Test 10

Kaze W.K Wong[1] **Miles Cranmer**[2]

Abstract

TEST 10. We present an automatic approach to discover analytic population models for gravitational-wave (GW) events from data. As more gravitational-wave (GW) events are detected, flexible models such as Gaussian Mixture Models have become more important in fitting the distribution of GW properties due to their expressivity. However, flexible models come with many parameters that lack physical motivation, making interpreting the implication of these models challenging. In this work, we demonstrate symbolic regression can complement flexible models by distilling the posterior predictive distribution of such flexible models into interpretable analytic expressions. We recover common GW population models such as a power-law-plus-Gaussian, and find a new empirical population model which combines accuracy and simplicity. This demonstrates a strategy to automatically discover interpretable population models in the ever-growing GW catalog, which can potentially be applied to other astrophysical phenomena.

1. Introduction

The rate of gravitational wave (GW) detections is increasing exponentially as the sensitivity of GW detectors improves (Abbott et al., 2021b) since the discovery in 2015 (Abbott et al., 2016). The most update catalog of GW contains 79 events, the majority of them being Binary Black Holes (BBHs). As more GW events are observed, more features in the distribution of GW events await to be discovered. Currently,

one of the most common ways to understand the observed GW population is through phenomenological modeling. For example, to understand the distribution of primary mass in observed BBHs, one would hand-craft a model of the primary mass distribution, such as a power-law distribution (Abbott et al., 2021c). By comparing the model to the data, one can measure the parameters which characterize the model, such as the spectral index.

The main issue with this paradigm is the lack of scalability to a more complex dataset, since construction of sensible phenomenological models requires manual intervention. To illustrate the shortcoming of this paradigm: when 11 events were announced in the first Gravitational-Wave Transient Catalog (GWTC1) in 2019 (Abbott et al., 2019), a power law was sufficient to explain the primary mass distribution. In GWTC2 (which is announced in 2020) (Abbott et al., 2021a), the community modified the simple power law to either a broken power law with two spectral indices, or with the addition of a Gaussian distribution. Finally, the state-of-the-art model used in analyzing GWTC3 (Abbott et al., 2021c) further adds a spline on top of the power-law and Gaussian model to fit the residual (Edelman et al., 2022). Since the complexity of the catalog grows with the size of the catalog, it becomes increasingly difficult to develop an interpretable model that is flexible enough to explain the increasing rich set of features in the catalog .

Alternatively, there exist studies using much more flexible models such as Gaussian Mixture Models (GMMs) to analyze the catalog (Tiwari, 2021). While such a flexible model is powerful in fitting complex datasets, it is difficult to interpret the physical meaning of the fitted parameters. There stand two core reasons why such flexible models are hard to interpret: first, these models often have many parameters; in GWTC3, the GMM has over 100 parameters. With so many degrees of freedom, it is non-trivial to analyze contributions from each mode and attempt to disentangle meaningful patterns. Second, flexible models often employ bases that are not physically motivated. For example, the GMM models has 11 independent Gaussian components in the mass-spin space. While one or two

