



Learning Equations from Data: Exhaustive Symbolic Regression



Seminarium astrofizyczne Komisji Astrofizyki PAU

Deaglan Bartlett

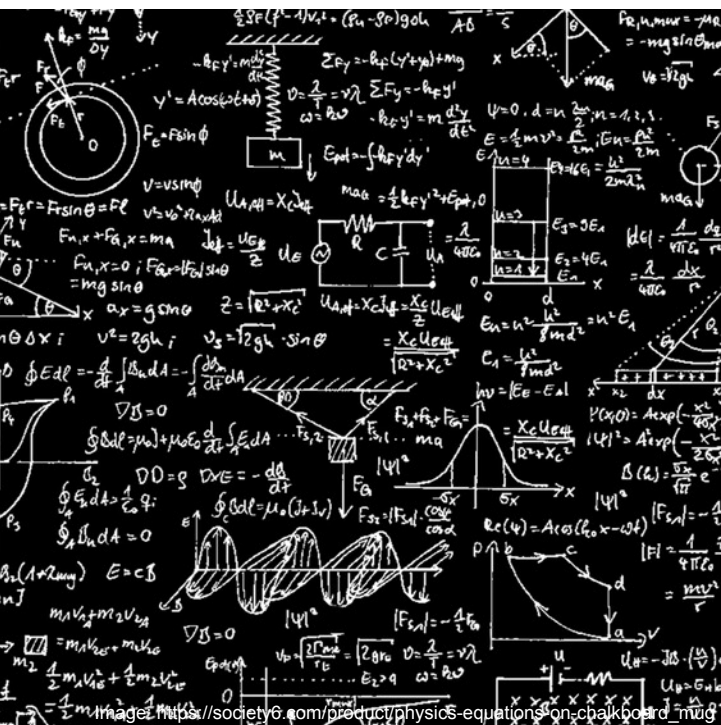
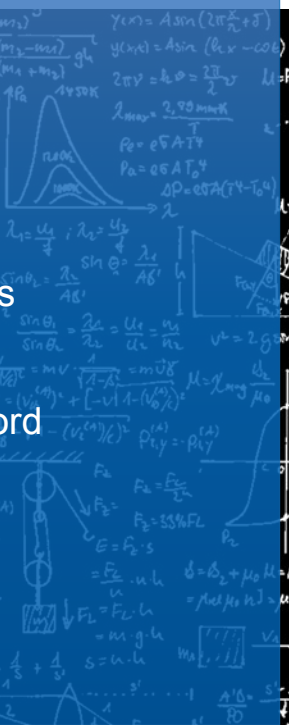
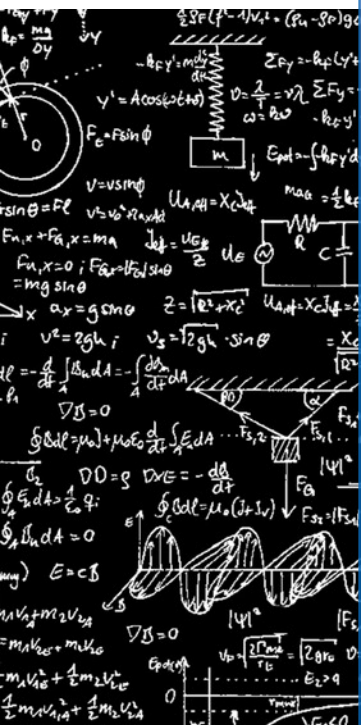
deaglan.bartlett@iap.fr

Institut d'Astrophysique de Paris
CNRS & Sorbonne Université

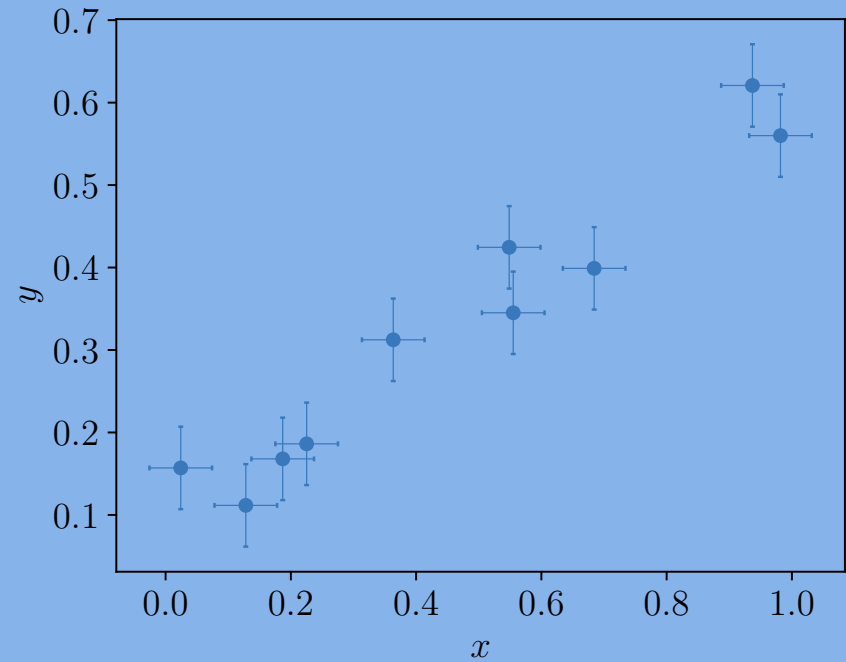
Astrophysics, University of Oxford

Harry Desmond & Pedro Ferreira

10 May 2023



Learning Equations from Data



Symbolic Regression - Overview

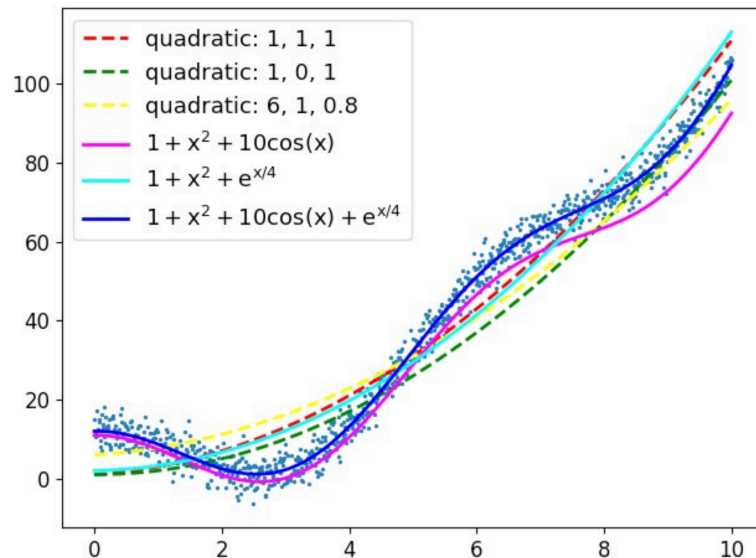
Find the best function rather than just the best parameters of a given function

Numerical Regression:

$$y = 6 + 1x + 0.8x^2$$

Symbolic Regression:

$$y = 1 + x^2 + 10 \cos(x)$$



Why Symbolic Regression?

Goals:

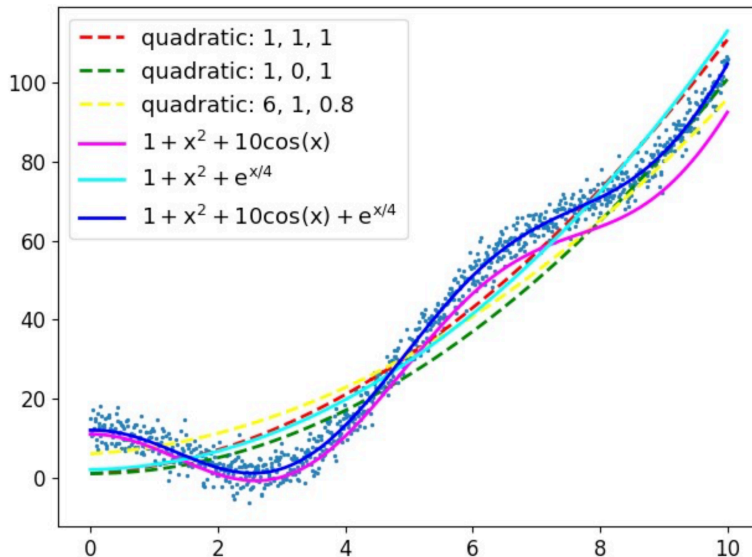
- Discover “physical laws” - necessarily analytic
- Guide theory - discover what limits a model should have
- Effective prescriptions of complex, non-linear systems

Advantages:

- More general - reduce confirmation bias
- Prevent overfitting
- Highly interpretable

Difficulties:

- Larger search space - convergence harder
- Optimisation methods of numerical regression not applicable



Traditional Symbolic Regression

The Task

Given some dictionary of functions, how do I combine these to produce an **accurate** and **simple** function which describes my data?

Lots of tools on the market

Genetic Algorithms

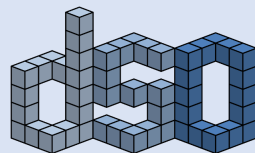


Cranmer et al. 2020 (arXiv:2006.11287)

**EVOLVED
ANALYTICS**

EUREQA

Neural Network Based



deep
symbolic
optimization

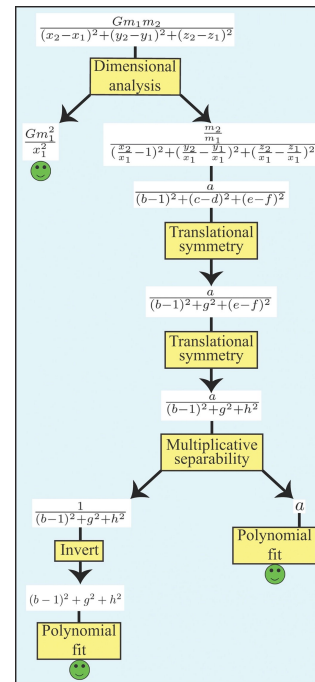
Petersen et al. 2019 (arXiv:1912.04871)

Simulated Annealing



TuringBot

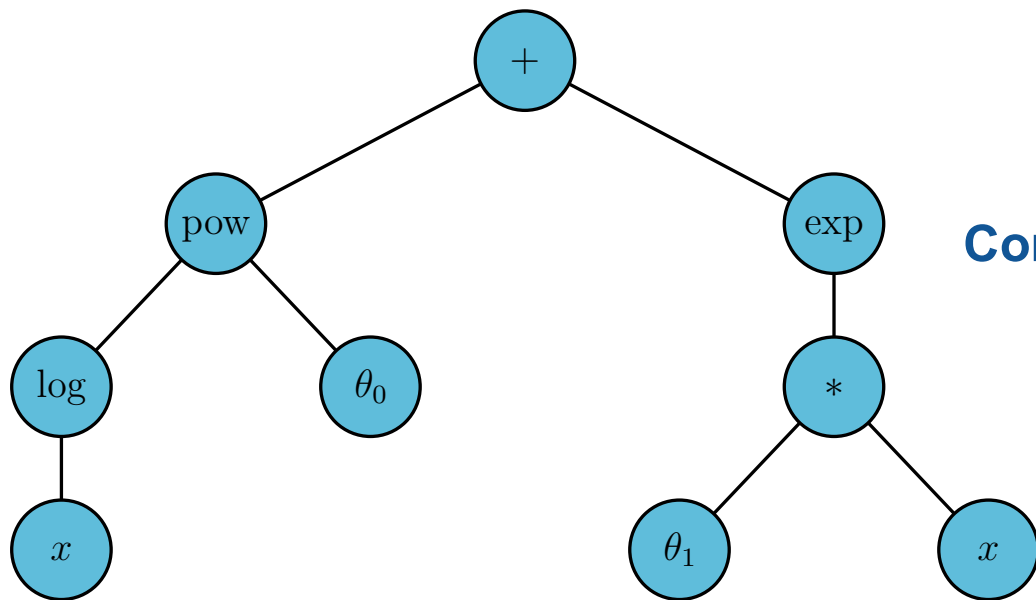
Physics Inspired (AlFeynman)



