

Symbolic regression benchmarked for cosmology

M. E. Thing and S. M. Koksang

CP3-Origins, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark

E-mail: thing@cp3.sdu.dk, koksang@cp3.sdu.dk

Abstract. TO BE REMOVED BEFORE SUBMISSION: APPROPRIATE REFEREE SUGGESTIONS: Deaglan Bartlett, William La Cava,

Contents

1	Introduction	1
2	Data	2
2.1	Redshift-distance relation in standard cosmology	2
2.2	Redshift drift in standard cosmology	3
2.3	Cosmic backreaction and two-region models	4
2.3.1	Redshift drift in 2-region models	7
3	Benchmarking	8
3.1	Benchmarking method	8
3.2	Benchmarking results	8
4	Summary and conclusions	8

1 Introduction

It is a foundational aspect of physics that we are able to use symbolic expressions to describe how the world around us works. It is therefore a significant challenge when we meet a physical system, observable or an experimental dataset that we cannot describe through mathematical formulas. This problem arises various places in astrophysics and cosmology where, for instance, the precise halo mass function is unknown [1], stellar density profiles are estimated based on various approximations but are generally unknown [2], a derived expression for the non-linear matter power spectrum does not yet exist [3], and where intricacies of general relativity leave useful expressions for some observables such as the mean redshift drift [4, 5] still waiting to be found (mean redshift drift will be introduced in section 2).

Fittingly, a significant number of freely available algorithms aimed at solving regression problems have been developed in the recent years. The algorithms are based on a variety of machine learning and deep learning approaches ranging from the more traditional pure genetic algorithms (such as [6]) to tree-based genetic programming [7], methods incorporating language models (e.g. [8]), models relying on deep neural networks (such as [9, 10]), just to name a few (see e.g. [11, 12] for reviews of symbolic regression and their different strategies and performance on standard datasets). Several of these algorithms have even been developed with physics in mind (e.g. [9, 10, 13]) and one motivation for the development of symbolic regression algorithms is exactly to be able to discover hitherto unknown physics formulae (noted in e.g. [9, 13]). Along this vein, symbolic regression has indeed been shown to be able to identify known physical relations [9, 10, 14], been used to study Hubble data [13, 15] and learn about galactic dynamics [16], studying inflationary potentials [17] and galaxy halos [18, 19]. In addition, symbolic regression has been used to attempted revealing unknown functional forms in cosmology, such as identifying an expression for the mean redshift drift and cosmic backreaction [4, 5], and the non-linear matter power spectrum [20].

In the works listed above, where symbolic regression has been used for cosmology, the approach has generally been to use one or very few specific algorithms. For instance, the work in [4, 5] (by one of the current authors) utilized the symbolic regression algorithms AI Feynman [9, 10] based on its user friendliness and because it was developed by physicists with

discovering physics formulae in mind and showed an improvement compared to earlier state-of-the-art algorithms. The work presented in [20] instead utilized OPERON [21] which was chosen due to its speed, memory efficiency and because it is based on genetic algorithms which tend to do well in symbolic regression benchmark studies and competitions (e.g. [12, 22]).

In general, when cosmologists choose an algorithm for symbolic regression, we naturally make a choice we expect representative of state-of-the-art for the task. However, symbolic regression algorithms are usually evaluated and compared by studying their efficiency at identifying symbolic expressions corresponding to certain ensembles of datasets such as the Penn Machine Learning Benchmarks [23, 24], collections of mathematical expressions (See e.g. the appendix A of [11]) and some even use a dataset based on 100 equations from the Feynman Lectures on Physics [25], the so-called Feynman Symbolic Regression Database¹. Although these ensembles often do include physics datasets (not the least the Feynman Symbolic Regression Database), this does not necessarily mean that benchmarking with them result in benchmarkings that are particularly enlightening when it comes to the applicability to real cosmology symbolic regression tasks. One hindrance is that real world cosmology problems are typically based on very complex physics which means that the corresponding datasets can be difficult to obtain, understand and can end up being complicated from a physical point of view, e.g. including several dependent variables. This type of datasets simply does not seem to be represented significantly in common machine learning/symbolic regression databases.