---

*Equal contribution [1]Center for Computational Astrophysics, Flatiron Institute, New York, NY 10010, USA [2]Department of Astrophysical Sciences, Princeton University, Princeton, New Jersey 08544, USA. Correspondence to: Kaze Wong <kwong@flatironinstitute.org>.
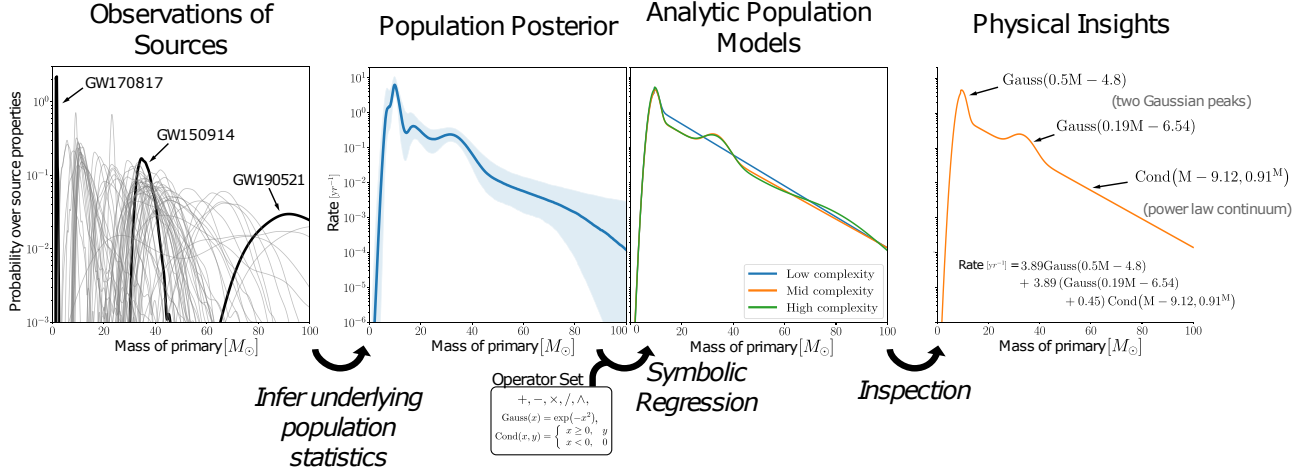
Figure 1. An illustration of our proposed pipeline, which first (*far left*) requires observations of individual source properties (here, events shown from GWTC3); then (*middle left*) infers the deconvolved population posterior density; followed by (*middle right*) using symbolic regression to search for an analytic model which approximates the density, using operators common to existing population models; and, finally (*far right*) distilling physical insights from the recovered population model.

Gaussians in mass or spin space alone can be physically motivated, the additional components are included in the model fit for unexpected features. Therefore, interpreting the importance of each component could be complicated by the components that are not physically motivated.

Symbolic regression (SR) is a machine learning technique which searches to space of analytic expressions to fit a dataset, which complements the weakness of a flexible model in understanding the distribution of GW. However, since SR does not take advantage of using gradient information when searching for expressions, and SR searches a combinatorial space over expressions, it is slow in comparison to traditional machine learning techniques such as Gaussian Mixture Models (GMMs). Therefore, rather than fitting SR directly to the complex and expensive likelihoods found in GW event analysis, we first fit the GMM as a fast-to-evaluate approximation of the true posterior. Then, we run SR to approximate the output of this GMM, which is feasible only because the loss function is then simply mean-squared error against the GMM, rather equal to an evaluation of the raw and expensive likelihood.

In this work, we apply SR to the mixture Gaussian model used to fit the primary mass distribution of GWTC3. Specifically, we approximate the posterior predictive distribution produced of the Gaussian mixture model with SR. This paper is structured as follows: we describe the GW data used in this work in 3. We briefly review SR in section 2. We present our main

result in section 4. We discuss the potential usage and extension of this work in section 5.

2. Method

2.1. Symbolic regression

While most ML algorithms optimize a fixed set number of real parameters, symbolic regression is a machine learning (ML) algorithm which searches for analytic expressions which optimize some loss function. The search typically employs a "genetic algorithm," which arranges symbols in a tree, evaluates the resultant expression, and repeats this process until a particular target is met. Since models for physical phenomena are often specified in a language of concise mathematical expressions, symbolic regression can produce interpretable models compared to traditional ML, as the "trained model" generated by symbolic regression is in the same *language* as existing physical models (i.e., analytic expressions). Therefore, symbolic regression can produce insights for science in a way that traditional ML cannot.

Algorithm. There are several different strategies for performing symbolic regression, all of them offering different advantages. Some methods, such as *SINDy* (Brunton et al., 2016), are dictionary-based, and search for a sparse linear combination of hand-crafted terms. Other methods, such as *eureqa* (Schmidt & Lipson, 2009), are genetic algorithm-based, and, though slower, can find any expression that can be represented

