SYMBOLIC REGRESSION FOR PDES USING PRUNED DIFFERENTIABLE PROGRAMS

Ritam Majumdar, Vishal Jadhav & Anirudh Deodhar
Tata Consultancy Services Research, India
{ritam.majumdar,vishal.jadhav,anirudh.deodhar}@tcs.com

Shirish Karande, Lovekesh Vig & Venkataramana Runkana
Tata Consultancy Services Research, India
{shirish.karande,lovekesh.vig,venkat.runkana}@tcs.com

ABSTRACT

Physics-informed Neural Networks (PINNs) have been widely used to obtain accurate neural surrogates for a system of Partial Differential Equations (PDE). One of the major limitations of PINNs is that the neural solutions are challenging to interpret, and are often treated as black-box solvers. While Symbolic Regression (SR) has been studied extensively, very few works exist which generate analytical expressions to directly perform SR for a system of PDEs. In this work, we introduce an end-to-end framework for obtaining mathematical expressions for solutions of PDEs. We use a trained PINN to generate a dataset, upon which we perform SR. We use a Differentiable Program Architecture (DPA) defined using context-free grammar to describe the space of symbolic expressions. We improve the interpretability by pruning the DPA in a depth-first manner using the magnitude of weights as our heuristic. On average, we observe a 95.3% reduction in parameters of DPA while maintaining accuracy at par with PINNs. Furthermore, on an average, pruning improves the accuracy of DPA by 7.81% . We demonstrate our framework outperforms the existing state-of-the-art SR solvers on systems of complex PDEs like Navier-Stokes: Kovasznay flow and Taylor-Green Vortex flow. Furthermore, we produce analytical expressions for a complex industrial use-case of an Air-Preheater, without suffering from performance loss viz-a-viz PINNs.

1 INTRODUCTION

Symbolic Regression is the task of generating a mathematical expression that best fits a given dataset. SR is an important problem as it helps understand underlying relationships and patterns in data, with application in scientific discovery [1, 2], engineering design [3, 4], and financial forecasting [5], just to name a few. SR helps reduce the complexity of the models and provides interpretable solutions, thereby improving the transparency and accountability of AI systems. Recently, Virgolin et al. in [6] proved SR to be an NP-hard problem. Historically, SR has been attempted using genetic programming methods [7], purely Deep-learning methods like sequence generation [8, 9], tree search [10, 11, 12], and a combination of both Deep-learning and Genetic programming methods [13, 14].

While SR has been applied for PDE equation discovery using Genetic Programming [15], Fast-function extraction [3], replacing activation functions of NNs with primitive functions [16], sequence to sequence equation generation using Transformers [17], very few works [18, 19] attempts to directly model the final analytical solution of the governing PDE. Inspired by [20] which generates differentiable programs, Majumdar et al. in [19] introduced Physics Informed Symbolic Networks (PISN) to generate analytical expressions for PDEs. Given context-free grammar, they approximate a production rule by taking a linear weighted approximation of the rules. While PISNs performed on par with PINNs, the analytical expressions generated were large and weren't interpretable. In this work, we use the original differentiable program architecture (DPA) in [20] for performing symbolic regression over generated data points by PINNs. We improve the transparency of the sym-

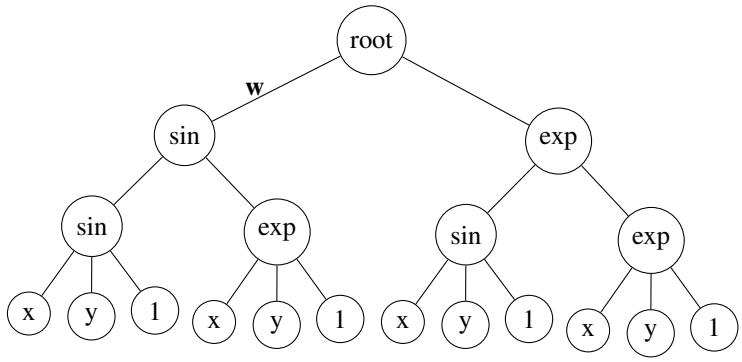

Figure 1: Program Derivation Graph of a Grammar with operators: sin, exp, x, y, c of Depth 2

bolic expressions by pruning the DPA in a depth-first manner, using the magnitude of weights as the heuristic. Pruning allows us to obtain sparser representations for PDEs that are easily interpretable.