It is not clear that symbolic regression algorithms optimized with typical machine learning database problems are also optimized for real world cosmology problems. To understand whether or not this is indeed the case, we here present a benchmarking of symbolic regression algorithms based on specific cosmology datasets. While typical benchmark tasks can consist of one or several hundreds of datasets, we here restrict the benchmark to only a few datasets which are, however, genuine cosmology datasets related to and representing genuine symbolic regression tasks in cosmology.

In section 2 below, we introduce the datasets we used for the benchmarking, including an introduction to the cosmological settings they represent. In section 3 we present the benchmarking setup and results while section 4 serves as a summary where we also provide a discussion of future possible developments.

2 Data

This section serves to describe the data used for the benchmarking. In the first subsection below, we introduce the redshift-distance relation used in standard cosmology. The subsequent subsection is dedicated to the observable known as redshift drift and its Friedmann-Lemaitre-Robertson-Walker (FLRW) limit. In the final two subsections we introduce a specific family of toy cosmological models called two-region models and discuss how to compute the redshift drift as well as a quantity known as cosmic backreaction in these models.

2.1 Redshift-distance relation in standard cosmology

Earlier work has included showcasing the use of new symbolic regression algorithms for astrophysics by applying them to supernovae data (e.g. [13, 15]). Supernova data is one of the most important data sets in cosmology and is the closest to direct evidence for accelerated expansion of the Universe currently available. For our benchmark, we do not wish to use real

¹<https://space.mit.edu/home/tegmark/aifeynman.html>

data where the ground truth of the corresponding symbolic expression is unknown. In order to understand the significance of the showcase in [13, 15] we will, however, include a mock supernova dataset as our first benchmark dataset.

Standard cosmology is based on the Friedman-Lemaitre-Robertson-Walker (FLRW) space-times, where the redshift-distance relation can be written as

$$D_L(z) = \frac{1+z}{\sqrt{K/R_0^2}} \sin \left(\sqrt{K/R_0^2} c \int_0^z \frac{dz'}{H(z')} \right), \quad (2.1)$$

where $K = \pm 1, 0$ is the sign of the curvature and R_0 is the curvature radius of the model. In Λ CDM models we can write the Hubble parameter, $H(z)$, as

$$H(z)^2 = \frac{\Omega_{m,0}}{a^3} + (1 - \Omega_{m,0}). \quad (2.2)$$

Supernovae data is usually presented in terms of the distance modulus, μ , and we therefore covert the luminosity distance, D_L , to μ according to

$$\mu = 5 \log_{10} \left(\frac{D_L}{1 \text{ Mpc}} \right) + 25. \quad (2.3)$$

We include 2 datasets based on the distance modulus in our benchmark. Following [13, 15] we will consider flat Λ CDM models only. Using this set of models, we generate 3 datasets. First, we generate two datasets meant to mimic real supernovae datasets by specifying a single model with $H_0 = 70 \text{ km/s/Mpc}$ and $\Omega_{m,0} = 0.3$. One dataset will consist of (μ, z) without noise while the second dataset will contain 10% Gaussian noise on μ . For each dataset, we compute 1000 data points distributed equidistantly in the redshift interval $z \in [0, 2]$, roughly representing the redshift interval of real supernova datasets. We will also consider a more complicated dataset, where we let H_0 and $\Omega_{m,0}$ vary. We thus generate datasets of the form $(\mu, z, H_0, \Omega_{m,0})$ where we let $H_0 \in [20, 100]$ in units of km/s/Mpc and $\Omega \in [0.1, 0.5]$. Again, we use equidistant points, now with 100 points for each feature. This dataset, as well as those discussed in the following subsections, are motivated by the ambition to decipher which symbolic regression methods are most suitable for discovering hitherto unknown observational relations. Our rationale is that observational relations must be valid for more than a single set of model parameters i.e. not just for a single Λ CDM model but rather for, e.g., *all* FLRW models. We therefore wish to test symbolic regression methods on how well they can describe data representing one or more families of cosmological models such as sub-families of FLRW models rather than just a single FLRW model. Since this type of dataset will be inherently more complicated than datasets with only a single feature, we will not add noise to these datasets.