Complexity	Score	Loss	Expression for Rate$[\mathrm{yr}^{-1}](M = M/M_\odot)$
16	0.126	0.212	$3.72\mathrm{Cond}\big(\mathrm{M} - 9.27, 0.9^{\mathrm{M}}\big) + 3.72\mathrm{Gauss}(0.52\mathrm{M} - 4.85)$
25	0.501	0.0352	$3.89\big(\mathrm{Gauss}(0.19\mathrm{M} - 6.54) + 0.45\big)\mathrm{Cond}\big(\mathrm{M} - 9.12, 0.91^{\mathrm{M}}\big)$ $+ 3.89\mathrm{Gauss}(0.5\mathrm{M} - 4.8)$
48	0.00771	0.0108	$\big(\mathrm{Cond}\big(\mathrm{M} - 9.42, 0.62 \cdot 0.9^{\mathrm{M}}\big) + 1.44\mathrm{Gauss}(0.51\mathrm{M} - 4.88)\big)$ $\big(5.11\mathrm{Gauss}(0.06\mathrm{M} - 4.67) + 7.82\mathrm{Gauss}(0.17\mathrm{M} - 5.86) + 3.26\big)$

Table 1. Expressions obtained through symbolic regression with *PySR*. In the search we perform, there are 30 equations with different complexities. We select three representative equations from the Pareto front by setting three successive complexity ranges, and selecting the highest scoring expression in each range.

as a tree of predefined operators, variables, and constants.

We use the algorithm and software *PySR*[1] (Cranmer, 2020) for performing symbolic regression. *PySR* uses a genetic algorithm-style symbolic regression algorithm (see Koza, 1994, for an early example). Loosely inspired by *eureqa*, *PySR* searches the space of trees containing a basis set of specified operators, variables, and arbitrary real constants to find expressions which fit relations in a given dataset. *PySR* uses "regularized evolution" for its search strategy (Real et al., 2019; 2020), which itself is a variant on classic tournament selection (Brindle, 1980; Goldberg & Deb, 1991). This algorithm simulates natural selection: each expression is represented by a tree, corresponding to its "genetic code." Expressions are subsampled, and the fittest expression in each group (using a metric which combines accuracy and simplicity) is used to breed new mutated expressions, as well as cross-bred with other top expressions. *PySR* adds several additional techniques onto these classic algorithms, such as explicit constant optimization using the *BFGS* optimization algorithm (Fletcher, 1988), and an adaptive complexity penalty, while allowing the use of custom operators and losses which is important for this work.

Internal to *PySR*, a multi-objective optimization problem is solved: the evolution seeks to find expressions which are as accurate as possible, while being as simple as possible. Formally, for our setting, *PySR* solves the following optimization problem:

$$s_c = \underset{\substack{s \in \mathcal{S} \\ |s| = c}}{\arg\min} \mathcal{L}(s), \quad \text{for each } c \in \mathcal{C}, \qquad (1)$$

for the loss function $\mathcal{L}(s) = \sum_i \ell(s(\mathbf{x}_i), y_i)$, where s is an expression, $\mathcal{S}$ is a space of possible expressions considered, $|s|$ is the complexity of an expression s, ℓ is the loss function, and $\{(x_i, y_i)\}$ is the dataset for x_i each input and y_i the label. The most accurate

expression at each c, denoted as s_c, is computed for a given set range of c, which we choose as $c \in \mathcal{C} \equiv \{1, \ldots, 50\}$. Complexity is defined here as the number of symbols in the expression (including repeats), and ℓ as $\ell(\hat{y}_i, y_i) = (\hat{y}_i - y_i)^2$.

Next, given the set of the best expressions at each complexity $\mathcal{S}^* = \{s_c\}_{c \in \{1, \ldots, 80\}}$, Next, we can construct the *Pareto front*:

$$\hat{\mathcal{S}}^* = \{s_c \mid \mathcal{L}(s_c) < \mathcal{L}(s_q) \forall q < c\}, \qquad (2)$$

such that each remaining expression is more accurate than all simpler expressions. Using this, we can define the *score* of an expression as the slope of the Pareto front, equal to: $\text{score} = -\Delta(\log \mathcal{L}(s_c))/\Delta c$, which is a heuristic for the quality of a particular equation, used in the symbolic regression literature (see, e.g., Schmidt & Lipson, 2009; Cranmer et al., 2020).

