$

q is solution polynomial, $\{\boldsymbol{x}_i\}_{i=1}^{4}$ is a set of observations

Search over the deg-5 polynomials q derivable using deg-6 certificates
MIP with 18958 continuous and 6 discrete variables, solves in 4 seconds
Solution:

$$m_1 m_2 G p^2 - m_1 d_1 d_2^2 - m_2 d_1^2 d_2 - 2 m_2 d_1 d_2^2 = 0,$$

Certificate:

$$
\begin{aligned}
-d_2^2p^2w^2, \\
-p^2, \\
d_1^2p^2 + 2d_1d_2p^2 + d_2^2p^2, \\
d_1^2p^2 + 2d_1d_2p^2 + d_2^2p^2, \\
m_1d_1d_2^2pw + m_2d_1^2d_2pw + 2m_2d_1d_2^2pw + m_1d_1d_2^2 + m_2d_1^2d_2 + 2m_2d_1d_2^2,
\end{aligned}
$$

Conclusion

Strengths:
- Few data points
- Real data
- Logical reasoning to distinguish the correct formula from a set of plausible formulas with similar error on the data

Limitations:
- Scalability
- Rely on correctness & completeness of background theory

Main challenges:
- Need more real-data datasets (with realistic amount/type of noise)
- Need more numerical datasets with associated background theory

Future directions:
- Consider restricted classes of axioms and derived formulas

References

1. C. Cornelio, T. Josephson, S. Dash, J. Goncalves, V. Austel, K. Clarkson, N. Megiddo, L. Horesh, Combining data and theory for derivable scientific discovery with AI-Descartes, Nature Communications **14** (2023), Article 1777. IBM blog post, Webpage
2. R. Cory-Wright, B. El Khadir, C. Cornelio, S. Dash, L. Horesh, AI Hilbert: From Data and Background Knowledge to Automated Scientific Discovery via Polynomial Optimization, Technical Report, IBM, 2023.