2.2 Redshift drift in standard cosmology

One of the biggest mysteries in cosmology is the physical nature of dark energy i.e. the source of accelerated expansion of the Universe [26]. The apparent accelerated expansion of the universe has, however, never been measured directly. The observable *redshift drift* represents a possibility to remedy this. The redshift of a luminous astronomical object changes with time. Since the change is slow, it has been dubbed the redshift *drift*. Redshift drift was first discussed in [27, 28] from 1962 but it was there assessed that the effect was too small to be measureable. Due to technological advances, it is presently expected that the effect can

be measured with e.g. the Square Kilometer Array (SKA)[29]. In standard cosmology, the redshift drift δz can be written as

$$\delta z = \delta t_0[(1+z)H_0 - H(z)], \quad (2.4)$$

where $H(z)$ is the Hubbel parameter, H_0 the Hubble constant and δt_0 is the time interval of the observation such that an object observed to have redshift z at time t_0 will be observed to have the redshift $z + \delta z$ at time $t_0 + \delta t_0$.

The expression for the redshift written above is fairly simple. Nonetheless, it represents an interesting symbolic regression problem because the right hand side of the equation contains a feature $H(z)$ which depends on one of the other features (z) in a non-trivial way. Earlier studies into machine and deep learning algorithms have included datasets where two or more variables were dependent, these dependencies are typically linear or otherwise simple (see e.g. [30] for an example). However, to the authors' knowledge, there have not been conducted systematic studies of this and there has not been conducted studies considering datasets where features have more complicated dependencies such as in the case of the Hubble parameter and its dependence on the redshift. In cosmology, observable relations are often more convenient to consider in terms of several dependent features though. Besides the redshift drift, this is for instance the case for redshift-distance relations which are usually written as an integral over the Hubble parameter in terms of the redshift multiplied by factors containing the redshift itself as well. In our benchmarking dataset we will therefore include the redshift drift as written above, i.e. with the target being $\delta z/\delta t_0$ and the features being z and $H(z)$. However, we will consider only flat, single-component FLRW models when generating the datasets. In this case, a simple rewriting gives that the redshift drift can be written as

$$\delta z = \delta t_0 H_0 [(1+z) - (1+z)^{3(1+\omega)/2}], \quad (2.5)$$

where ω is the equation-of-state parameter for the single component of the content of the model universe.

Using the formalism discussed above, we have generated two redshift drift datasets, $(\delta z/\delta t_0, z, H(z))$ and $(\delta z/\delta t_0, z, \omega)$. The two models contain the same number of features, has the same target and contains the same number of data points generated using 50 equidistant points in each of the two intervals $z \in [0.1, 1]$ and $\omega \in [-0.99, 1/3]$. When choosing the redshift interval, we deliberately avoid the very lowest redshifts to avoid a target very close to zero. We choose the upper limit $z = 1$ since this is the upper limit expected for main redshift drift measurements from SKA [29]. For ω we choose the upper limit of $1/3$ which corresponds to the equation-of-state parameter for radiation while we choose a lower limit close to that of a cosmological constant, although neglecting the value -1 to keep the overall form of the Hubble parameter the same for all models².

2.3 Cosmic backreaction and two-region models

The expression for the redshift drift given in the previous section is valid only for FLRW spacetimes, i.e. in spacetimes which have no structures on any scales. Standard cosmology is based on the assumption that the large-scale dynamics of the Universe can be approximated well by the FLRW models but it has also been discussed whether the structures on smaller

²Remember that for flat, single component universes, $H(z) = \frac{2}{3(1+\omega)t}$ except for the case $\omega = -1$ where the Hubble parameter is constant.

scales can affect the large-scale/averaged dynamics. This possibility is known as cosmic backreaction [31–34] and it has even been suggested that the seemingly accelerated expansion of the Universe is an artifact due to this phenomenon (the co-called backreaction conjecture – see e.g. [35–38]). Backreaction and the backreaction conjecture are most commonly considered through the Buchert equations presented in [39–41] where dynamical equations describing the average (large-scale) evolution of the Universe are obtained by averaging Raychaudhuri’s equation and the Hamiltonian constraint using the averaging definition

$$s_D := \frac{\int_D s dV}{\int_D dV}, \quad (2.6)$$

where s is some scalar, dV the proper infinitesimal volume element on the spatial domain D assumed to be larger than the homogeneity scale. This definition can be used when considering spacetimes foliated with spatial hypersurfaces orthogonal to the fluid flow (possible for irrotational fluids such as dust) and where the line element is written as

$$ds^2 = -dt^2 + g_{ij}dx^i dx^j. \quad (2.7)$$

The resulting averaged Hamiltonian constraint and Raychaudhuri equation can be written as (setting $c = 1$)

$$\begin{aligned} 3H_D^2 &= 8\pi G\rho_D - \frac{1}{2}R_D - \frac{1}{2}Q_D \\ 3\frac{\ddot{a}_D}{a_D} &= -4\pi G\rho_D + Q_D. \end{aligned} \quad (2.8)$$