Implementation for GWs Symbolic regression is an automated type of empirical model discovery—something which has driven the development of many theories in the physical sciences. For example, Kepler's laws were an empirical model crafted by hand to fit patterns in the orbits of planets, and ultimately inspired Newton to discover a formula for gravitational force which could explain Kepler's laws (see Hawking, 2004, for a review of this history). Astronomy has seen increasing use of symbolic regression to find new empirical models in an automated way (e.g., Graham et al., 2013; Cranmer et al., 2020; Wadekar et al., 2020; Delgado et al., 2021; Cranmer et al., 2021; Shao et al., 2021).

Here, we use symbolic regression for the first time to learn population density models in astrophysics. One of the main advantages of symbolic regression is to learn a model into an interpretable language, and, therefore, we define a set of operators which frequently occur in GW population models: addition, subtraction, division, power laws, square, Gaussians (unary operator: $f(x) = \exp(-x^2)$), and conditional state-

ments (binary operator: $g(x, y) = \text{if}(x \geq 0, y, 0)$). To simplify our assumptions, all operators are assigned the same complexity of 1. We also introduce additional constraints on how operators may be nested inside each other to rule out uninterpretable expressions; these are defined in Appendix A.

3. Gravitational-Wave Data

GW data comes in the form of time series, in which each time sample carries very little physical meaning. It is easier to understand the physical implication of a GW event through its inferred source properties (Veitch et al., 2015). In the case of a BBH, the source properties include the masses of the two merging black holes. On top of extracting physics from individual events, we can also learn about astrophysical processes through understanding of the distribution of source properties, such as the mass distribution of black holes (Thrane & Talbot, 2019; Vitale et al., 2020).

In principle, one can apply symbolic regression to fit these distributions directly. Unfortunately, GW observations are extremely noisy. As a result, all inferred properties about the source such as the observed masses or distance to the source come with sizable uncertainty, often comparable to the prior used in inferring their source properties. This also means the posterior distribution of these inferred properties often has a non-trivial shape that the typical Gaussian approximation of "error bar'' does not hold. Moreover, since there are only 79 announced events, the shot noise makes it challenging to extract meaningful information if we first stack the events and fit the distribution with symbolic regression.

In order to maximally extract information from the GW catalog, the community has employed Hierarchical Bayesian Analysis (HBA) to analyze the distribution of GW events. HBA takes advantage of the fact that measurement uncertainties for each event are conditionally independent from the population model, so we can deconvolve the measurement uncertainty for each event (Bovy et al. 2011 describes this as extreme deconvolution) while fitting for the underlying distribution of GW events[2].

One product coming out of this HBA pipeline is the posterior predictive distribution for the properties of interest, i.e., a set of probability density functions that encapsulate the inferred source distribution and the

uncertainty on the population level. It is much easier to find an interpretable form of the source distribution by fitting the posterior predictive distribution with SR. In this study, we focus on understanding the distribution of binary black hole primary mass, i.e., the heavier mass in the binary. One of the most flexible models that is fitted to this set of data is a Gaussian mixture model (Tiwari, 2021). To generate the data that we feed to the SR pipeline, we select many evaluation points in the primary mass axis, then we take the median of the posterior predictive distribution given by the Gaussian mixture model as the value we are trying to predict. We also estimate the uncertainty of the value at each evaluation point using the 68% confidence interval.

4. Result

We show the workflow of this work in figure 1. On the leftmost panel, we show the posterior density in primary mass for every event in GWTC3. In the left-middle panel, we show a GMM fitted to the data given by (Tiwari, 2021). Given the posterior predictive distribution shown in the middle panel, we use symbolic regression to extract effective descriptions of the data. As we increase the complexity of the equation, it is natural to expect that the equation should fit the data better than a lower complexity equation since it has additional free parameters. At the same time, an equation with higher complexity could either be overfitting or difficult to interpret. *PySR* uses a score (see Cranmer et al. 2020 for details) as a metric normalized to complexity to show how well an equation is fitting the data. At each complexity, *PySR* outputs the equation with the lowest loss. To select the equations shown in figure 1, we first make an accuracy cut based on the loss of the function such that all expressions have a loss lower than 0.3, so the selected equation would still be a good fit to the data. We then consider the expressions with the highest score among three ranges: $c \in [1, 20]$, $c \in [20, 35]$, and $c \in [35, 50]$. These ranges are set to give us a variety of complexity expressions to study. The equations selected under these criteria are tabulated in table 1.