Udrescu & Tegmark 2019 (arXiv:1905.11481)
Udrescu et al. 2020 (arXiv:2006.10782)

Combining Functions - Represent Functions as Trees or Lists

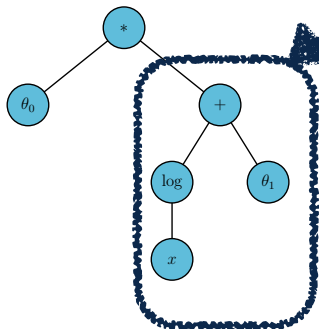
$$(\log(x))^{\theta_0} + \exp(\theta_1 x) = [+ , \text{pow} , \log , x , \theta_0 , \text{exp} , * , \theta_1 , x]$$



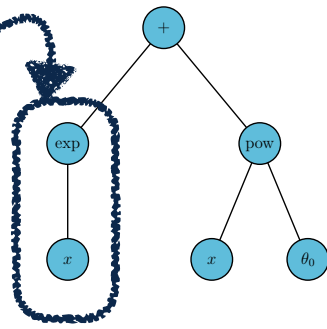
Complexity = Number of Nodes

The Genetic Algorithm

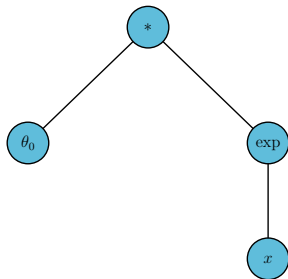
$$\theta_0 (\log(x) + \theta_1) = [* , \theta_0 , + , \log , x , \theta_1]$$



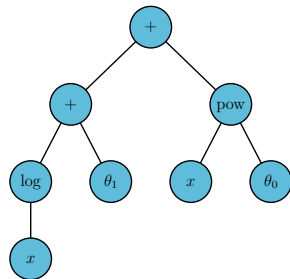
$$\exp(x) + x^{\theta_0} = [+ , \exp , x , \text{pow} , x , \theta_0]$$



$$\log(x) + \theta_1 + x^{\theta_0} = [* , \theta_0 , \exp , x]$$



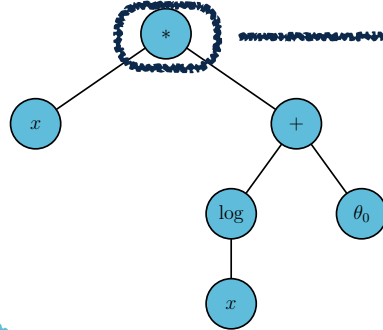
$$\log(x) + \theta_1 + x^{\theta_0} = [+ , + , \log , x , \theta_1 , \text{pow} , x , \theta_0]$$



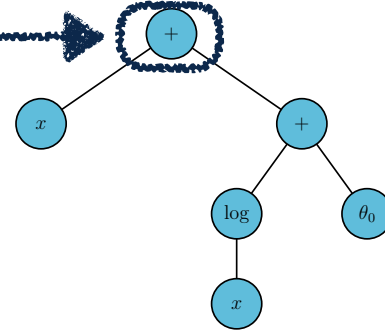
Breeding

Mutations

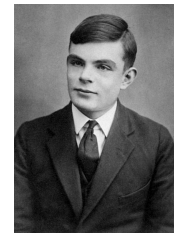
$$x (\log(x) + \theta_0) = [* , x , + , \log , x , \theta_0]$$



$$x + \log(x) + \theta_0 = [+ , x , + , \log , x , \theta_0]$$



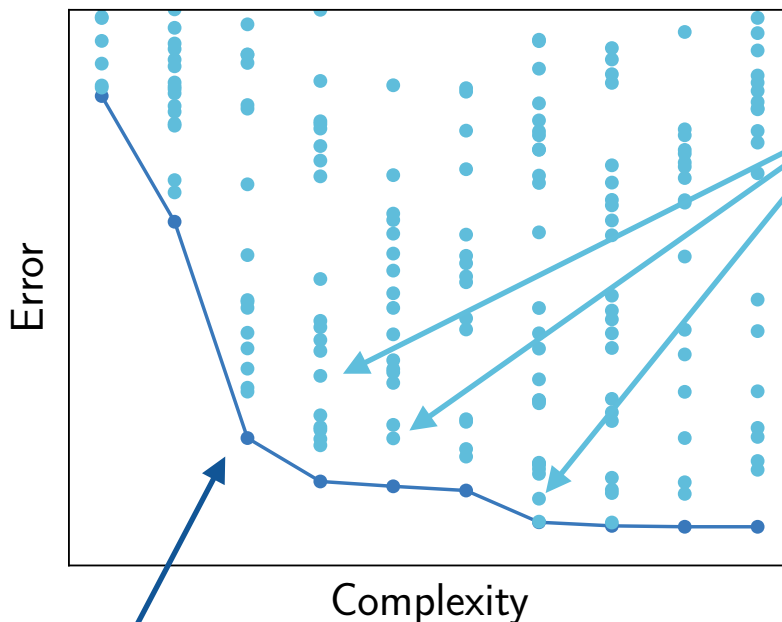
- Start with a population of functions
- Allow these to mutate and breed
- Optimise parameters
- Preferentially keep functions which fit data well
- Functions should evolve to fit data well



Alan Turing

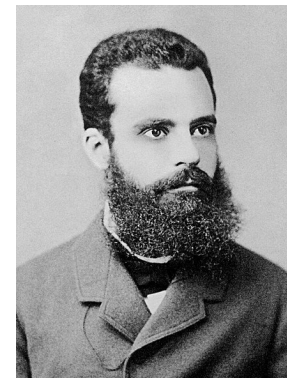
The Pareto Front - Balance of Simplicity and Accuracy

- **Pareto front:** most accurate solution at a given complexity
- Problem: Can make function arbitrarily complex to get zero error
- Which function on the Pareto front is the “best”?



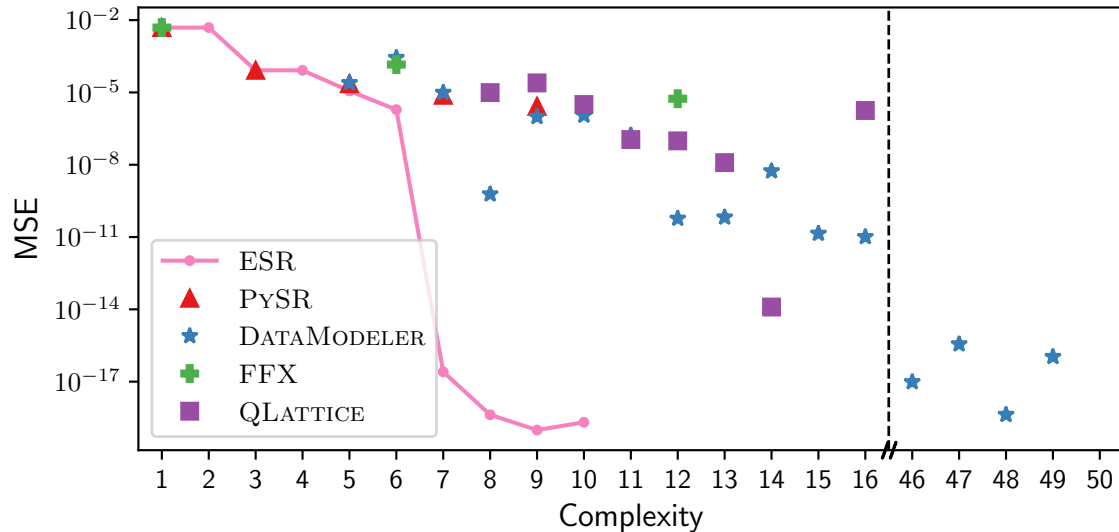
Feasible Solutions

Pareto front



Vilfredo Pareto

Traditional methods do not always find the true optimum



Task:

Univariate function with 100,000
noiseless data points

Best equation found:

$$y = \theta_1 \theta_0^{x^2}$$

$$\theta_0 \approx 1/\sqrt{e}, \quad \theta_1 \approx 1/\sqrt{2\pi}$$

**None of these stochastic
methods found this!**

Two Big Problems

1. Traditional methods do not necessarily find the best function
2. How do you balance accuracy with simplicity?

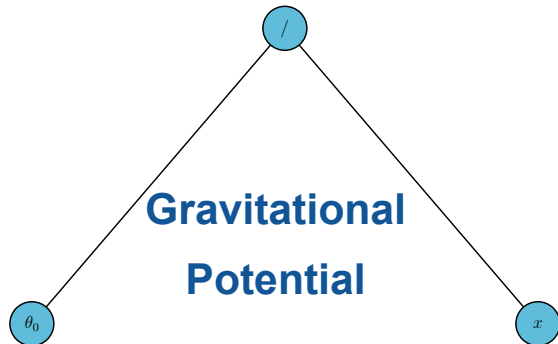
Exhaustive Symbolic Regression

Bartlett, Desmond & Ferreira 2022
(arXiv:2211.11461)

1. Finding the true optimum

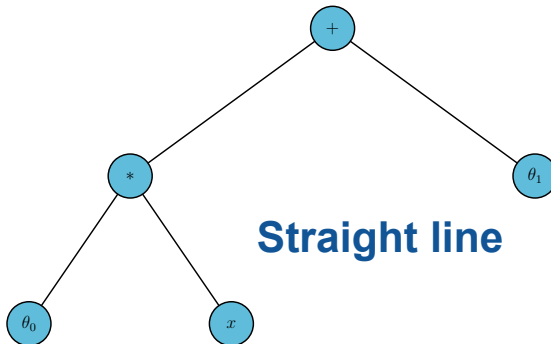
Motivation: Physics often needs simple equations

$$\frac{\theta_0}{x} = [/, \theta_0, x]$$



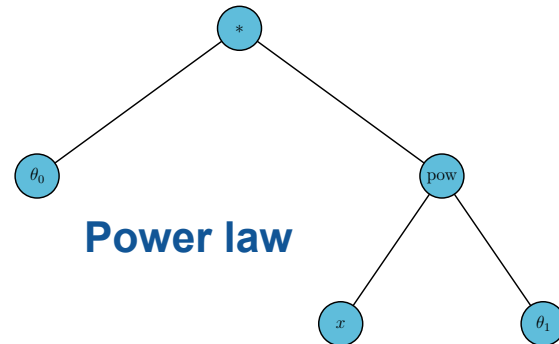
Gravitational Potential

$$\theta_0 x + \theta_1 = [+ , * , \theta_0 , x , \theta_1]$$



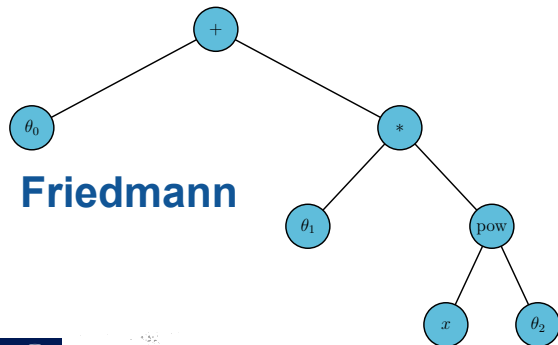
Straight line

$$\theta_0 x^{\theta_1} = [* , \theta_0 , \text{pow} , x , \theta_1]$$



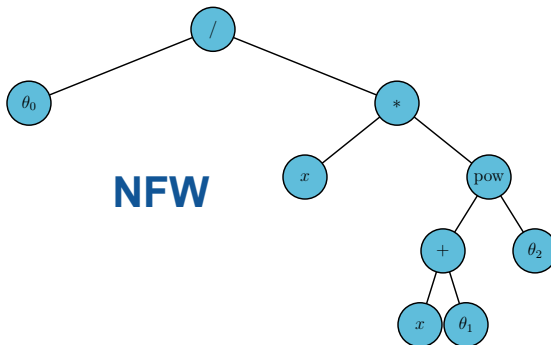
Power law

$$\theta_0 + \theta_1 x^{\theta_2} = [+ , \theta_0 , * , \theta_1 , \text{pow} , x , \theta_2]$$