Our key contributions are as follows: 1) We use a DPA to perform Symbolic Regression on PDEs. 2) Our pruning strategy reduces 95.3% of the parameters of the program architecture with a performance at par with PINNs. 3) Our framework demonstrates excellent performance on complex PDEs like Navier Stokes and industrial systems like Air-preheater [21, 22] which have no predefined analytical solution.

The rest of the paper is organized as follows. Section 2 consists of Methodology, followed by Observations and Discussions. Section 4 consists of the limitations. The Appendix is organized as follows. Section A.1 consists of PDE-details, followed by Experiment details. Section A.3 consists of training schedule, followed by generated symbolic expressions and a comparison between pruned and unpruned DPA. Finally, we conclude by providing a visualization of the proposed pruning algorithm.

2 METHODOLOGY

Algorithm 1 Symbolic Regression for Partial Differential Equation

1: Train a PDE solver to solve for the PDE
2: Generate input-output data points using the learned PDE solver
3: Perform Regression on generated data points using Differentiable Program Architecture
4: Prune the Differentiable Program Architecture

Algorithm 1 describes our end-to-end procedure for performing symbolic regression on Partial Differential Equations. In the first step, we use a PDE solver to solve for the given PDE setup. In this work, we use PINNs as our PDE solver. The PINN solver can be replaced by any numerical PDE solver suitable for the problem at hand. The second step involves preparing the dataset by generating input-output data points using the trained PINN. Symbolic regression is then performed using a Differentiable Program Architecture defined based on the context-free grammar [23] described in Equation 1. We take sin, exp, log, power 2, and power 3 as our unary operators, and Addition and Multiplication as our binary operators. Figure 1 provides an example of expanding the differentiable program architecture till depth 2 using sin, exp, and leaf nodes as operators. x,y,t,c are the terminal symbols. Finally, we prune the DPA as described in Algorithm 2.

$$\alpha ::= \textbf{sin}\ \alpha_1 \mid \textbf{exp}\ \alpha_1 \mid \textbf{log}\ \alpha_1 \mid \textbf{pow2}\ \alpha_1 \mid \textbf{pow3}\ \alpha_1 \mid \textbf{Add}\ \alpha_1\ \alpha_2 \mid \textbf{Multiply}\ \alpha_1\ \alpha_2 \mid x \mid y \mid t \mid c \quad (1)$$

Algorithm 2 represents our pruning strategy for DPA. We initialize the final DPA with unpruned DPA weights and loss as the mean-relative-L2-error on data points generated using the PDE solver. We postulate the importance of a term in the mathematical expression is directly proportional to the magnitude of the edge. Thus, we recursively visit every child of a node starting with the child having

Algorithm 2 Depth First Search Pruning Strategy of Differentiable Program Architecture (DPA)

Global initialize $W \leftarrow$ Unpruned DPA weights, loss $\leftarrow$ SCORE (W)
Function DFS $(node)$
Initialize $visited \leftarrow node$, $children \leftarrow$ children of $node$ sorted by absolute value
if $children$ is $None$ **then**
    $W^{'} \leftarrow$ PRUNE $(node)$
    $W^{'} \leftarrow$ FINETUNE $(W^{'})$
    finetuned-loss $\leftarrow$ SCORE $(W^{'})$
    if finetuned-loss $\leq$ loss **then**
        loss $\leftarrow$ finetuned-loss
        $W \leftarrow W^{'}$
    end if
    Return W,loss
end if
for all $child$ in $children$ **do**
    if $child$ not in $visited$ **then**
        DFS($child$)
    end if
end for

a minimum value, in a depth-first manner. On encountering a leaf, we prune that weight and finetune the DPA. If the resulting DPA performs on par or better, we accept the prune, else we reset the prune and move to the next child. We perform this operation recursively until all nodes are visited.