The lowest complexity (16) equation returned is a decaying exponential function with a Gaussian bump at the lower end of the mass distribution. This agrees with a recent study pointing out there is a very strong excess of events at $m_1 \sim 10 \ M_\odot$, which corresponds to a smooth turn-on of the black hole mass spectrum (Talbot & Thrane, 2018). In order to account for the bump around 30 $M_\odot$, the exponential part of this equation has a much higher normalization com-

---

[2]For interested readers, (Mandel et al., 2019; Vitale et al., 2020; Gaebel et al., 2019) are excellent references with practical examples on how to apply the HBA framework to GW data.

pared to the others. This agrees with the analysis done in GWTC3, where a single power law seems to have trouble accounting for the excess of events around $m_1 \sim 30\ M_\odot$(Abbott et al., 2021c).

Compared to the lowest complexity equation, the middle complexity (25) equation instead prefers a lower normalization of the exponential function, but account for the bump around 30 $M_\odot$ with a Gaussian component, which is proposed to be driven by a pile-up from pulsational pair-instability supernovae (Talbot & Thrane, 2018). This equation is similar to one of the simplest phenomenological models fitted to GWTC3, which is a power law mixed with a Gaussian bump. The difference between this equation and the one presented in GWTC3 is this equation has an additional bump at $m_1 \sim 10\ M_\odot$, which is also present in the lower complexity equation. This suggests the first bump at lower mass has more statistical significance than the second bump. Although the middle complexity equation is more complex, the loss is about one order of magnitude lower than the next best one with lower complexity (16). As a result, the score of this equation is higher than the lower complexity equation, making it a favorable equation.

At the highest complexity (48), one more Gaussian component is added to the equation account for the tapering at the higher mass end. Interestingly, the new component added in this equation does not try to fit the wiggles in between the first bump and the second bump. The significance of any features other than the two bumps and the continuum is still an open question. Our result hints the wiggles between the bumps seem to be less significant than the extra tapering of the continuum. In addition, none of the equations prefers to fit the extra hump before the 10 $M_\odot$ peak. This could mean the little hump could be an artifact in the Gaussian mixture model instead of a truly statistically significant feature. However, the loss of this equation is similar to the middle complexity (25) equation, at the same time being almost twice as complex. As a result, the score is much lower, which means the tapering is not necessarily needed.

5. Discussion

In this paper, we use symbolic regression to extract interpretable equations from the posterior predictive distribution of a flexible model fitted to the catalog of GWs. We show SR can discover equations which are similar to state-of-the-art phenomenological models. On top of finding forms that have already been discovered, our result also shed insights on an open question—whether there exist additional features other than the two prominent peaks in the primary mass distribution (Tiwari, 2022). Since the highest complexity equation does not significantly improve the fitting accuracy, one should take caution when interpreting the tapering of the continuum or wiggles between the two peaks as their significance are unclear.

The main message of this work is that by combining flexible models with symbolic regression, we can discover interpretable equations that can describe the increasingly complex GW catalog. Compared to traditional methods such as phenomenological modeling, which relies on an individual scientist's insight, our pipeline requires minimal human intervention. This technique is not a model selection method by itself, but a pipeline to discover models for downstream tasks such as model selection and parameter estimation. Once we obtain a set of equations, we can insert the expressions to a standard HBA pipeline to estimate the parameters of the model and select between models. This allows the community to discover interpretable patterns in the GW catalog in a data-driven instead of hand-crafted model-driven manner. In this sense, our method is more scalable to more complex and higher-dimensional datasets.

In this work, we focus on analyzing only the primary mass distribution. Our method is trivially generalizable to higher-dimensional data, such as the joint space of both the primary mass and the secondary mass. So far there are relatively few phenomenological models which can capture correlations between observables, mainly due to the difficulty of manually constructing reasonable models to account for these correlations. At the same time, the most advanced flexible model in the literature is an n-dimensional histogram, which is hard to interpret. Our method provides a natural way to explore and to explain these correlations.