$H_D := \dot{a}_D/a_D$ denotes the average Hubble parameter and is related to the local fluid expansion scalar θ by $H_D = \theta_D/3$. The volume/averaged scale factor a_D is unlike the scale factor in FLRW spacetimes not related to the metric (which is not considered in the Buchert averaging scheme) but is instead defined as $a_D := (V_D/V_{D0})^{1/3}$, where V_D is the proper volume of the spatial domain D and subscripts zero indicate evaluation at present time.

The dynamical equations for the average universe shown above are reminiscent of the Friedmann equation and acceleration equation but contain an extra term, namely $Q_D := 2/3 [(\theta^2)_D - (\theta_D^2)] - (\sigma_{\mu\nu}\sigma^{\mu\nu})_D$, where $\sigma_{\mu\nu}$ is the shear tensor of the fluid. In addition, the curvature term R_D can deviate from the FLRW behavior where the curvature must be proportional to the inverse of the squared scale factor. Q_D is sometimes referred to as the kinematical backreaction while the deviation of R_D from having FLRW evolution is referred to as intrinsic backreaction. Combined, these two differences compared to the FLRW models mean that the large-scale dynamics of an inhomogeneous universe does not necessarily follow FLRW dynamics even when averaging above the homogeneity scale.

The two backreaction terms are coupled through the integrability condition

$$a_D^{-6}\partial_t(a_D^6 Q_D) + a_D^{-2}\partial_t(a_D^2 R_D) = 0. \quad (2.9)$$

This equation ensures that the two Buchert equations are in accordance with each other. However, the two Buchert equations and the integrability condition do not form a closed set so cannot be used to predict backreaction, not even qualitatively. Therefore, it is not known how backreaction e.g. can be parameterized in terms of the scale factor (or equivalently, the mean redshift $\langle z \rangle + 1 \approx 1/a_D$ [42, 43]). As discussed in e.g. [4, 5], this means that backreaction cannot be sensibly constrained with current cosmological data. This motivated

an attempt (presented in [4, 5]) to obtain parameterizations for backreaction using toy cosmological models where backreaction can be computed, combined with symbolic regression to identify symbolic expression for backreaction. That study is extended here, where several datasets based on the so-called two-region models introduced in [44, 45] will be used in the benchmarking.

Two-region models are toy-cosmological models consisting of an ensemble of two different FLRW models. Since the ensemble is disjoint, the model is not an exact solution to Einstein's equation. The model is nonetheless convenient to use for studying backreaction because the model is numerically fast and simple to work with. Following the procedure in [44, 45] we will consider a two-region model where one model is the empty (Milne) solution while the other region is an overdense (density larger than the critical density) matter+curvature model. In this case, the scale factors, a_o and a_u , of the two regions can be related by a common time coordinate, t , by using a parameter, ϕ (sometimes denoted the development angle), according to

$$\begin{aligned} t &= t_0 \frac{\phi - \sin(\phi)}{\phi_0 - \sin(\phi_0)} \\ a_u &= \frac{f_u^{1/3}}{\pi} (\phi - \sin(\phi)) \\ a_o &= \frac{f_o^{1/3}}{\pi} (1 - \cos(\phi)). \end{aligned} \tag{2.10}$$

The parameters f_u and f_o denote the present time relative volume fractions of the two regions in the ensemble and are related by $f_u = 1 - f_o$. As in [44, 45], present time is defined to correspond to $\phi_0 = 3/2\pi$. As noted in e.g. [5], once f_u or f_o has been fixed, there is still one parameter which needs to be set in order to specify a two-region model uniquely. This is seen by noting that $H_D = H_u(1 - v + vh)$, where v is the time dependent volume fraction of the overdense region and $h := H_o/H_u$. Since $H_u = 1/t$ we therefore have $t_0 = (1 - v_0 + v_0 h_0)/H_{D_0}$ and thus need to fix either t_0 or H_{D_0} . We will choose $H_{D_0} = 70\text{km/s/Mpc}$.