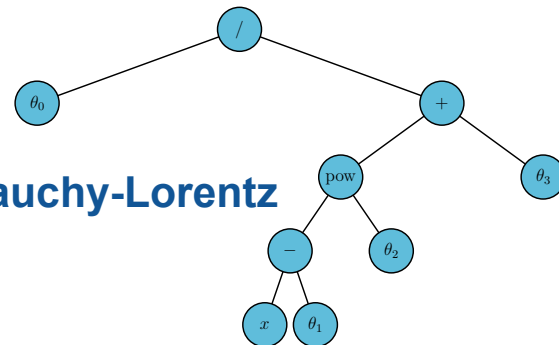
Friedmann

$$\frac{\theta_0}{x(x+\theta_1)^{\theta_2}} = [/, \theta_0 , * , x , \text{pow} , + , x , \theta_1 , \theta_2]$$



NFW

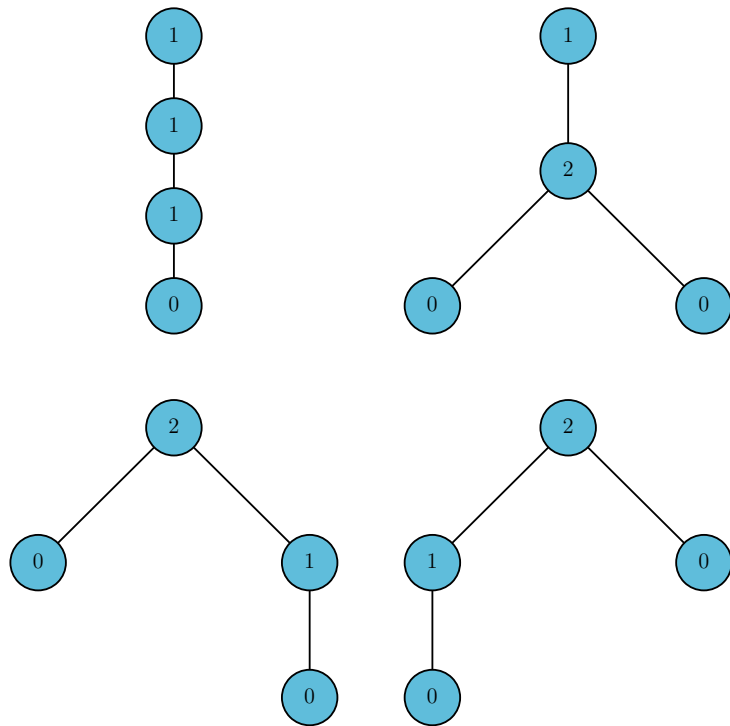
$$\frac{\theta_0}{(x-\theta_1)^{\theta_2+\theta_3}} = [/, \theta_0 , + , \text{pow} , - , x , \theta_1 , \theta_2 , \theta_3]$$



Cauchy-Lorentz

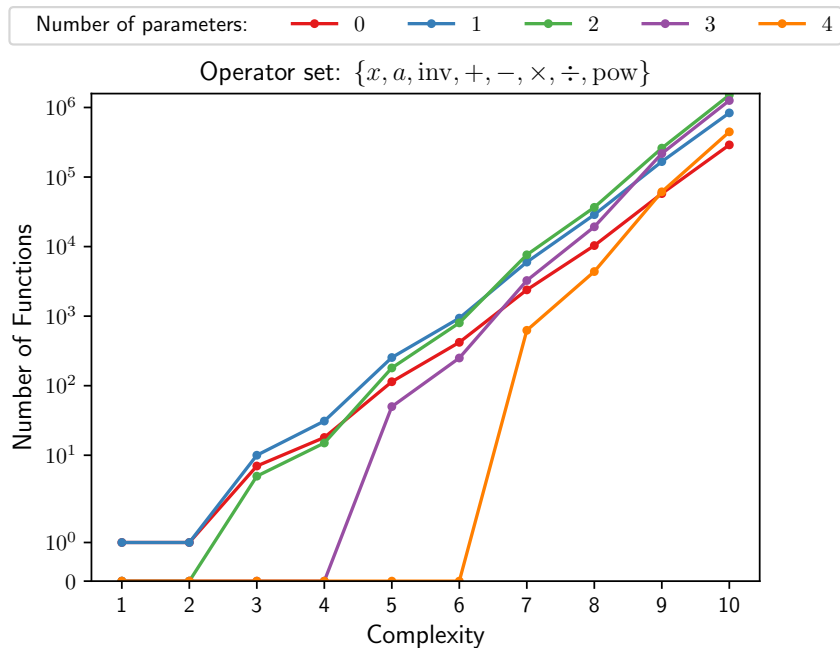
Find all valid trees

- First find all valid *shapes* of trees
- Equivalent to all valid lists of 0, 1 and 2
- Not all lists can form a tree - only keep ones that do
- Decorate the trees



All valid trees at complexity 4: Only 4, but $3^4 = 81$ lists

Problem: Number of functions grows exponentially with complexity



168 million

Naïve estimate: $\sum_{j=1}^n j^k = H_n^{(-k)}$

5.2 million

Total number of decorated trees

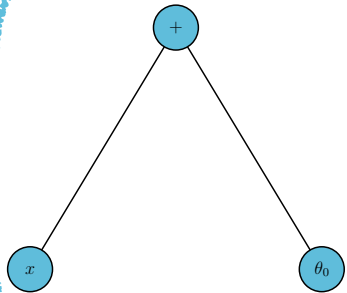
Even if we can't fix the exponential growth:

$$n_{\text{fun}} \sim A \times B^{\text{Complexity}},$$

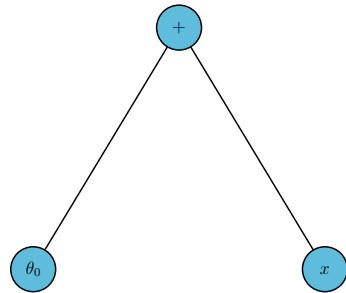
Can we reduce A to make this feasible?

Remove duplicate equations

$x + \theta_0 = [+ , x , \theta_0]$

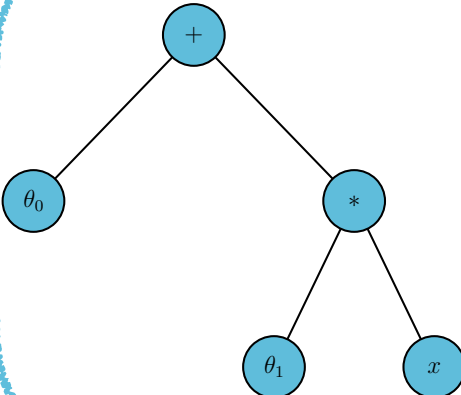


$\theta_0 + x = [+ , \theta_0 , x]$

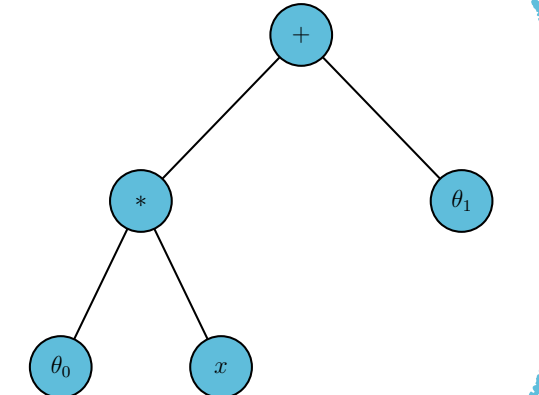


Tree reordering

$\theta_0 + \theta_1 x = [+ , \theta_0 , * , \theta_1 , x]$

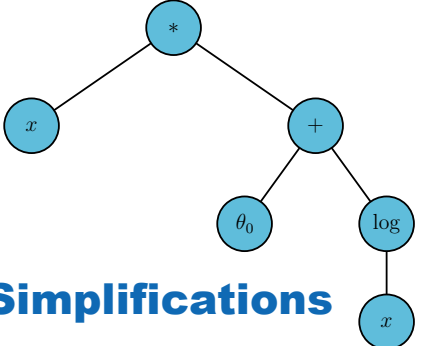


$\theta_0 x + \theta_1 = [+ , * , \theta_0 , x , \theta_1]$

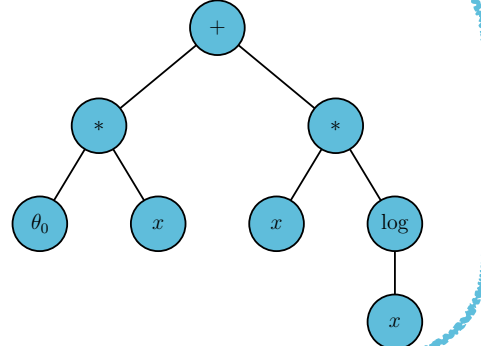


Parameter permutations

$x (\theta_0 + \log(x)) = [* , x , + , \theta_0 , \log , x]$



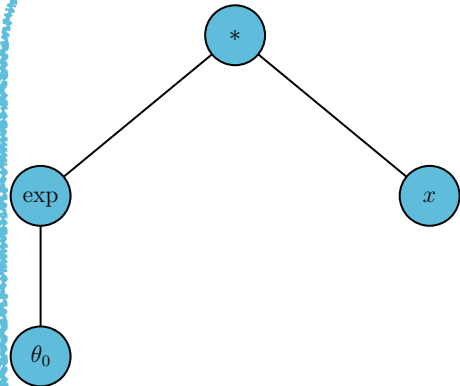
$\theta_0 x + x \log(x) = [+ , * , \theta_0 , x , * , x , \log , x]$



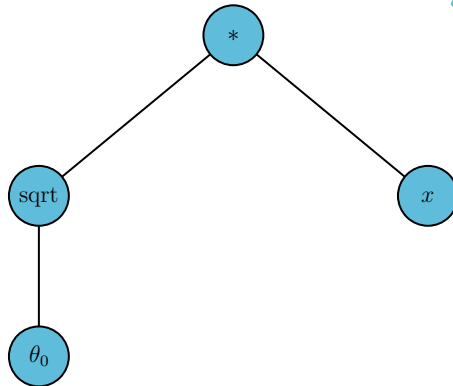
Simplifications

Remove duplicate equations

$$\exp(\theta_0)x = [* , \exp, \theta_0, x]$$



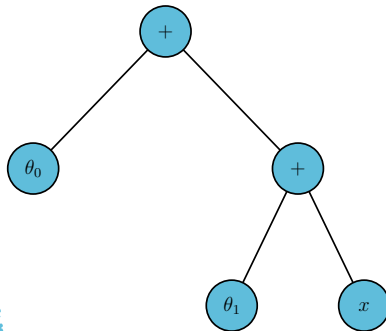
$$\sqrt{\theta_0}x = [* , \text{sqrt}, \theta_0, x]$$