Although we focus on GW in this work, this method can be easily extended to other population analysis problems. For example, given the recently released Gaia DR3 dataset (Collaboration, 2016), it might be possible to apply this method to understanding various subpopulations of stars in the Milky Way. As size of datasets grows larger in astronomy, this work provides a simple and scalable pathway to uncover trends and relations in astrophysical catalogs.

References

Abbott, B. P. et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Phys. Rev. Lett.*, 116(6):061102, 2016. doi: 10.1103/ PhysRevLett.116.061102.

Abbott, B. P. et al. GWTC-1: A Gravitational-Wave Transient Catalog of Compact Binary Mergers Observed by LIGO and Virgo during the First and Second Observing Runs. *Phys. Rev. X*, 9(3):031040, 2019. doi: 10.1103/PhysRevX.9.031040.

Abbott, R. et al. Population Properties of Compact Objects from the Second LIGO-Virgo Gravitational-Wave Transient Catalog. *Astrophys. J. Lett.*, 913(1): L7, 2021a. doi: 10.3847/2041-8213/abe949.

Abbott, R. et al. GWTC-3: Compact Binary Coalescences Observed by LIGO and Virgo During the Second Part of the Third Observing Run. 11 2021b.

Abbott, R. et al. The population of merging compact binaries inferred using gravitational waves through GWTC-3. 11 2021c.

Bovy, J., Hogg, D. W., and Roweis, S. T. Extreme deconvolution: Inferring complete distribution functions from noisy, heterogeneous and incomplete observations. *The Annals of Applied Statistics*, 5(2B): 1657 – 1677, 2011. doi: 10.1214/10-AOAS439. URL https://doi.org/10.1214/10-AOAS439.

Brindle, A. *Genetic Algorithms for Function Optimization*. PhD thesis, 1980.

Brunton, S. L., Proctor, J. L., and Kutz, J. N. Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences*, 113 (15):3932–3937, 2016. ISSN 0027-8424. doi: 10.1073/pnas.1517384113.

Collaboration, G. The gaia mission. 2016.

Cranmer, M. PySR: Fast & Parallelized Symbolic Regression in Python/Julia. Zenodo, September 2020.

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. Discovering Symbolic Models from Deep Learning with Inductive Biases. *NeurIPS*, June 2020.

Cranmer, M., Kreisch, C., Pisani, A., Villaescusa-Navarro, F., Spergel, D. N., and Ho, S. Histogram Pooling Operators: An Interpretable Alternative for Deep Sets. *ICLR 2021 SimDL Workshop*, pp. 10, June 2021.

Delgado, A. M., Wadekar, D., Hadzhiyska, B., Bose, S., Hernquist, L., and Ho, S. Modeling the galaxy-halo connection with machine learning. November 2021.

Edelman, B., Doctor, Z., Godfrey, J., and Farr, B. Ain't No Mountain High Enough: Semiparametric Modeling of LIGO–Virgo's Binary Black Hole Mass Distribution. *Astrophys. J.*, 924(2):101, 2022. doi: 10.3847/1538-4357/ac3667.

Fletcher, R. Practical Methods of Optimization. 1988. doi: 10.2307/2008742.

Gaebel, S. M., Veitch, J., Dent, T., and Farr, W. M. Digging the population of compact binary mergers out of the noise. MNRAS, 484(3):4008–4023, April 2019. doi: 10.1093/mnras/stz225.

Goldberg, D. E. and Deb, K. A Comparative Analysis of Selection Schemes Used in Genetic Algorithms. In Rawlins, G. J. E. (ed.), *Foundations of Genetic Algorithms*, volume 1, pp. 69–93. Elsevier, January 1991. doi: 10.1016/B978-0-08-050684-5.50008-2.

Graham, M. J., Djorgovski, S. G., Mahabal, A. A., Donalek, C., and Drake, A. J. Machine-assisted discovery of relationships in astronomy. 431(3):2371–2384, May 2013. doi: 10.1093/mnras/stt329.

Hawking, S. *On the Shoulders of Giants: The Great Works of Physics and Astronomy*. Running, Philadelphia, Pa.; London, 2004. ISBN 978-0-7624-1898-5.

Koza, J. R. Genetic programming as a means for programming computers by natural selection. *Statistics and Computing*, 4(2):87–112, June 1994. ISSN 1573-1375. doi: 10.1007/BF00175355.