Since both Q_D and R_D have unknown parameterizations in terms of a_D , we have used two two-region datasets for the benchmarking, namely $(Q_D, z, \Omega_{m,0}, \Omega_{R,0}, \Omega_{Q,0})$ and $(R_D, z, \Omega_{m,0}, \Omega_{R,0}, \Omega_{Q,0})$, where we have defined $\Omega_{m,0} := 8\pi/(3H_{D_0}^2)\rho_{D,0}$, $\Omega_{R,0} := -R_{D,0}/(6H_{D,0}^2)$ and $\Omega_{Q,0} := -Q_{D,0}/(6H_{D,0}^2)$. Based on the findings in [4, 5], the data points were generated using equidistant points in the intervals $z \in [0, 1]$ and $f_o \in [0.1, 0.25]$.

Since a ground truth for the symbolic expression for Q_D and R_D in terms of the features in the datasets are unknown, we assess the success of the symbolic regression algorithms by 1) evaluating their precision on the training model subspace, and 2) evaluating their precision on data generated outside the model subspace used for training. Specifically, we will consider models in the larger interval $z \in [0, 5]$ and $f_o \in [0.05, 0.3]$.

As detailed in [44, 45], the averaged deceleration parameter q_D can be written as

$$\begin{aligned} q_D &:= -\frac{1}{H_D^2} \frac{\ddot{a}_D}{a_D} = q_u \frac{1-v}{(1-v+hv)^2} + q_o \frac{vh^2}{(1-v+hv)^2} - 2 \frac{v(1-v)(1-h)^2}{(1-v+hv)^2} \\ &= q_o \frac{vh^2}{(1-v+hv)^2} - 2 \frac{v(1-v)(1-h)^2}{(1-v+hv)^2}, \end{aligned} \quad (2.11)$$

where the second line comes from the empty region being coasting and hence $q_u = 0$. The last term is the backreaction term due to Q_D . By comparing with the equations 2.8 we see that we can write Q_D as e.g.

$$\begin{aligned} Q_D &= 6H_D^2 \frac{v(1-v)(1-h)^2}{(1-v+hv)^2} \\ &= -6 \left(\frac{\ddot{a}_D}{a_D} + \frac{1}{2a_D^3} H_{D_0}^2 \Omega_{m,0} \right). \end{aligned} \quad (2.12)$$

The above two expressions are fairly straight forward analytical expressions. However, both expressions contain several dependent variables since v, h, H_D, \ddot{a}_D and a_D all depend on time (differently) which may make the regression task more complicated. We will therefore include (Q_D, H_D, v, h) and $(Q_D, H_{D_0}, \Omega_{m,0}, a_D, \ddot{a}_D)$ as datasets used for the benchmarking to complement the earlier datasets for the target Q_D where there is no known ground truth expression.

2.3.1 Redshift drift in 2-region models

The redshift drift is a particularly interesting observable in relation to backreaction and the backreaction conjecture because it can be used as a smoking gun for distinguishing between accelerated expansion due to dark energy versus backreaction [46–49]. The foundation for the smoking gun signal is that the mean redshift drift is not in general equal to the drift of the mean redshift in an inhomogeneous spacetime even if that spacetime is statistically homogeneous (where “mean” observations here indicate observations after taking the mean over many lines of sight in order to remove statistical fluctuations). Mathematically, we can write this statement as

$$\langle \delta z \rangle \neq \delta \langle z \rangle = \delta t_0 [(1 + \langle z \rangle) H_{D,0} - H_D(\langle z \rangle)], \quad (2.13)$$

where we set $\langle z \rangle + 1 = 1/a_D$. As discussed in [4, 5], an expression for $\langle z \rangle$ in inhomogeneous cosmological models exhibiting backreaction is unknown. The redshift drift can nonetheless be computed in concrete cosmological models exhibiting backreaction as was done in [46, 47], with the former utilizing two-region models.

The procedure for computing the redshift drift in two-region models is to arrange copies of the two FLRW regions consecutively after one another along a light ray propagated according to the equations

$$\frac{dt}{dr} = -a \quad (2.14)$$

$$\frac{dz}{dr} = (1+z)\dot{a} \quad (2.15)$$

$$\frac{d\delta z}{dr} = \dot{a}\delta z + (1+z)\ddot{a}\delta t \quad (2.16)$$

$$\frac{d\delta t}{dr} = -\dot{a}\delta t, \quad (2.17)$$

where scale factors are evaluated according to the local model the light ray is in at a given time.