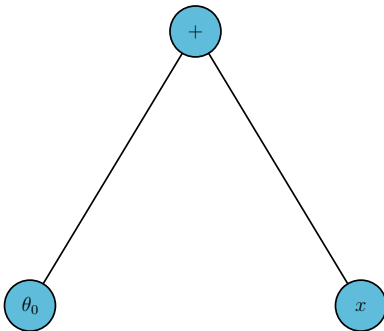
Reparametrisation invariance

Parameter combinations

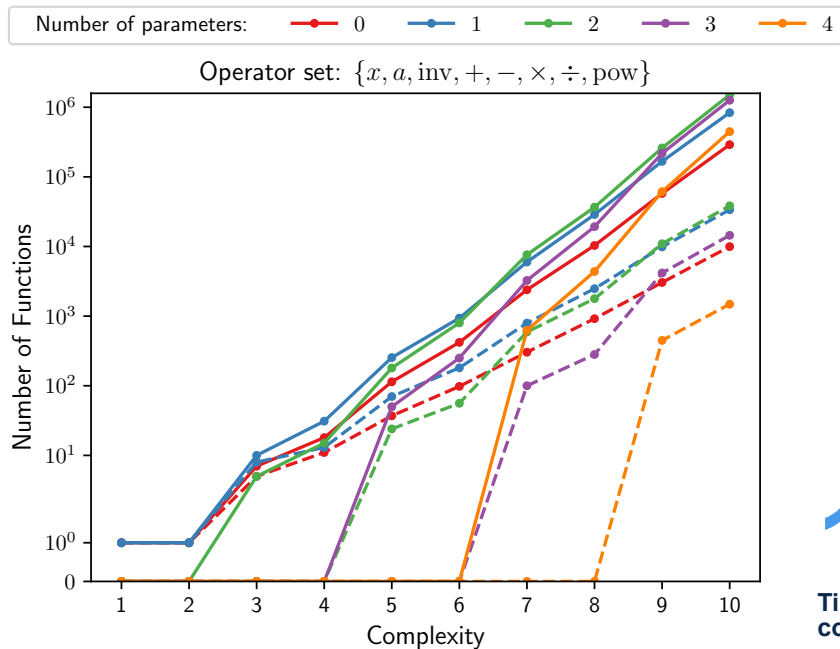
$$\theta_0 + \theta_1 + x = [+ , \theta_0, + , \theta_1, x]$$



$$\theta_0 + x = [+ , \theta_0, x]$$



Simplifications mean that an exhaustive search is feasible



168 million

Naïve estimate: $\sum_{j=1}^n j^k = H_n^{(-k)}$

5.2 million

Total number of decorated trees

134,234

Number of unique equations

1400

Times fewer equations to consider than our naïve guess

119,861

Number of equations containing at least one parameter

Exhaustive Symbolic Regression

Bartlett, Desmond & Ferreira 2022
(arXiv:2211.11461)

2. Model selection

A recap of Bayesian model comparison

$$P(\theta|D) = \frac{\mathcal{L}(D|\theta) \varpi(\theta)}{Z(D)}$$

Bayes' theorem

$$Z(D|\mathcal{M}) \longrightarrow Z(D) = \int \mathcal{L}(D|\theta) \varpi(\theta) d\theta$$

“Evidence”

$$P(\mathcal{M}|D) = \frac{Z(D|\mathcal{M}) \mathcal{P}(\mathcal{M})}{\mathcal{Z}(D)}$$

Bayes' theorem

$$\frac{P(\mathcal{M}_1|D)}{P(\mathcal{M}_2|D)} = \frac{\mathcal{P}(\mathcal{M}_1)}{\mathcal{P}(\mathcal{M}_2)} \frac{Z(D|\mathcal{M}_1)}{Z(D|\mathcal{M}_2)}$$

Model comparison

Problem with Bayesian evidence in symbolic regression

$$\log K = \log Z(D | \mathcal{M}_1) - \log Z(D | \mathcal{M}_2)$$

- Neglects prior on model
- Consider two models:

$$y = \theta \times f_1(x), \quad y = \theta \times f_2(x)$$

- If $f_1(x)$ and $f_2(x)$ are indistinguishable in the range of our data, they will have the same evidence
- But $f_1(x)$ may be more “complex” than $f_2(x)$
- The evidence won’t prefer $f_2(x)$, although we definitely would

Minimum Description Length

Why do we want to fit a function anyway?

- Compress the data
- If I have a perfect function, I don't need to tell someone the y values, but just the function

How much information do I need to transmit?

$$\text{Description length} \quad L(D) = \text{Hypothesis} \quad L(H) + \text{Residuals} \quad L(D | H)$$

The best function compresses the data the most $\rightarrow$ Minimise $L(D)$

Description length of a function

$$L(D) = L(D | H) + L(H)$$

$$L(D) = -\log(\mathcal{L}(\hat{\theta})) + k \log(n) + \sum_i^p \log(|\hat{\theta}_i|/\Delta_i) + p \log(2) + \sum_j \log(c_j)$$

Shannon-Fano coding

k nodes with n basis functions
needs $\log(n^k)$ nats

Send p parameters,
each with precision Δ_i

Integer constants

The Description Length of a Function

$$L(D) = -\log(\mathcal{L}(\hat{\theta})) + k \log(n) - \frac{p}{2} \log(3) + \sum_i^p \left(\frac{1}{2} \log(\mathbf{I}_{ii}) + \log(|\hat{\theta}_i|) \right)$$

This is the thing to optimise

Combined accuracy and simplicity into a single number

Favours functions which

- Fit the data accurately
- Contain only a few parameters
- Do not contain many operators
- Do not need finely tuned parameters

Two Solutions

1. By trying all functions, you are guaranteed to find the best
2. Minimise description length - naturally balances accuracy with simplicity

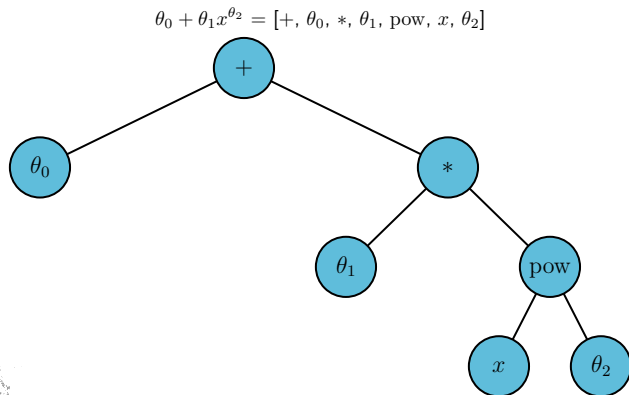
The Fundamental Law of Cosmic Expansion

Bartlett, Desmond & Ferreira 2022
(arXiv:2211.11461)

Bartlett, Desmond & Ferreira 2023
(arXiv:2304.06333)

Friedmann is a simple (complexity = 7) function - can we discover it?

- Can measure Hubble rate (or integrals of this) from standard clocks or candles
- Provides evidence for accelerated expansion of the Universe (2011 Nobel Prize)
- Usually fit terms of the Friedmann equation - assumes a theory of gravity and constituents of Universe

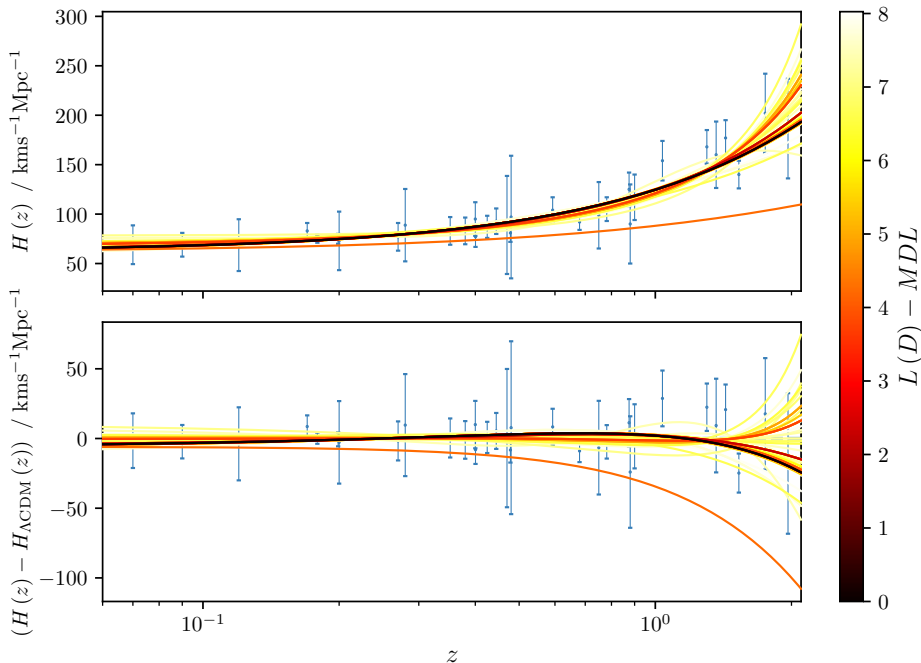


$$y(x \equiv 1 + z) \equiv H^2(z) = \theta_0 + \theta_1 x^{\theta_2}$$

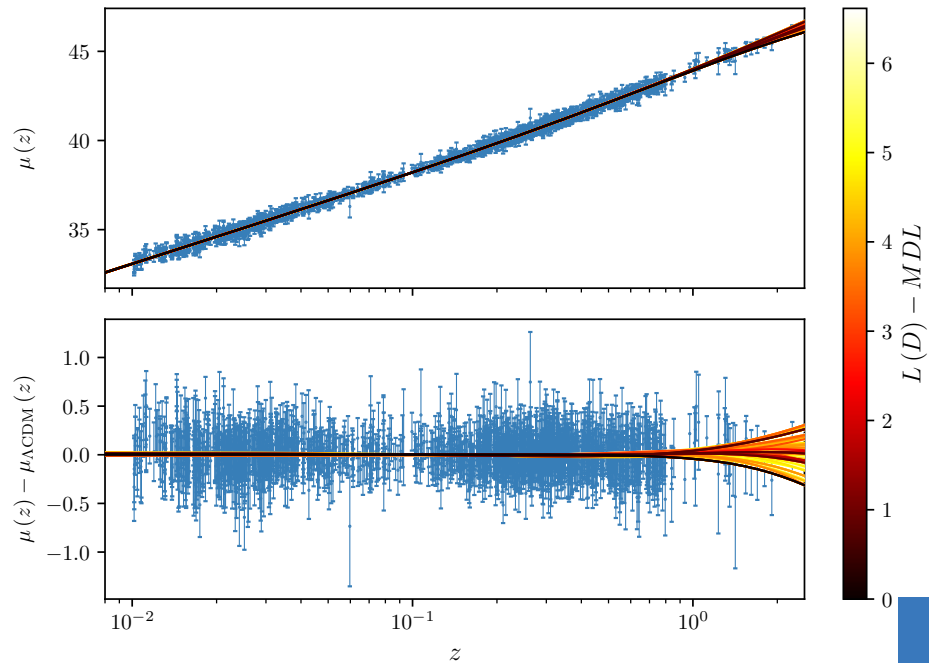
Can we derive the law of the Universe's expansion without having to assume General Relativity?

Many functions can fit the data, but they diverge outside this range

Cosmic Chronometers



Pantheon+ Supernovae



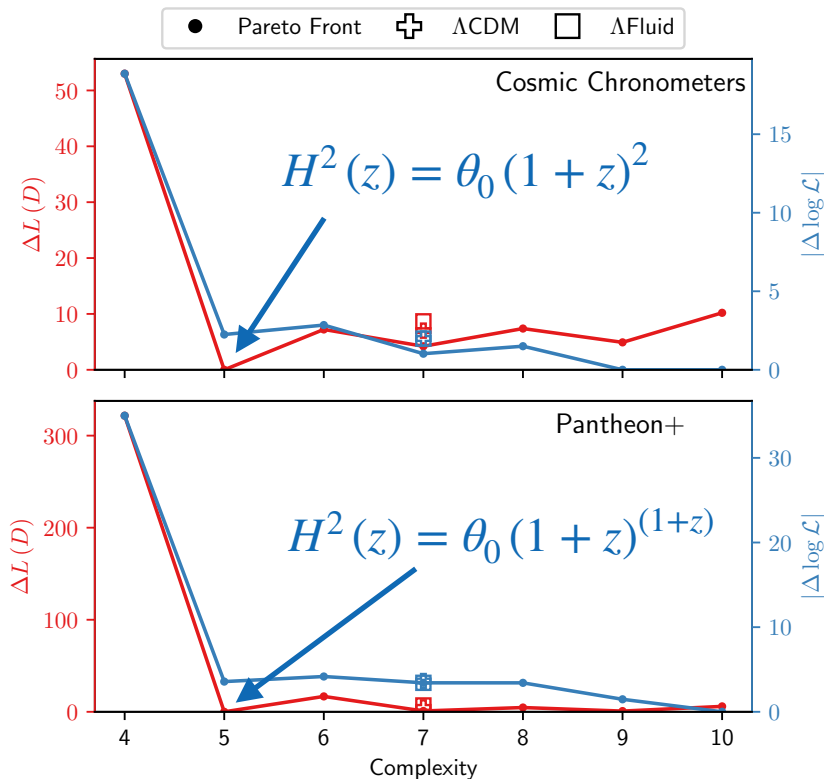
Jimenez & Loeb 2001 (arXiv:astro-ph/0106145), Moresco 2015 (arXiv:1503.01116), Moresco et al. 2016 (arXiv:1601.01701), Ratsimbazafy et al. 2017 (arXiv:1702.00418), Stern et al. 2010 (arXiv:0907.3149), Simon et al. 2004 (arXiv:astro-ph/0412269), Borghi et al. 2021 (arXiv:2110.04304), Zhang et al. 2012 (arXiv:1207.4541), Moresco et al. 2012 (arXiv:1201.3609), Moresco et al. 2022 (arXiv:2201.07241), Scolnic et al. 2021 (arXiv:2112.03863), Riess et al. 2022 (arXiv:2112.04510)