3 OBSERVATIONS AND DISCUSSION

Table 1 compares the performance of the output variables of interest. We compare the performance of pruned DPA with the trained PINN, unpruned DPA, and our benchmarks, AI-Feynmann (AIF), SymbolicGPT (SGPT), and Deep Symbolic Regression (DSR). Except for Air-Preheater (APH), we compare the relative-L2 error between the solutions generated with their corresponding true analytical solution. In the case of APH, true analytical solutions don't exist, and we use numerical simulations by Finite Difference Method to generate the ground-truth temperature distribution [24].

		PINN	DPA-Unpruned	DPA-Pruned	AIF	SGPT	DSR
Diffusion	u	7.32e-3	8.54e-3	**8.16e-3**	**8.00e-4**	0.54	1.16
Kovasznay	u	**9.64e-3**	1.15e-2	**9.85e-3**	4.50e-1	0.53	0.64
	v	**1.44e-2**	1.56e-2	**1.47e-2**	4.80e-1	0.55	0.54
	p	1.75e-2	2.09e-2	**1.77e-2**	**1.69e-2**	0.75	1.33
Taylor Green	u	**3.02e-2**	3.71e-2	**3.08e-2**	5.60e-1	0.59	1.52
	v	**2.73e-2**	3.59e-2	**2.76e-2**	6.70e-1	0.62	1.83
	p	**3.64e-2**	4.66e-2	**3.75e-2**	7.40e-1	0.76	0.91
Diffusion Reaction	u	**1.34e-2**	1.68e-2	**1.64e-2**	3.50e-1	0.56	0.82
APH	T_{fg}	**2.03**	2.14	**2.06**	31.93	35.21	22.45
	T_{mg}	**2.53**	2.55	**2.54**	34.23	38.24	31.27
	T_{fa_1}	**3.08**	3.27	**3.09**	28.45	46.65	44.48
	T_{ma_1}	**2.81**	3.02	**2.85**	47.71	41.24	49.67
	T_{fa_2}	**2.97**	3.10	**3.00**	39.64	37.41	45.21
	T_{ma_2}	**3.02**	3.08	**3.04**	38.26	41.69	37.73

Table 1: Performance comparison of Differentiable program architecture with existing benchmarks. In every row, bold denotes the Top-2 best performing methods.

Across all tasks, Unpruned-DPA is slightly worse than PINNs on the same training samples, in spite of the best expression lying in the defined CFG. The reason is, NNs have one standard activation throughout, while DPA has multiple operators with unique convergence characteristics. For

example, Sine introduces periodicity, leading to the gradients being periodic, exponential increases rapidly for larger values and log decreases sharply for very small values. Currently, popular optimizers like Adam have difficulty in converging all the operators simultaneously [25], as they all have different convergence rates. Loss-curve characteristics of DPA need to be studied and custom optimizers have to be developed further. Across all examples, we observe pruning to improve the accuracy of DPA. The accuracy boost happens because pruning of weights reduces overfitting occurring due to complex expressions at higher depths, leading to better generalization. From table 3, we observe on an average, 95.3% reduction in # weights of DPA after pruning. Furthermore, pruned DPA comprehensively outperforms the benchmarks. Across the 14 output variables in 5 PDEs, AIF is marginally better on just 2 variables whose underlying mathematical expressions are Depth-1 expressions and are easy to capture. The other 12 variables have expressions with higher depths which the benchmarks struggle to recapture. Failure of benchmarks on expressions with higher depth highlights the superior representation capacity of program architecture.