By solving the above equations, we generate datasets of the form $(\langle \delta z \rangle / \delta t_0, \langle z \rangle, \Omega_{m,0}, \Omega_{R,0}, \Omega_{Q,0})$, $(\langle \delta z \rangle / \delta t_0, \langle z \rangle, H_D(\langle z \rangle), \Omega_{m,0}, \Omega_{R,0}, \Omega_{Q,0})$ and $((\langle \delta z \rangle - \delta \langle z \rangle) / \delta t_0, \langle z \rangle, \Omega_{m,0}, \Omega_{R,0}, \Omega_{Q,0})$, where the two latter datasets are inspired by the fact that we know the symbolic expression for the FLRW redshift drift requires either an equation of state parameter or Hubble parameter. The datasets were generated using 50 equidistant points of $f_o \in [0.1, 0.25]$ and X values of the redshift in the interval $z \in [0.1, 1]$.

As with the earlier two-region model datasets, we will assess the quality of the symbolic expressions obtained by the regression algorithms by seeing how well they generalized outside the model subspace the algorithms were trained on.

3 Benchmarking

We need some overall considerations regarding symbolic regression here, evt. med udgangspunkt i de algoritmer vi tester med highlight af, hvordan de adskiller sig fra hinanden.

3.1 Benchmarking method

3.2 Benchmarking results

Remember to comment on units in final expressions. E.g. [13] claim that symbolic expressions are nonesensical from a physical point of view if the units do not match and therefore introduce a “unit” prior in their symbolic regression algorithm. But that is too strong a claim because units can be “hidden” in the prefactors. NOTE: WE COULD EVEN TRY TO RE-CAST PREFACTORS IN TERMS OF c , G and π OF SYMBOLIC EXPRESSIONS IF ANY SEEM REPRESENTATIVE OF A GROUND TRUTH.

4 Summary and conclusions

E.g. discuss something about machine learning not aimed at precision: 2210.13447. Also discuss that simplicity is often optimised in SR algorithms which is sensible from a physical perspective because physical laws generally are not very complex. But at the same time, note that the complexity is no hindrance for utilising expressions for e.g. parameter determination (2307.16468). Therefore, it may be that better machine learning algorithms for cosmology can be achieved by being less strict about reducing complexity of symbolic expressions.

Comment on whether your benchmarking shows a significant difference in which models are best/worst compared to the original benchmarking of 2107.14351.

Acknowledgments

We thank the authors of [12] and in particular Fabrício Olivetti de França and William La Cava for correspondence.

SMK is funded by VILLUM FONDEN, grant VIL53032. Part of the numerical work done for this project was performed using the UCloud interactive HPC system managed by the eScience Center at the University of Southern Denmark.

Author contribution statement:

References

- [1] Z. Xu, “Dark matter halo mass functions and density profiles from mass and energy cascade,” *Scientific Reports*, vol. 13, Oct. 2023.
- [2] A. Sonnenfeld, “A robust two-parameter description of the stellar profile of elliptical galaxies,” *Astronomy & Astrophysics*, vol. 641, p. A143, Sept. 2020.
- [3] D. J. Bartlett, L. Kammerer, G. Kronberger, H. Desmond, P. G. Ferreira, B. D. Wandelt, B. Burlacu, D. Alonso, and M. Zennaro, “A precise symbolic emulator of the linear matter power spectrum,” 2023.
- [4] S. M. Kocsbang, “Machine learning cosmic backreaction and its effects on observations,” *Physical Review Letters*, vol. 130, May 2023.
- [5] S. M. Kocsbang, “Cosmic backreaction and the mean redshift drift from symbolic regression,” *Physical Review D*, vol. 107, May 2023.
- [6] T. Stephens, “gplearn.” <https://gplearn.readthedocs.io/en/stable/>. Accessed: 2023-12-01.
- [7] M. Virgolin, T. Alderliesten, and P. A. N. Bosman, “Linear scaling with and within semantic backpropagation-based genetic programming for symbolic regression,” *Proceedings of the Genetic and Evolutionary Computation Conference*, pp. 1084–1092, 2019.
- [8] D. Bartlett, H. Desmond, and P. Ferreira, “Priors for symbolic regression,” in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, ACM, July 2023.
- [9] S.-M. Udrescu and M. Tegmark, “Ai feynman: A physics-inspired method for symbolic regression,” *Science Advances*, vol. 6, no. 16, p. eaay2631, 2020.
- [10] S.-M. Udrescu, A. Tan, J. Feng, O. Neto, T. Wu, and M. Tegmark, “Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity,” 2020.
- [11] N. Makke and S. Chawla, “Interpretable scientific discovery with symbolic regression: A review,” 2023.
- [12] W. L. Cava, P. Orzechowski, B. Burlacu, F. O. de França, M. Virgolin, Y. Jin, M. Kommenda, and J. H. Moore, “Contemporary symbolic regression methods and their relative performance,” 2021.
- [13] W. Tenachi, R. Ibata, and F. I. Diakogiannis, “Deep symbolic regression for physics guided by units constraints: toward the automated discovery of physical laws,” 2023.
- [14] P. Lemos, N. Jeffrey, M. Cranmer, S. Ho, and P. Battaglia, “Rediscovering orbital mechanics with machine learning,” 2022.
- [15] D. J. Bartlett, H. Desmond, and P. G. Ferreira, “Exhaustive symbolic regression,” *IEEE Transactions on Evolutionary Computation*, p. 1–1, 2023.
- [16] H. Desmond, D. J. Bartlett, and P. G. Ferreira, “On the functional form of the radial acceleration relation,” *Monthly Notices of the Royal Astronomical Society*, vol. 521, p. 1817–1831, Feb. 2023.
- [17] T. Sousa, D. J. Bartlett, H. Desmond, and P. G. Ferreira, “The simplest inflationary potentials,” 2023.
- [18] A. M. Delgado, D. Wadekar, B. Hadzhiyska, S. Bose, L. Hernquist, and S. Ho, “Modeling the galaxy-halo connection with machine learning,” 2022.
- [19] D. Wadekar, L. Thiele, J. C. Hill, S. Pandey, F. Villaescusa-Navarro, D. N. Spergel, M. Cranmer, D. Nagai, D. Anglés-Alcázar, S. Ho, and L. Hernquist, “The sz flux-mass (y-m) relation at low-halo masses: improvements with symbolic regression and strong constraints on