EXHAUSTIVE SYMBOLIC REGRESSION

DEAGLAN BARTLETT 29

Basis: $\{x \equiv 1 + z, \theta, \text{inv}, +, -, \times, \div, \text{pow}\}$

~40 functions fit the data more economically than Λ CDM



- Λ CDM in top 40 ($\theta_0 + \theta_1 x^3$)
- Λ Fluid in top 100 ($\theta_0 + \theta_1 x^{\theta_2}$)

Why?

$$y_{\Lambda\text{CDM}} = H_0^2 (1 + 3\Omega_{m0}z + 3\Omega_{m0}^2z^2 + \Omega_{m0}^3z^3)$$

$$\theta_0 x^x = \theta_0 \left(1 + z + z^2 + \frac{1}{2}z^3 + \mathcal{O}(z^4) \right)$$

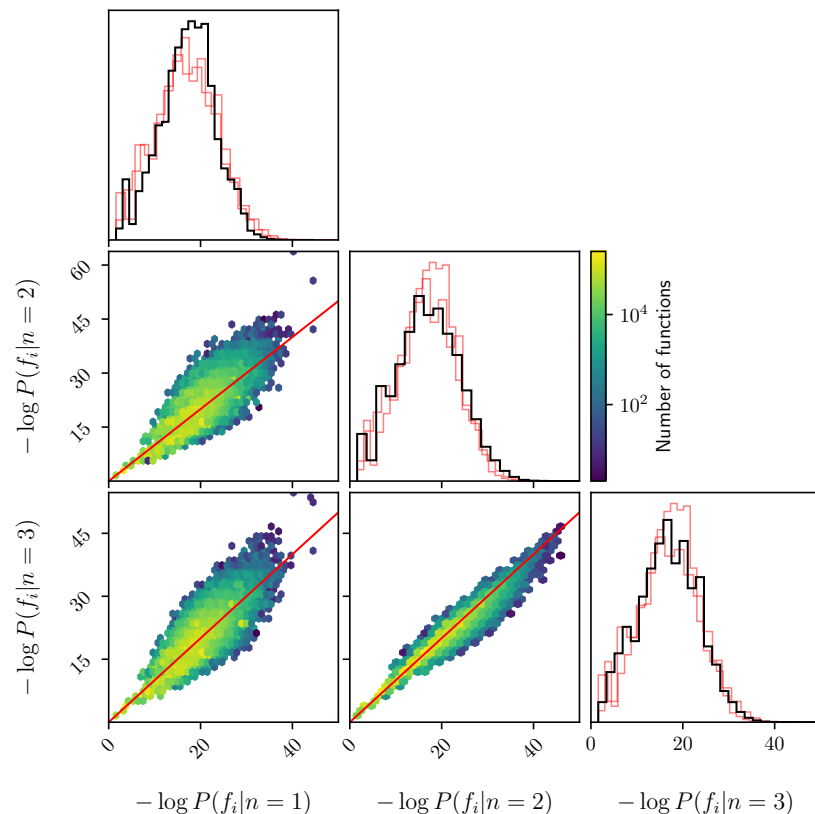
$$\Omega_{m0} \approx \frac{1}{3} \text{ so agree up to and including second order}$$

Benefit of ESR:

- Started with 5.2 million equations
- Λ CDM not on Pareto front - most standard SR packages would not have discovered this

Priors on functions

- Term $k \log n$ in MDL does not care about which operators being used
- Problem: Same prior probability to equations $\sin(f(x)) + \sin(g(x))$ and $\sin(\sin(f(x) + g(x)))$
- We would naturally prefer the former
- **Idea:** Weight functions based on known “good” equations based on how operators are combined
- Do this through a language model



Prior on functions removes the physically unreasonable cases

Method	Rank 1	Rank 2		Rank 3		Rank 4	
	Function	Function	Value	Function	Value	Function	Value
Likelihood	$\left \theta_1 + \frac{1}{\theta_0 - x}\right \theta_2 ^x$	$ \theta_0 - x\theta_1 ^{\theta_2 - x}$	1.48	$-\theta_2 + \theta_0 (\theta_1 - x) ^x$	3.15	$\frac{ \theta_1 + x\theta_2 ^x}{\theta_0}$	3.18
Score	$\theta_0 x$	$\theta_0 \theta_1 ^{-x}$	11.54	$ \theta_0 - x\theta_1 ^{\theta_2 - x}$	14.32	$\left \theta_1 + \frac{1}{\theta_0 - x}\right \theta_2 ^x$	14.59
MDL	$\theta_0 x^x$	$ \theta_0 ^x \theta_1^{\theta_1}$	0.78	$\theta_0 \theta_1 ^{-x}$	0.83	$\theta_0 x^x \theta_1^{\theta_1}$	0.91
MDL+LM	$\frac{x(2x + \frac{1}{x})}{\theta_0}$	$\frac{2x^2 + \frac{1}{x}}{\theta_0}$	1.40	$\frac{x^2 + 2x}{\theta_0}$	1.63	$\theta_0 + \theta_1 x^3$	1.72
MDL+FBF+LM	$ \theta_0 ^{\frac{1}{\theta_1 + 2x}}$	$ \theta_0 \theta_1 ^x$	0.16	$\theta_0 \theta_1 ^x$	1.77	$\frac{\theta_1 - x^2}{\theta_0}$	2.19
Bayes+FBF+LM	$ \theta_0 ^{\frac{1}{\theta_1 + 2x}}$	$ \theta_0 \theta_1 ^x$	0.16	$ \theta_0 ^{-\theta_1 + x}$	1.03	$\frac{x^3 + \frac{1}{\theta_0}}{\theta_1}$	1.95

The Fundamental Law of Cosmic Expansion

1. ~40 functions fit the data more economically than the Friedmann equation
2. Low-redshift data do not necessarily prefer a Λ CDM expansion history
3. Many traditional SR algorithms would not have found Λ CDM

Theory of Gravity

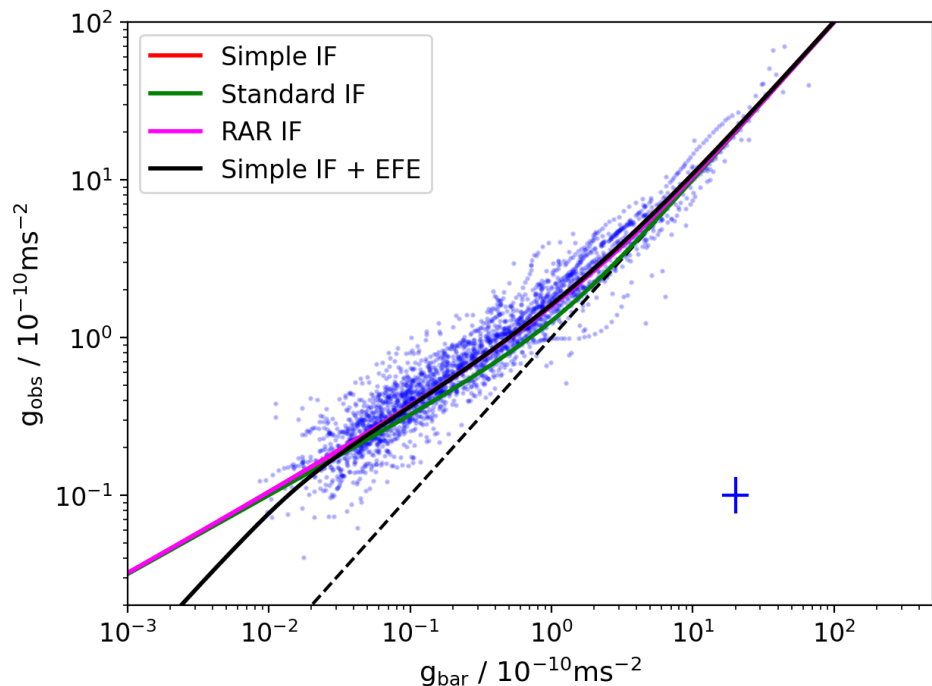
Desmond, **Bartlett** & Ferreira 2022
(arXiv:2301.04368)

The Radial Acceleration Relation (RAR)

- Relates centripetal acceleration due to baryons (g_{bar}) to total dynamical acceleration (g_{obs})
- Regular and low scatter
- Deviates from $g_{\text{obs}} = g_{\text{bar}}$ at low g_{bar}
 - **Dark matter** - fortuitous?
 - **Modification to gravity**
- Various Interpolating Functions (IFs) have been suggested. All have two limits:

Newtonian: $g_{\text{obs}} = g_{\text{bar}}, \quad g_{\text{bar}} \rightarrow \infty$

MOND: $g_{\text{obs}} = \sqrt{g_{\text{bar}} g_0}, \quad g_{\text{bar}} \rightarrow 0$



e.g. Milgrom 1983 (ApJ 270 365, ApJ 270 371, ApJ 270 384), Lelli et al. 2017 (arXiv:1610.08981), Skordis & Zlošnik 2021 (arXiv:2007.00082), Famaey & McGaugh 2012 (arXiv:1112.3960), Banik & Zhao 2022 (2110.06936)

Does the RAR support MOND?

If the RAR supports MOND, then any function

$g_{\text{obs}}(g_{\text{bar}})$ must have a Newtonian and a MOND limit

$$s_- \equiv \lim_{g_{\text{bar}} \rightarrow 0} \frac{d \log(g_{\text{obs}})}{d \log(g_{\text{bar}})}, \quad s_+ \equiv \lim_{g_{\text{bar}} \rightarrow \infty} \frac{d \log(g_{\text{obs}})}{d \log(g_{\text{bar}})},$$

A MOND function needs $s_- = \frac{1}{2}$ and $s_+ = 1$

Typical IFs:

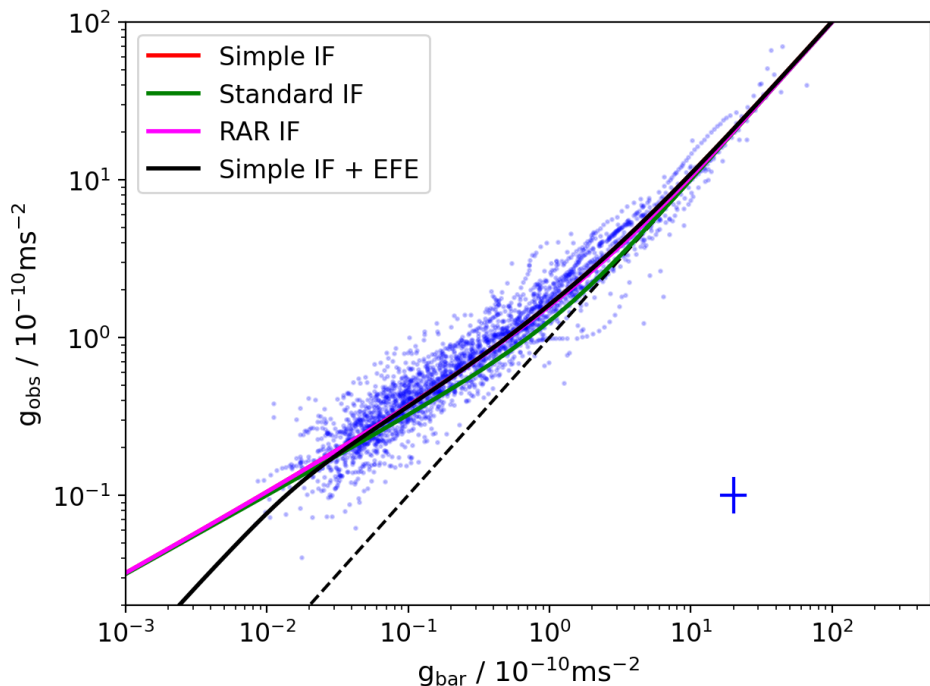
$$g_{\text{obs}} = g_{\text{bar}}/2 + \sqrt{g_{\text{bar}}^2/4 + g_{\text{bar}}g_0} \quad \text{“Simple” IF}$$