Table 4 and 5 represent the expressions obtained by Pruned-DPA and benchmark methods respectively. On careful observation, we notice SGPT and DSR to be biased towards the logarithm and nested sin and cos operators. AIF produces over-simplified expressions, hence struggles to find a good fit to datapoints. We provide an example in Appendix A.5 of unpruned-DPA for the Diffusion equation. Our pruning heuristics allows us to prune 95% of the weights and bring down 20 line expressions to 1-2 lines, drastically improving their explainability. Furthermore, for complex examples like Kovasznay flow, pruned-DPA expressions come very close to the ground-truth, and equivalencies can be proved. u_{true} and u differ in $cos(2\pi y)$ and $sin(6.28y - 1.57)$ terms, and it's well known from trignometric identities relation, $cos(2\pi y) = sin(2\pi y - pi/2)$. v_{true} and v differ in $exp(\lambda x)/2\pi$ and $(0.29 - 0.54x + 0.46x^2 - 0.28x^3 + 0.12x^4)$, where one can verify, the differed expression is in fact the Taylor-series approximation of the ground-truth upto the 4th order.

The higher relative L2-error of PINNs and DPA in Taylor Green Vortex is because of the difficulty in the underlying physics dynamics, as the flow is unsteady with decaying vortices [26]. Here, the advantages of pruning are more evident, as there is a 0.79% boost in accuracy over pruned-DPA, indicating significant improvement in function generalization arising from disentanglement from complex expressions, thereby improving interpretability. In the Diffusion-Reaction example, the ground-truth is an example of Depth-6. We were able to obtain a symbolic expression using a Depth-3 DPA, highlighting the expressive capacity of DPA. However, it comes at a cost of larger complex expressions with terms like log, 2nd and 3rd power terms which don't appear in the ground truth. In contrast to other examples, there isn't a large boost in accuracy from pruning here, as there are large inter-dependencies amongst operators to fit a Depth-6 expression into a Depth-3 DPA. Our framework shows promise for systems with no-analytical solution, as evident from the example of Air-Preheaters, where benchmark methods fail. Pruned-DPA gives 1-line symbolic expressions for generalizing temperature distributions in the entire domain, while the temperature MAE w.r.t. numerical simulations are comparable to that of PINNs.

4 LIMITATIONS

While our pruning strategy reduces the size of DPA, the ground-truth expressions are far more concise than the obtained expressions. One of the reasons is the greedy nature of the pruning algorithm. To provide an intuition, suppose a lesser weight is assigned to the operator when the DPA hasn't been pruned yet, which can ultimately provide the most concise expression. There is a high possibility of that weight getting pruned, as a different expression obtained by the remaining operators can still generalize over the dataset. Thus, our pruning strategy is sub-optimal in nature, and better pruning strategies need to be explored to obtain even more concise expressions. Nevertheless, this DFS-based pruning is a good starting point. Additionally, the convergence of DPA is non-trivial due to the varying mathematical properties of primitive operators. Detailed investigation on optimization guarantees of DPA and theoretical studies on its convergence and error bounds remain.

REFERENCES

[1] Yousef M. F. El Hasadi and Johan T. Padding. Solving fluid flow problems using semi-supervised symbolic regression on sparse data. *AIP Advances*, 9(11):115218, 2019.

[2] Mehrad Ansari, Heta A. Gandhi, David G. Foster, and Andrew D. White. Iterative symbolic regression for learning transport equations. *AIChE Journal*, 68(6):e17695, 2022.

[3] Kourosh Danai and William La Cava. Controller design by symbolic regression. *Mechanical Systems and Signal Processing*, 151:107348, 04 2021.

[4] Harsha Vaddireddy, Adil Rasheed, Anne E. Staples, and Omer San. Feature engineering and symbolic regression methods for detecting hidden physics from sparse sensor observation data. *Physics of Fluids*, 32(1):015113, 2020.

[5] Mahsa Mostowfi. A Study on Financial Time Series Forecasting and Symbolic Regression by means of a Hybrid Probabilistic Model-Building Cartesian Genetic Programming Methodology. 11 2022.