- baryonic feedback,” *Monthly Notices of the Royal Astronomical Society*, vol. 522, p. 2628–2643, Apr. 2023.
- [20] D. J. Bartlett, L. Kammerer, G. Kronberger, H. Desmond, P. G. Ferreira, B. D. Wandelt, B. Burlacu, D. Alonso, and M. Zennaro, “A precise symbolic emulator of the linear matter power spectrum,” 2023.
 - [21] B. Burlacu, G. Kronberger, and M. Kommenda, “Operon c++: an efficient genetic programming framework for symbolic regression,” *Proceedings of the 2020 Genetic and Evolutionary Computation Conference Companion*, 2020.
 - [22] B. Burlacu, “Gecco’2022 symbolic regression competition: Post-analysis of the operon framework,” in *Proceedings of the Companion Conference on Genetic and Evolutionary Computation*, GECCO ’23 Companion, (New York, NY, USA), p. 2412–2419, Association for Computing Machinery, 2023.
 - [23] J. D. Romano, T. T. Le, W. L. Cava, J. T. Gregg, D. J. Goldberg, N. L. Ray, P. Chakraborty, D. Himmelstein, W. Fu, and J. H. Moore, “Pmlb v1.0: An open source dataset collection for benchmarking machine learning methods,” 2021.
 - [24] R. S. Olson, W. L. Cava, P. Orzechowski, R. J. Urbanowicz, and J. H. Moore, “Pmlb: A large benchmark suite for machine learning evaluation and comparison,” 2017.
 - [25] R. P. Feynman, R. B. Leighton, and M. Sands, *The Feynman lectures on physics; New millennium ed.* New York, NY: Basic Books, 2010. Originally published 1963-1965.
 - [26] J. A. Frieman, M. S. Turner, and D. Huterer, “Dark energy and the accelerating universe,” *Annual Review of Astronomy and Astrophysics*, vol. 46, p. 385–432, Sept. 2008.
 - [27] A. R. Sandage, “The change of redshift and apparent luminosity of galaxies due to the deceleration of selected expanding universes.,” *The Astrophysical Journal*, vol. 136, p. 319, 1962.
 - [28] G. C. Mcvittie, “Appendix to the change of redshift and apparent luminosity of galaxies due to the deceleration of selected expanding universes.,” *The Astrophysical Journal*, vol. 136, p. 334, 1962.
 - [29] H. R. Klöckner, D. Obreschkow, C. Martins, A. Raccanelli, D. Champion, A. Roy, A. Lobanov, J. Wagner, and R. Keller, “Real time cosmology - a direct measure of the expansion rate of the universe,” 2015.
 - [30] Z. Liu, E. Gan, and M. Tegmark, “Seeing is believing: Brain-inspired modular training for mechanistic interpretability,” 2023.
 - [31] T. Buchert and S. Räsänen, “Backreaction in late-time cosmology,” *Annual Review of Nuclear and Particle Science*, vol. 62, p. 57–79, Nov. 2012.
 - [32] C. Clarkson, G. Ellis, J. Larena, and O. Umeh, “Does the growth of structure affect our dynamical models of the universe? the averaging, backreaction, and fitting problems in cosmology,” *Reports on Progress in Physics*, vol. 74, p. 112901, Oct. 2011.
 - [33] G. F. R. Ellis, “Inhomogeneity effects in cosmology,” *Classical and Quantum Gravity*, vol. 28, p. 164001, Aug. 2011.
 - [34] G. F. Ellis and T. Buchert, “The universe seen at different scales,” *Physics Letters A*, vol. 347, p. 38–46, Nov. 2005.
 - [35] T. Buchert, “On average properties of inhomogeneous cosmologies,” 2000.
 - [36] C. Wetterich, “Can structure formation influence the cosmological evolution?,” *Physical Review D*, vol. 67, feb 2003.
 - [37] D. J. Schwarz, “Accelerated expansion without dark energy,” 2002.
 - [38] S. Räsänen, “Dark energy from back-reaction,” *Journal of Cosmology and Astroparticle Physics*, vol. 2004, p. 003–003, Feb. 2004.

- [39] T. Buchert, “On average properties of inhomogeneous fluids in general relativity: Dust cosmologies,” *General Relativity and Gravitation*, vol. 32, p. 105–125, Jan. 2000.
- [40] T. Buchert, “On average properties of inhomogeneous fluids in general relativity: Perfect fluid cosmologies,” *General Relativity and Gravitation*, vol. 33, p. 1381–1405, Aug. 2001.
- [41] T. Buchert, P. Mourier, and X. Roy, “On average properties of inhomogeneous fluids in general relativity iii: general fluid cosmologies,” *General Relativity and Gravitation*, vol. 52, Mar. 2020.
- [42] S. Räsänen, “Light propagation in statistically homogeneous and isotropic dust universes,” *Journal of Cosmology and Astroparticle Physics*, vol. 2009, p. 011–011, Feb. 2009.
- [43] S. Räsänen, “Light propagation in statistically homogeneous and isotropic universes with general matter content,” *Journal of Cosmology and Astroparticle Physics*, vol. 2010, p. 018–018, Mar. 2010.
- [44] S. RÄSÄNEN, “Cosmological acceleration from structure formation,” *International Journal of Modern Physics D*, vol. 15, p. 2141–2146, Dec. 2006.
- [45] S. Räsänen, “Accelerated expansion from structure formation,” *Journal of Cosmology and Astroparticle Physics*, vol. 2006, p. 003–003, Nov. 2006.
- [46] S. M. Koksang, “Another look at redshift drift and the backreaction conjecture,” *Journal of Cosmology and Astroparticle Physics*, vol. 2019, p. 036–036, Oct. 2019.
- [47] S. M. Koksang, “Observations in statistically homogeneous, locally inhomogeneous cosmological toy-models without flrw backgrounds,” 2020.
- [48] A. Heinesen, “Redshift drift as a model independent probe of dark energy,” *Physical Review D*, vol. 103, Apr. 2021.
- [49] S. M. Koksang, “Searching for signals of inhomogeneity using multiple probes of the cosmic expansion rate $H(z)$,” *Physical Review Letters*, vol. 126, June 2021.