Famery & Binney 2005 (arXiv:astro-ph/0506723)

$$g_{\text{obs}} = g_{\text{bar}}(1 - \exp(-\sqrt{g_{\text{bar}}/g_0})) \quad \text{“RAR” IF}$$

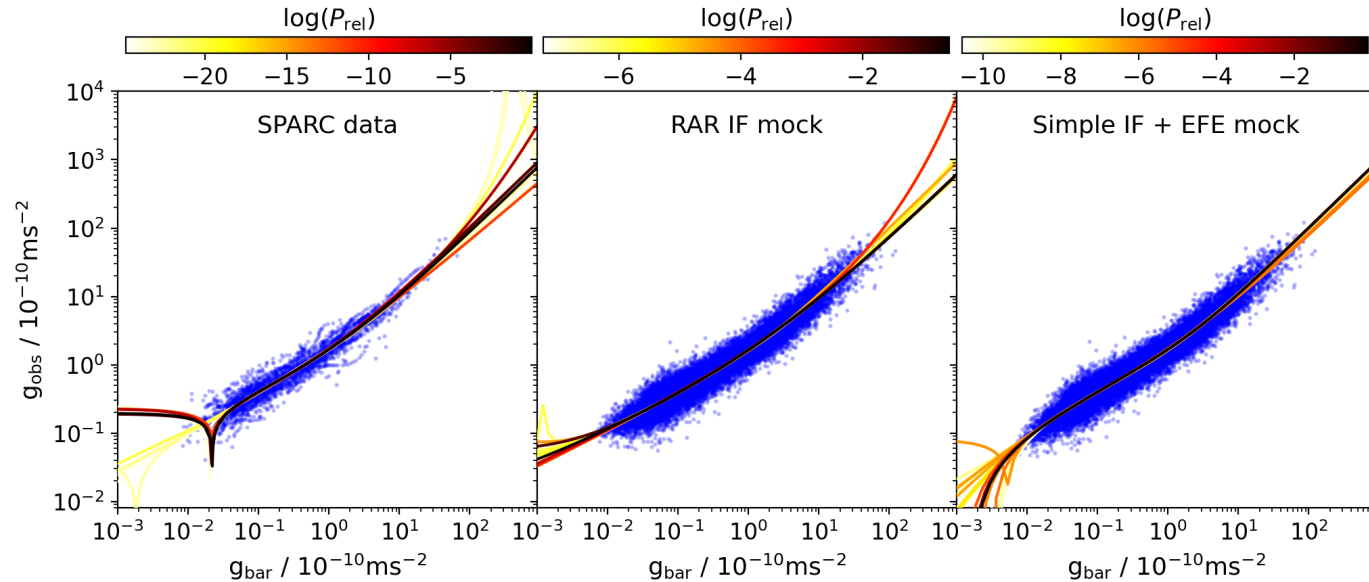
Lelli et al. 2017 (arXiv:1610.08981)

Basis: $\{g_{\text{bar}}, \theta, \exp, \sqrt{\cdot}, \text{square}, \text{inv}, +, -, \times, /, \text{pow}\}$



Best functions don't have MOND limit (even for MOND mocks)

1. RAR does not require MOND limit for low acceleration
2. Even if there was a MOND limit, the data are not good enough to find it
3. RAR neither demonstrates nor rules out a modification to gravity at low accelerations



<http://astroweb.cwru.edu/SPARC/>

Theory of Gravity

1. Best-fitting equations recover a Newtonian limit at high accelerations
 $(g_{\text{obs}} = g_{\text{bar}})$.
2. Observed galaxies do not prefer functions with the deep-MOND limit
 $(g_{\text{obs}} \propto \sqrt{g_{\text{bar}}})$ at low acceleration. They prefer $g_{\text{obs}} \rightarrow \text{const.}$
3. Mocks suggest SPARC data have uncertainties which are too large and an insufficient dynamic range to recover the true low-acceleration limit.

Conclusions

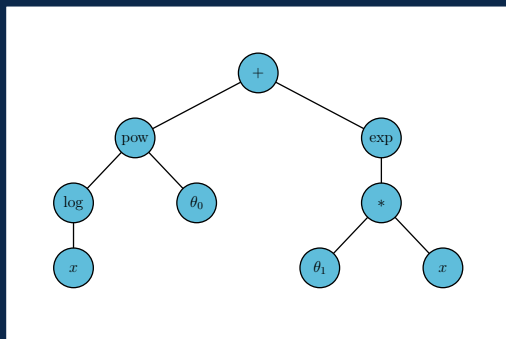
Bartlett, Desmond & Ferreira 2022 (arXiv:2211.11461)

Desmond, Bartlett & Ferreira 2022 (arXiv:2301.04368)

Bartlett, Desmond & Ferreira 2023 (arXiv:2304.06333)

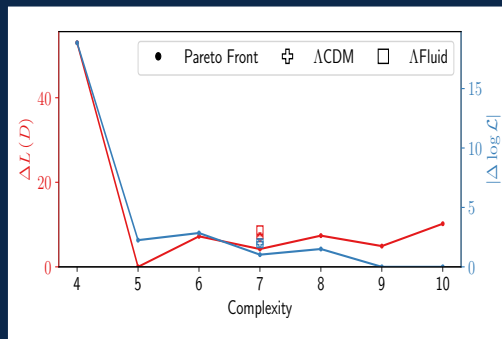
<https://github.com/DeaglanBartlett/ESR>

<https://github.com/DeaglanBartlett/katz>



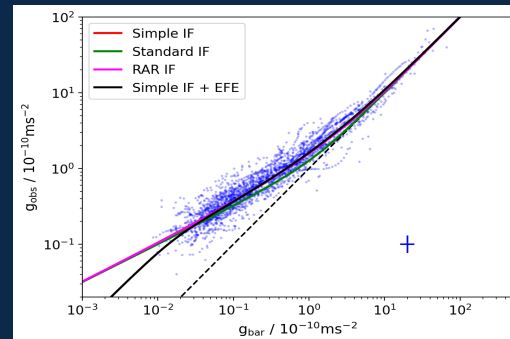
EXHAUSTIVE SR

- Guaranteed to find true optimum
- Combine simplicity and accuracy in a principled way



COSMIC EXPANSION

- ~40 better functions than standard Friedmann
- Low-redshift data do not necessarily prefer a Λ CDM expansion history



RADIAL ACCELERATION RELATION

- Best functions look Newtonian at large acceleration
- Best functions prefer non-MOND lower limit

Additional Slides

Precision of constants - tradeoff between accuracy and information needed

True $\hat{\theta}_i$ uniformly distributed within $\pm \Delta_i/2$ of transmitted value - may not transmit true MLP

Taylor expand log-likelihood about true value:

$$-\log(\mathcal{L}(\hat{\theta} + \mathbf{d})) \approx -\log(\mathcal{L}(\hat{\theta})) + \frac{1}{2} \mathbf{d}^T \mathbf{I} \mathbf{d}$$

Giving expected contribution to description length

$$L(\Delta) = \frac{1}{2} \sum_{ij} \langle \mathbf{I}_{ij} d_i d_j \rangle - \sum_i \log(\Delta_i)$$

Minimise this:

$$L(\Delta_i) = \frac{1}{24} \mathbf{I}_{ii} \Delta_i^2 - \log(\Delta_i) \quad \Rightarrow \quad \Delta_i = \left(\frac{12}{\mathbf{I}_{ii}} \right)^{1/2}$$

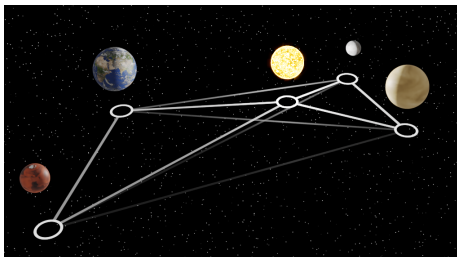
The Description Length of a Function

$$L(D) = -\log(\mathcal{L}(\hat{\theta})) + k \log(n) - \frac{p}{2} \log(3) + \sum_i^p \left(\frac{1}{2} \log(\mathbf{I}_{ii}) + \log(|\hat{\theta}_i|) \right)$$

$$\frac{1}{2} \left(p \log(N) - 2 \log(\mathcal{L}) \right) = \frac{1}{2} \text{BIC} \quad (\text{for large number of data points, } N)$$

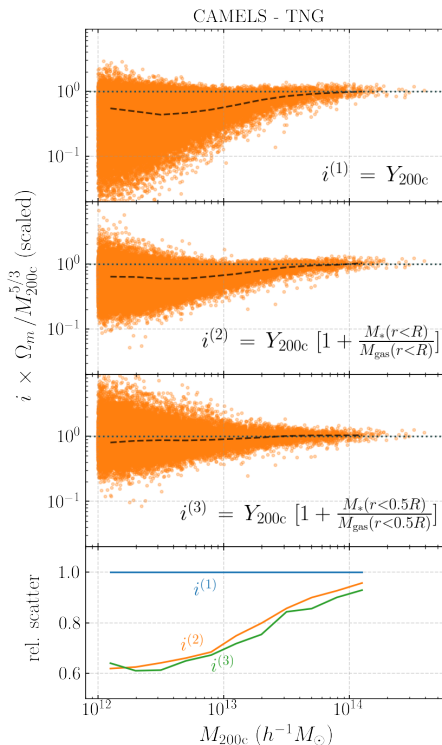
- Description length looks like BIC plus corrections due to structural complexity (prior on model)
- For large N , equivalent to minimising the BIC (an approximation to the evidence)

Previous applications of symbolic regression in physics



Rediscover Newtonian gravity from solar system data

Lemos et al. 2022 (arXiv:2202.02306)



tSZ Y-M relation which is more robust to feedback

Wadekar et al. 2022 (arXiv:2209.02075)

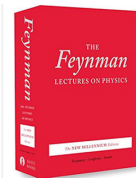


$$L = \frac{\hbar \omega^3}{\pi^2 c^3 (e^{\hbar \omega / k_B T} - 1)}$$

$$F = \frac{G m_1 m_2}{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

$$r = \frac{a(1 - e^2)}{1 + e \cos(\theta_1 - \theta_2)}$$

$$\frac{d\sigma}{d \cos \theta} = \frac{\pi \alpha^2 \hbar^2}{m^2 c^2} \left(\frac{\omega'}{\omega} \right)^2 \left(\frac{\omega'}{\omega} + \frac{\omega}{\omega'} - \sin^2 \theta \right)$$



Learn 100 equations from the Feynman Lectures on Physics

Udrescu & Tegmark 2019 (arXiv:1905.11481)

Cosmic Chronometers - Standard Clocks

$$H(z) = -\frac{1}{1+z} \frac{dz}{dt} \approx -\frac{1}{1+z} \frac{\delta z}{\delta t}$$



Image credit: AAS NOVA

- Passively evolving stellar population are standard clocks - measure δt
- Directly measure δz
- (Cosmological) model-independent measurement of $H(z)$
- We use sample of 32 CC $H(z)$ measurements

Jimenez & Loeb 2001 (arXiv:astro-ph/0106145)
 Moresco 2015 (arXiv:1503.01116)
 Moresco et al. 2016 (arXiv:1601.01701)
 Ratsimbazafy et al. 2017 (arXiv:1702.00418)
 Stern et al. 2010 (arXiv:0907.3149)
 Simon et al. 2004 (arXiv:astro-ph/0412269)
 Borghi et al. 2021 (arXiv:2110.04304)
 Zhang et al. 2012 (arXiv:1207.4541)
 Moresco et al. 2012 (arXiv:1201.3609)
 Moresco et al. 2022 (arXiv:2201.07241)

