SF2A 2023
M. N'Diaye, A. Siebert, N. Lagarde, O. Venot, K. Baillié, M. Béthermin, E. Lagadec, J. Malzac, J. Richard (eds)

SYMBOLIC REGRESSION DRIVEN BY DIMENSIONAL ANALYSIS FOR THE AUTOMATED DISCOVERY OF PHYSICAL LAWS AND CONSTANTS OF NATURE

W. Tenachi[1], R. Ibata[1] and F. I. Diakogiannis[2]

Abstract. Given the abundance of empirical laws in astrophysics, the rise of agnostic and automatic methods to derive them from data is of great interest. This concept is embodied in symbolic regression, which seeks to identify the best functional form fitting a dataset. Here we present a protocol for deducing both physical laws but also the constants of nature appearing in those with their associated units. Our method is grounded in the Physical Symbolic Optimization framework, which integrates dimensional analysis with deep reinforcement learning. We showcase our approach on a panel of equations from (astro)-physics.

Keywords: Symbolic regression, Dimensional Analysis, Reinforcement learning

1 Introduction

Physical theories, including astrophysical ones, often originate from empirical laws. Typically, physicists observe phenomena, derive empirical laws, and then formulate overarching theories that encompass these laws, e.g., Newton's law of universal gravitation Newton (1687) which describes both the movement of objects on Earth and the movement of planets themselves. However, with the rise of deep learning, many empirical laws are now represented as neural networks, complicating their integration into broader theories. While neural networks are indispensable for modeling complex or high-dimensional data, such as images, there are instances where these models can be distilled into simpler analytical expressions (see e.g., Tenachi et al. (2023a)), bridging the gap between empirical laws and theoretical understanding.

Our research project aims to build a method for agnostically bridging this gap. In Section 2, we briefly summarize how we employ dimensional analysis alongside reinforcement learning to derive physical analytical expressions through our Physical Symbolic Optimization framework detailed in Tenachi et al. (2023b). Subsequently, in Section 3, we introduce a protocol to simultaneously identify concise physical laws and the constants that should appear in those with appropriate units such that the resulting law is accurate, succinct, and adheres to dimensional analysis principles.

2 Recovering physical laws

Symbolic regression (SR) consists in the inference of a free-form symbolic analytical function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ that fits $y = f(\mathbf{x})$ given $(\mathbf{x}, y)$ data. It is distinct from numerical parameter optimization procedures in that it consists in a search in the space of functional forms themselves by optimizing the arrangement of mathematical symbols (e.g. x, $+$, $-$, $\times$, $/$, sin, exp, log, ...).

Here we adopt the Physical Symbolic Optimization (Φ-SO) framework detailed in Tenachi et al. (2023b) which was built from the ground up for physics. In this framework, the search space is reduced by leveraging physical units constraints (e.g., cos is dimensionless, $\text{velocity} + \text{length}/\square \iff \square$ is a time etc.) and proposing physically meaningful expressions only. Φ-SO relies on a recurrent neural network (RNN) to generate multiple trial analytical expressions. Fit quality of these expressions can then be assessed against data. The best expressions are then reinforced and the process is repeated until the RNN converges and yields a set of high quality expressions.

[1] Université de Strasbourg, CNRS, Observatoire astronomique de Strasbourg, UMR 7550, F-67000 Strasbourg, France
[2] Data61, CSIRO, Kensington, WA 6155, Australia

Φ-SO was evaluated using a standard benchmark of 120 equations from the Feynman Lectures on Physics (Feynman et al. 1971) following the `SRBench` procedure (La Cava et al. 2021), with the objective of accurately recovering the expressions from their associated datasets. It achieved state-of-the-art performance in the presence of noise (exceeding 0.1%) and showed robust performances even in the presence of substantial (10%) noise.

3 Composing natural constants

For new scientific discovery, there are instances where the appropriate free parameters and their corresponding units are not immediately evident. In such situations, we propose a protocol wherein Φ-SO is allowed one free parameter for each input variable, sharing the same units, and another free parameter reflecting the units of the output variable. Specifically, for an SR problem consisting in the deduction of y from $\{x_1, ...x_n\}$, we would permit the inclusion of $\{\theta_y, \theta_{x_1}, ...\theta_{x_n}\}$ as free constants. This grants Φ-SO the flexibility to selectively combine or omit these free parameters to construct new parameters that align with dimensional analysis constraints. In this setup, we demonstrate that Φ-SO can adeptly resolve the SR challenges outlined in Table 1, yielding both the precise symbolic expressions and their corresponding physical constants with accurate units.

Case	Expression
Ideal Gas Law	$P = \frac{nRT}{V}$
Free Fall Terminal Velocity	$v_t = \sqrt{\frac{2mg}{\rho A C_d}}$
Classical Gravity	$F = \frac{Gm_1 m_2}{r^2}$
Black Body Photon Count	$n = 1/(e^{\frac{h\nu}{k_b T}} - 1)$

Table 1. Target expressions. Input variables are colored in red with recovered natural constants to recover left in black.

For illustration, Φ-SO successfully derives the equation describing the equation of state of an ideal gas $P = C\frac{nT}{V}$ with $C = \frac{\theta_P \theta_V}{\theta_n \theta_T}$ having units $M.L^2.T^{-2}.K^{-1}.N^{-1}$ effectively rediscovering the ideal gas constant usually denoted by R. Similarly, Φ-SO is able to recover the expression for the terminal velocity of a free falling object as a function of its mass m, its surface area A and the density of the medium it traverses ρ as $v_t = \sqrt{C\frac{m}{\rho A}}$ by unveiling its proportionality to the square root of an acceleration $\sqrt{C}$, formulated by Φ-SO as $\sqrt{\theta_{v_t}/\sqrt{\theta_A}}$, corresponding to the Earth surface gravity $\sqrt{g}$ and other scale factors.

Interestingly, while the discussion on interpretability often revolves around the agnostic recovery of functional forms (as in our main study Tenachi et al. (2023b)), incorporating dimensional analysis introduces fresh perspectives. Specifically, it offers a method to agnostically determine the most straight forward units of the degrees of freedom such that these units are the simplest combination of the problem's natural units, and that the resulting expression is not only predictive but also physically sound and concise.

Code availability

The documented code for the Φ-SO algorithm along with demonstration notebooks are available on GitHub github.com/WassimTenachi/PhySO.

RI acknowledges funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 834148).

References

Feynman, R. P., Leighton, R. B., Sands, M., et al. 1971, The Feynman lectures on physics, Vol. 1-3 (Addison-Wesley Reading, MA)

La Cava, W., Orzechowski, P., Burlacu, B., et al. 2021, in Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks, ed. J. Vanschoren & S. Yeung, Vol. 1 (Curran)

Newton, I. 1687, Philosophiae Naturalis Principia Mathematica

Tenachi, W., Ibata, R., & Diakogiannis, F. I. 2023a, arXiv e-prints, arXiv:2305.16845

Tenachi, W., Ibata, R., & Diakogiannis, F. I. 2023b, arXiv e-prints, arXiv:2303.03192