[6] Marco Virgolin and Solon P. Pissis. Symbolic regression is np-hard, 2022.

[7] Baihe He, Qiang Lu, Qingyun Yang, Jake Luo, and Zhiguang Wang. Taylor genetic programming for symbolic regression, 2022.

[8] Mojtaba Valipour, Bowen You, Maysum Panju, and Ali Ghodsi. Symbolicgpt: A generative transformer model for symbolic regression. *CoRR*, abs/2106.14131, 2021.

[9] Jacob Devlin, Jonathan Uesato, Surya Bhupatiraju, Rishabh Singh, Abdel-rahman Mohamed, and Pushmeet Kohli. Robustfill: Neural program learning under noisy I/O. *CoRR*, abs/1703.07469, 2017.

[10] Li Li, Minjie Fan, Rishabh Singh, and Patrick Riley. Neural-guided symbolic regression with semantic prior. *CoRR*, abs/1901.07714, 2019.

[11] Silviu-Marian Udrescu and Max Tegmark. Ai feynman: a physics-inspired method for symbolic regression. 2019.

[12] Silviu-Marian Udrescu, Andrew K. Tan, Jiahai Feng, Orisvaldo Neto, Tailin Wu, and Max Tegmark. AI feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. *CoRR*, abs/2006.10782, 2020.

[13] Brenden K Petersen, Mikel Landajuela Larma, Terrell N. Mundhenk, Claudio Prata Santiago, Soo Kyung Kim, and Joanne Taery Kim. Deep symbolic regression: Recovering mathematical expressions from data via risk-seeking policy gradients. In *International Conference on Learning Representations*, 2021.

[14] Daniel A. Abolafia, Mohammad Norouzi, and Quoc V. Le. Neural program synthesis with priority queue training. *CoRR*, abs/1801.03526, 2018.

[15] Michail Maslyaev, Alexander Hvatov, and Anna V. Kalyuzhnaya. Data-driven PDE discovery with evolutionary approach. *CoRR*, abs/1903.08011, 2019.

[16] Michael Zhang, Samuel Kim, Peter Y. Lu, and Marin Soljačić. Deep learning and symbolic regression for discovering parametric equations, 2022.

[17] Guillaume Lample and François Charton. Deep learning for symbolic mathematics. *CoRR*, abs/1912.01412, 2019.

[18] Maysum Panju, Kourosh Parand, and Ali Ghodsi. Symbolically solving partial differential equations using deep learning. *CoRR*, abs/2011.06673, 2020.

[19] Ritam Majumdar, Vishal Jadhav, Anirudh Deodhar, Shirish Karande, Lovekesh Vig, and Venkataramana Runkana. Physics informed symbolic networks, 2022.

[20] Guofeng Cui and He Zhu. Differentiable synthesis of program architectures. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 11123–11135. Curran Associates, Inc., 2021.

[21] Vishal Jadhav, Anirudh Deodhar, Ashit Gupta, and Venkataramana Runkana. Physics informed neural network for health monitoring of an air preheater. PHM Society European Conference, 7(1), 07 2022.

[22] Ritam Majumdar, Vishal Jadhav, Anirudh Deodhar, Shirish Karande, Lovekesh Vig, and Venkataramana Runkana. Real-time health monitoring of heat exchangers using hypernetworks and pinns, 2022.

[23] John E. Hopcroft, Rajeev Motwani, and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages, and Computation (3rd Edition)*. Addison-Wesley Longman Publishing Co., Inc., USA, 2006.

[24] Limin Wang, Yufan Bu, Dechao Li, Chunli Tang, and Defu Che. Single and multi-objective optimizations of rotary regenerative air preheater for coal-fired power plant considering the ammonium bisulfate deposition. *International Journal of Thermal Sciences*, 136:52–59, 2019.