Type Ia Supernovae - Standard Candles

$$d_L(z) = (1+z) \int_0^z \frac{dz'}{H(z')}$$

$$\mu(z) = 5 \log_{10} \left(\frac{d_L(z)}{10 \text{ pc}} \right)$$

$$\mu = m_B + \alpha x_1 - \beta c - M_0$$

Amplitude Stretch Colour Rest-frame magnitude

Pantheon+ sample with SH0ES Cepheid calibration

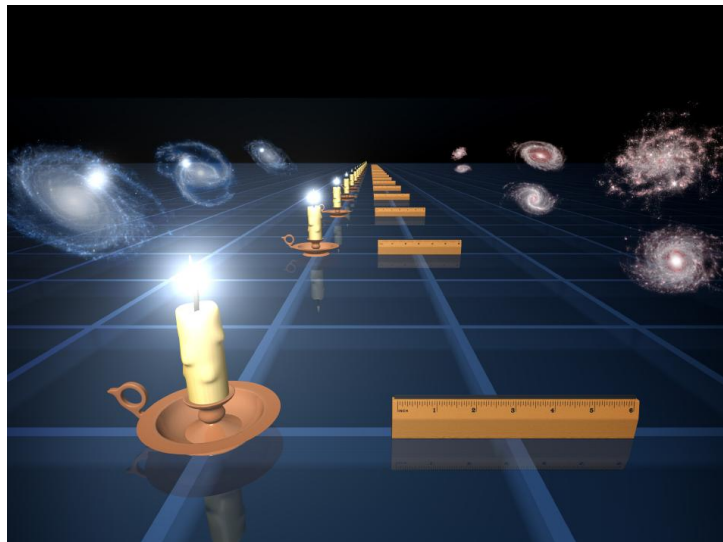


Image credit: NASA

Scolnic et al. 2021 (arXiv:2112.03863)
 Riess et al. 2022 (arXiv:2112.04510)

Best Functions - Cosmic Chronometers

Rank	$y(x) / \text{km}^2\text{s}^{-2}\text{Mpc}^{-2}$	Complexity	Parameters			Codelength			
			θ_0	θ_1	θ_2	Residuals ¹	Function ²	Parameter ³	Total
1	$\theta_0 x^2$	5	3883.44	-	-	8.36	5.49	2.53	16.39
2	$ \theta_0 ^{x^{\theta_1}}$	5	3982.43	0.22	-	7.97	5.49	5.24	18.70
3	$\theta_0 \theta_1 ^{-x}$	5	1414.43	0.31	-	7.57	6.93	5.58	20.08
4	$\theta_0 x^{\theta_1}$	5	3834.51	2.03	-	8.35	6.93	5.08	20.36
5	$x^2 (\theta_0 + x)$	7	3881.85	-	-	8.36	9.70	2.53	20.60
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
39	$\theta_0 + \theta_1 x^3$	9	3164.02	1481.71	-	7.28	12.48	3.76	23.51
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
84	$\theta_0 + \theta_1 x^{\theta_2}$	7	3322.96	1374.97	3.08	7.27	11.27	6.52	25.06

¹ $-\log \mathcal{L}(\hat{\theta})$

² $k \log(n) + \sum_j \log(c_j)$

³ $-\frac{p}{2} \log(3) + \sum_i^p (\frac{1}{2} \log(I_{ii}) + \log(|\hat{\theta}_i|))$

Best Functions - Pantheon+

Rank	$y(x) / \text{km}^2\text{s}^{-2}\text{Mpc}^{-2}$	Complexity	Parameters			Codelength			
			θ_0	θ_1	θ_2	Residuals ¹	Function ²	Parameter ³	Total
1	$\theta_0 x^x$	5	5345.02	-	-	706.18	6.93	5.11	718.22
2	$ \theta_0 ^{x^{\theta_1}}$	9	5280.11	0.16	-	705.11	5.49	8.41	719.01
3	$\theta_0 \theta_1 ^{-x}$	5	1694.95	0.32	-	701.79	6.93	10.33	719.05
4	$\theta_0 x^{x^{\theta_1}}$	7	5378.69	0.78	-	702.45	9.70	6.98	719.13
5	$ \theta_0 ^{ \theta_1 ^x}$	5	1898.47	1.14	-	701.88	5.49	12.64	720.02
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
37	$\theta_0 + \theta_1 x^3$	9	3591.09	1773.63	-	701.85	12.48	8.81	723.13
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
96	$\theta_0 + \theta_1 x^{\theta_2}$	7	3280.83	2069.32	2.73	701.64	11.27	12.19	725.10

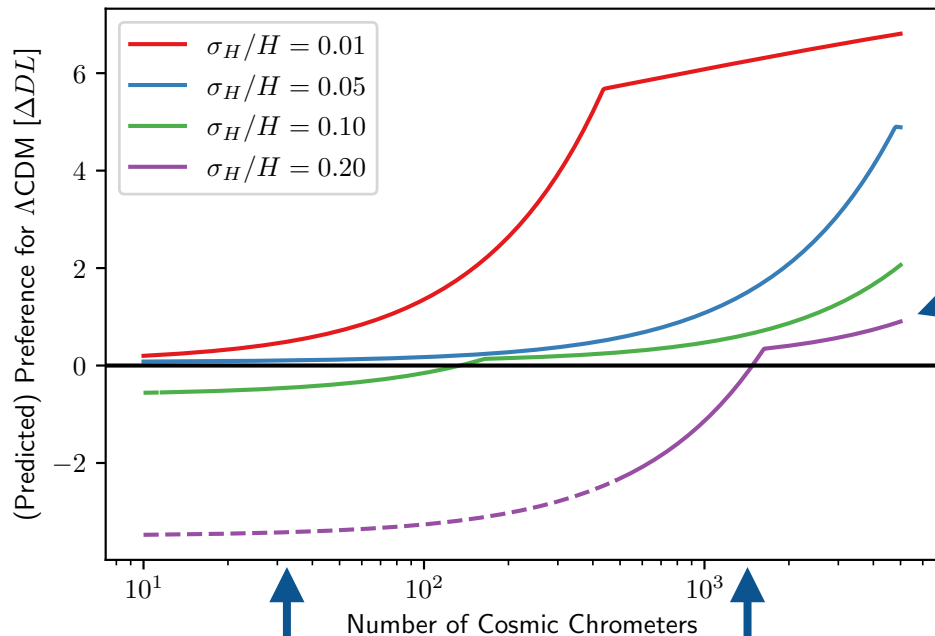
$$^1 - \log \mathcal{L}(\hat{\theta})$$

$$^2 k \log(n) + \sum_j \log(c_j)$$

$$^3 - \frac{p}{2} \log(3) + \sum_i^p (\frac{1}{2} \log(I_{ii}) + \log(|\hat{\theta}_i|))$$

Should we have seen Λ CDM? No.

Mock cosmic
chronometer
data assuming
 Λ CDM
(Planck18)



**Current
observational
uncertainties**

**We used
this many**

**Would need
this many**

Best Functions - SPARC

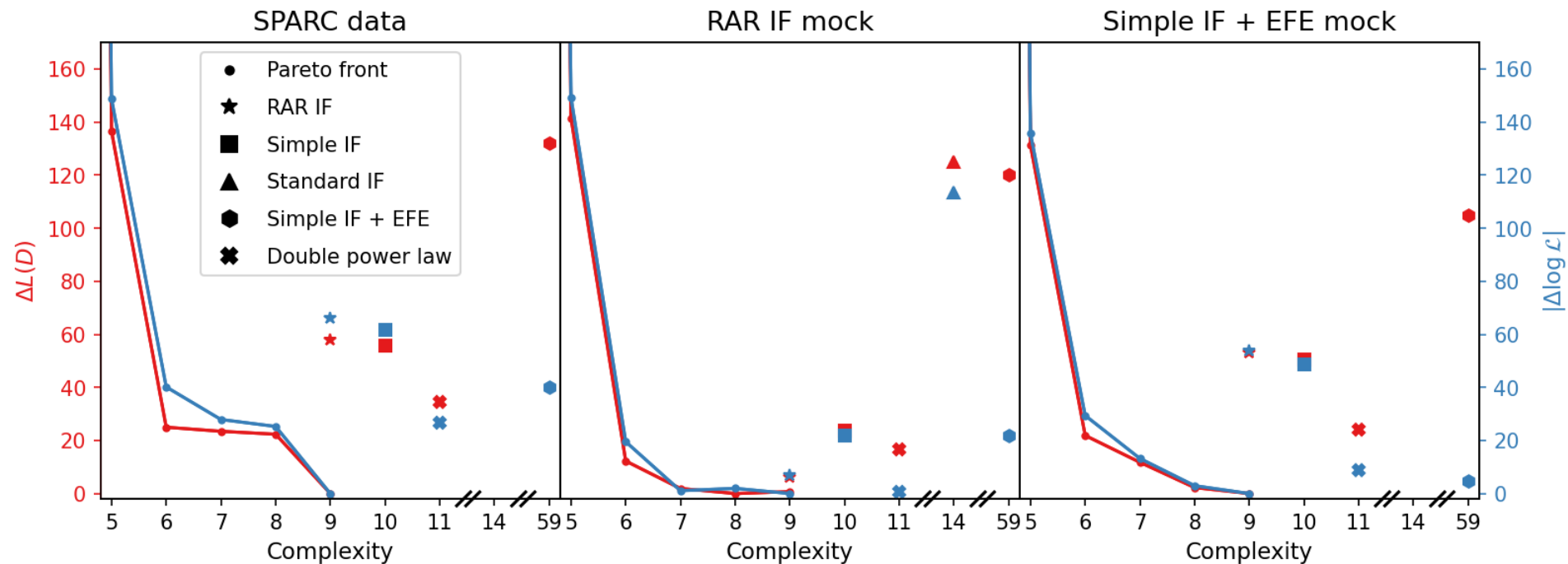
Rank	Function	Comp.	$P(f)$	Parameters				Description length			
				θ_0	θ_1	θ_2	θ_3	Resid. ¹	Func. ²	Param. ³	Total
1	$\theta_0 (\theta_1 + x ^{\theta_2} + x)$	9	9.3×10^{-1}	0.84	-0.02	0.38	—	-1279.1	14.5	14.0	-1250.6
2	$ \theta_1 ^x + \theta_0 ^{\theta_2} + x$	9	6.4×10^{-2}	-0.99	0.64	0.36	—	-1279.9	12.5	19.6	-1247.9
3	$ \theta_0 ^{ \theta_1 - x ^{\theta_2} - \theta_3}$	9	2.0×10^{-3}	-1.4×10^2	0.02	0.14	0.89	-1276.4	12.5	19.5	-1244.4
4	$ \theta_0(\theta_1 + x) ^{\theta_2} + x$	9	1.4×10^{-4}	0.35	-0.02	0.34	—	-1268.9	14.5	12.7	-1241.7
5	$ \theta_0 - \theta_1 - x ^{\theta_2} ^{\theta_3}$	9	1.0×10^{-5}	-0.30	0.02	0.42	2.14	-1271.1	12.5	19.5	-1239.1
6	$\sqrt{x} \exp\left(\frac{ \theta_0 + x ^{\theta_1}}{2}\right)$	9	1.5×10^{-9}	-0.02	0.36	—	—	-1257.9	17.5	10.0	-1230.3
7	$\left(\frac{ \theta_0 ^x}{x}\right)^{\theta_1} + x$	9	2.4×10^{-10}	1.87	-0.52	—	—	-1250.6	14.5	7.6	-1228.5
8	$\sqrt{ \theta_0 + x } + \theta_1 x$	8	1.8×10^{-10}	-1.8×10^{-3}	0.72	—	—	-1245.6	12.9	4.5	-1228.2
9	$\left \theta_0 + \frac{1}{\sqrt[3]{x}}\right ^{\theta_1}$	8	9.6×10^{-11}	-0.22	-2.14	—	—	-1251.1	14.3	9.2	-1227.6
10	$(\sqrt{x} + \frac{1}{x})^{\theta_0} + x$	9	8.2×10^{-11}	-0.53	—	—	—	-1248.3	16.1	4.8	-1227.4
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
17	$x/(\exp(\theta_0) - \theta_1 ^{\sqrt{x}})$	9	2.2×10^{-11}	0.03	0.44	—	—	-1250.9	17.5	7.3	-1226.1
—	Double power law	11	9.7×10^{-16}	4.65	3.96	1.03	0.60	-1252.3	17.7	18.5	-1216.1
—	Simple IF	10	5.5×10^{-25}	1.11	—	—	—	-1217.3	18.6	3.9	-1194.8
—	RAR IF	9	6.7×10^{-26}	1.13	—	—	—	-1212.8	16.1	3.9	-1192.7
—	Simple IF + EFE	59	5.0×10^{-69}	1.16	6.8×10^{-3}	—	—	-1238.9	139.9	5.6	-1093.4
—	Standard IF	14	9×10^{-150}	1.54	—	—	—	-939.5	27.9	4.1	-907.5

$$^1 - \log \mathcal{L}(\hat{\theta})$$

$$^2 k \log(n) + \sum_j \log(c_j)$$

$$^3 - \frac{p}{2} \log(3) + \sum_i^p (\log(I_{ii})^{1/2} + \log(|\hat{\theta}_i|))$$