Mandel, I., Farr, W. M., and Gair, J. R. Extracting distribution parameters from multiple uncertain observations with selection biases. *Mon. Not. Roy. Astron. Soc.*, 486(1):1086–1093, 2019. doi: 10.1093/mnras/stz896.

Real, E., Aggarwal, A., Huang, Y., and Le, Q. V. Regularized Evolution for Image Classifier Architecture Search. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):4780–4789, July 2019. ISSN 2374-3468. doi: 10.1609/aaai.v33i01.33014780.

Real, E., Liang, C., So, D., and Le, Q. AutoML-Zero: Evolving Machine Learning Algorithms From Scratch. In *International Conference on Machine Learning*, pp. 8007–8019. PMLR, November 2020.

Schmidt, M. and Lipson, H. Distilling Free-Form Natural Laws from Experimental Data. *Science*, 324 (5923):81–85, April 2009. ISSN 0036-8075, 1095-9203. doi: 10.1126/science.1165893.

Shao, H., Villaescusa-Navarro, F., Genel, S., Spergel, D. N., Angles-Alcazar, D., Hernquist, L., Dave, R., Narayanan, D., Contardo, G., and Vogelsberger, M. Finding universal relations in subhalo properties with artificial intelligence. September 2021.

Talbot, C. and Thrane, E. Measuring the binary black hole mass spectrum with an astrophysically motivated parameterization. *Astrophys. J.*, 856(2):173, 2018. doi: 10.3847/1538-4357/aab34c.

Thrane, E. and Talbot, C. An introduction to Bayesian inference in gravitational-wave astronomy: Parameter estimation, model selection, and hierarchical models. PASA, 36:e010, March 2019. doi: 10.1017/pasa.2019.2.

Tiwari, V. VAMANA: modeling binary black hole population with minimal assumptions. *Class. Quant. Grav.*, 38(15):155007, 2021. doi: 10.1088/1361-6382/ac0b54.

Tiwari, V. Exploring Features in the Binary Black Hole Population. *Astrophys. J.*, 928(2):155, 2022. doi: 10.3847/1538-4357/ac589a.

Veitch, J. et al. Parameter estimation for compact binaries with ground-based gravitational-wave observations using the LALInference software library. *Phys. Rev. D*, 91(4):042003, 2015. doi: 10.1103/PhysRevD.91.042003.

Vitale, S., Gerosa, D., Farr, W. M., and Taylor, S. R. Inferring the properties of a population of compact binaries in presence of selection effects. 7 2020. doi: 10.1007/978-981-15-4702-7_45-1.

Wadekar, D., Villaescusa-Navarro, F., Ho, S., and Perreault-Levasseur, L. Modeling assembly bias with machine learning and symbolic regression. *arXiv e-prints*, art. arXiv:2012.00111, November 2020.

A. Operator Constraints

With full flexibility over how symbolic operators can be arranged up to a maximum number, surprisingly many combinations exist which are simple and accurate, yet completely uninterpretable. Typically, a genetic algorithm might find that a particular deeply nested combination of operators might perform very well, with the overall complexity being low. However, we wish to find a model that is interpretable, and so we provide some additional constraints here.

First, we implement limits on which operators can be nested, and how deeply. We impose the following nested constraints:

1. No operator with arity of 1 (i.e., an operator taking a single argument) may be nested with another arity-1 operator, unless the operator is square.

   (a) However, we permit a square to be nested inside itself once, as such a combination may be used to make a quadratic operation with few operations.

2. The division operator may only be nested inside itself once.

3. The power operator may not contain any arity-1 operator.

We also impose the following constraints on maximum complexity within each operator:

1. The power operator may only have a maximum complexity of 9 in its base, and 3 in its exponent.

2. The division operator may have any complexity in the numerator, but a maximum complexity of 7 in the denominator.

3. The Gauss operator may only have a maximum complexity of 11 in its argument.

4. The conditional operator may only have a maximum complexity of 5 in its condition, and 11 in its resultant argument.

These prevent *PySR* from finding expressions which, despite being low-complexity and high-accuracy, are still uninterpretable to a human scientist due to unusual combinations of operators compared to typical expressions found throughout physics and astrophysics.