Can always find functions which fit better than standard IFs

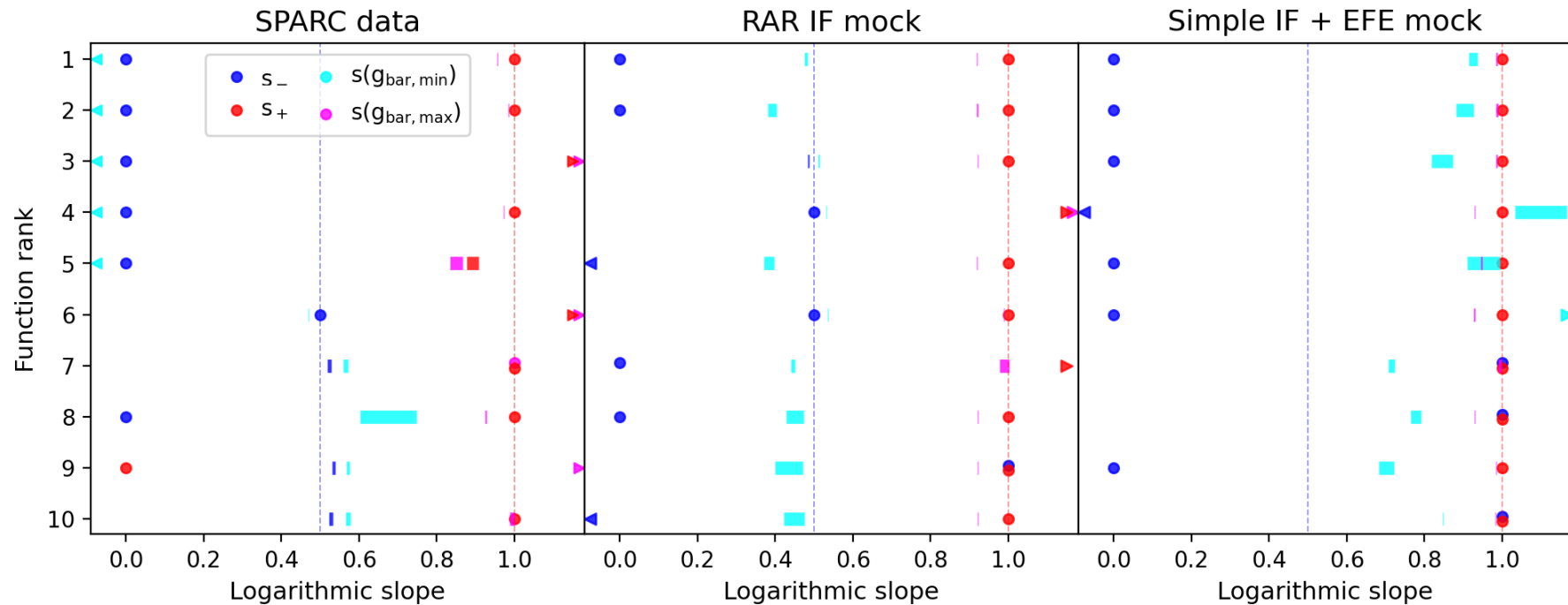


$$\theta_0 \left(|\theta_1 + x|^{\theta_2} + x \right)$$

$$\theta_0 + \theta_1 x + \sqrt{x}$$

$$\theta_0 + \sqrt{x^2 + 2x}$$

(Almost) always get Newtonian limit, (almost) never get MOND limit



$$\theta_0 \left(|\theta_1 + x|^{\theta_2} + x \right)$$

$$\theta_0 + \theta_1 x + \sqrt{x}$$

$$\theta_0 + \sqrt{x^2 + 2x}$$

Best Functions - RAR IF Mocks

Rank	Function	Comp.	$P(f)$	Parameters				Description length			
				θ_0	θ_1	θ_2	θ_3	Resid. ¹	Func. ²	Param. ³	Total
1	$\theta_0 + \theta_1 x + \sqrt{x}$	8	5.6×10^{-1}	9.1×10^{-3}	0.63	—	—	-2045.2	12.9	4.9	-2027.4
2	$\sqrt{ \theta_0 + x } + \theta_1 x$	8	2.8×10^{-1}	3.0×10^{-3}	0.64	—	—	-2044.4	12.9	4.8	-2026.7
3	$\theta_0 x + x^{\theta_1}$	7	8.2×10^{-2}	0.64	0.49	—	—	-2045.2	11.3	8.5	-2025.5
4	$\sqrt{x} \exp\left(\frac{x^{\theta_0}}{2}\right)$	7	3.5×10^{-2}	0.36	—	—	—	-2040.7	12.5	3.5	-2024.7
5	$(\theta_0 + x) \left(\theta_1 + \frac{1}{\sqrt{x}}\right)$	9	1.1×10^{-2}	1.3×10^{-3}	0.64	—	—	-2044.5	16.1	4.8	-2023.5
6	$\frac{1}{\sqrt{ \theta_0 + \frac{1}{x} }} + x$	8	8.8×10^{-3}	1.74	—	—	—	-2038.5	12.9	2.3	-2023.3
7	$(x \theta_0)^{(x \theta_1)^{\theta_2}}$	9	3.1×10^{-3}	-2.09	-1.4×10^{-4}	0.04	—	-2045.3	12.5	10.6	-2022.2
8	$\theta_0 x + \theta_1 + x ^{\theta_2}$	9	2.4×10^{-3}	0.64	1.4×10^{-3}	0.49	—	-2045.4	14.5	8.9	-2022.0
9	$x(\theta_0 - x ^{\theta_1} - \theta_2)$	9	2.3×10^{-3}	1.2×10^{-3}	-0.51	-0.64	—	-2045.3	14.5	8.9	-2021.9
10	$(\theta_0 - x)(\theta_1 - x^{\theta_2})$	9	2.2×10^{-3}	-6.5×10^{-4}	-0.64	-0.51	—	-2045.4	14.5	9.0	-2021.9
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
27	$x/(\exp(\theta_0) - \exp(-\sqrt{x}))$	9	3.2×10^{-4}	-0.01	—	—	—	-2039.3	17.5	1.9	-2020.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
41	$x/(\exp(\theta_0) - \theta_1 ^{\sqrt{x}})$	9	1.1×10^{-4}	-5.0×10^{-3}	0.38	—	—	-2042.1	17.5	5.7	-2018.9
—	RAR IF	9	1.0×10^{-3}	1.14	—	—	—	-2041.1	16.1	3.9	-2021.1
—	Double power law	11	3.4×10^{-8}	1.25	1.47	0.90	0.54	-2047.2	17.7	18.7	-2010.8
—	Simple IF	10	2.8×10^{-11}	1.12	—	—	—	-2026.2	18.6	3.9	-2003.7
—	Standard IF	14	2.9×10^{-55}	1.54	—	—	—	-1934.4	27.9	4.1	-1902.4
—	Simple IF + EFE	59	5.9×10^{-64}	1.12	0	—	—	-2026.2	139.9	3.9	-1882.4

$$^1 - \log \mathcal{L}(\hat{\theta})$$

$$^2 k \log(n) + \sum_j \log(c_j)$$

$$^3 - \frac{p}{2} \log(3) + \sum_i^p (\log(I_{ii})^{1/2} + \log(|\hat{\theta}_i|))$$

Best Functions - Simple IF + EFE Mock

Rank	Function	Comp.	$P(f)$	Parameters				Description length			
				θ_0	θ_1	θ_2	θ_3	Resid. ¹	Func. ²	Param. ³	Total
1	$\theta_0 + \sqrt{x^2 + 2x}$	9	8.9×10^{-1}	-0.06	—	—	—	-2017.7	14.5	3.1	-2000.0
2	$\theta_0 + \sqrt{x \theta_1 + x }$	8	9.3×10^{-2}	-0.06	1.97	—	—	-2017.9	12.9	7.3	-1997.8
3	$- \theta_0 \sqrt{x} + \theta_1 + x$	8	5.6×10^{-3}	0.26	0.95	—	—	-2017.9	12.9	10.1	-1995.0
4	$(\theta_0 - x)(\theta_1 - x^{\theta_2})$	9	3.3×10^{-3}	3.1×10^{-3}	-0.71	-0.53	—	-2019.7	14.5	10.7	-1994.4
5	$x^{\theta_0} - \theta_1(\theta_2 - x)$	9	2.4×10^{-3}	0.39	0.79	0.12	—	-2020.9	14.5	12.3	-1994.1
6	$ \theta_0 - x ^{\theta_1} - \theta_2 x$	9	2.0×10^{-3}	5.5×10^{-3}	0.48	-0.71	—	-2019.1	14.5	10.6	-1994.0
7	$x \theta_0 ^{- \theta_1 ^{x^{\theta_2}}}$	9	1.7×10^{-3}	0.04	-0.16	0.33	—	-2018.1	12.5	11.9	-1993.8
8	$x(\theta_0 + \theta_1 + x ^{\theta_2})$	9	1.5×10^{-3}	0.71	0.01	-0.53	—	-2018.7	14.5	10.6	-1993.7
9	$ \theta_0 \theta_1 ^{x^{\theta_2}} + x$	9	6.5×10^{-4}	7.0×10^{-6}	0.03	0.17	—	-2016.7	12.5	11.4	-1992.8
10	$\exp\left(\theta_0 - \frac{1}{\sqrt[3]{x}}\right) + x$	9	5.5×10^{-4}	0.57	—	—	—	-2014.0	17.5	3.9	-1992.6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
21	$x/(\exp(\theta_0) - \theta_1 \sqrt{x})$	9	1.8×10^{-5}	0.03	0.44	—	—	-2014.2	17.5	7.4	-1989.3
—	Double power law	11	3.4×10^{-11}	3.53	3.31	0.98	0.60	-2012.3	17.7	18.6	-1976.0
—	Simple IF	10	1.2×10^{-22}	1.11	—	—	—	-1972.1	18.6	3.9	-1949.6
—	RAR IF	9	7.0×10^{-24}	1.13	—	—	—	-1966.9	16.1	3.9	-1946.8
—	Simple IF + EFE	59	3.8×10^{-57}	1.19	8.6×10^{-3}	—	—	-2016.0	139.9	5.9	-1870.2
—	Standard IF	14	2×10^{-141}	1.54	—	—	—	-1708.3	27.9	4.1	-1676.3

$$^1 1 - \log \mathcal{L}(\hat{\theta})$$

$$^2 k \log(n) + \sum_j \log(c_j)$$

$$^3 -\frac{p}{2} \log(3) + \sum_i^p (\log(I_{ii})^{1/2} + \log(|\hat{\theta}_i|))$$