[25] Bin Ding, Huimin Qian, and Jun Zhou. Activation functions and their characteristics in deep neural networks. In *2018 Chinese Control And Decision Conference (CCDC)*, pages 1836–1841, 2018.

[26] Tim De Ryck, Ameya D Jagtap, and Siddhartha Mishra. Error estimates for physics-informed neural networks approximating the Navier–Stokes equations. *IMA Journal of Numerical Analysis*, 01 2023. drac085.

[27] Adolf Fick. Ueber diffusion. *Annalen der Physik*, 170(1):59–86, 1855.

[28] Tony W.H. Sheu, S.K. Wang, and R.K. Lin. An implicit scheme for solving the convection–diffusion–reaction equation in two dimensions. *Journal of Computational Physics*, 164(1):123–142, 2000.

A Appendix

A.1 PDE information

Table 2 represents the PDE setups for our experiments. The second column consists of information on governing conditions, initial conditions, boundary conditions, and the domain of the spatial and temporal variables. The final column represents the ground-truth analytical solutions of output variables of interest.

Air-Preheater: We consider the non-dimensional form of APH. Equation 2 represents Conduction while 3 represents convection heat transfer. There are six outputs to this PDE system, three fluid temperatures (T) and three metal temperature (T_m) for given co-ordinates (θ, z). NTU and Pe stand for the number of transfer units and Peclet number respectively.

$$\frac{\partial T_{m_j}}{\partial \varphi} = NTU_{m_j}(T_j - T_{m_j}) + \frac{1}{Pe_{m_j}} \frac{\partial^2 T_{m_j}}{\partial z^2} \tag{2}$$

$$\frac{\partial T_j}{\partial z} = NTU_{m_j}\left(T_{m_j} - T_j,\right) \quad j = 1, 2, 3 \tag{3}$$

$$T_j\left(\varphi, z = 0\right) = T_{in,j}, \quad j = 1, 2, 3 \tag{4}$$

$$T_{m_1}(\varphi = 0, z) = T_{m_3}(\varphi = 1, 1 - z) \tag{5}$$

$$T_{m_1}(\varphi = 1, z) = T_{m_2}(\varphi = 0, 1 - z) \tag{6}$$

$$T_{m_2}(\varphi = 1, z) = T_{m_3}(\varphi = 0, z) \tag{7}$$

	Governing Conditions	Ground-truth expressions
Diffusion	$u_t = u_{xx} - e^{-t}sin(\pi x)(1-\pi^2)$ $u(x,0) = sin(\pi x)$ $u(-1,t) = u(1,t) = 0$	$u_{true} = e^{-t}sin(\pi x)$
Kovasznay	$u \cdot \nabla u + \nabla p = \nu \Delta u$ in $[0,1]^2$ $div(u) = 0$ in [0,1]	$u_{true} = 1 - e^{\lambda x}cos(2\pi y)$ $v_{true} = \lambda e^{\lambda x}sin(2\pi y)/2\pi$ $p_{true} = (1 - e^{2\lambda x})/2$
Taylor-Green	$u_t + u \cdot \nabla u + \nabla p = \nu \Delta u$ in $[0,2]^2 \times$[0,1] $div(u) = 0$ in $[0,2]^2 \times$[0,1] $u(t=0) = u_0$ in $[0,2]^2$	$u_{true} = -cos(\pi x)sin(\pi y)e^{-2\pi^2 \nu t}$ $v_{true} = sin(\pi x)cos(\pi y)e^{-2\pi^2 \nu t}$ $p_{true} = -\frac{(cos(2\pi x)+cos(2\pi y))e^{-2\pi^2 \nu t}}{4}$
Diffusion Reaction	$u_t = u_{xx} + e^{-t}f(x)$ $u(x,0) = p(x) + q(x)$ $u(t,-\pi) = u(t,\pi) = 0$ in $[-\pi,\pi] \times$[0,1] $f(x) = \frac{36sin(2x)+64sin(3x)+90sin(4x)+189sin(8x)}{24}$	$u_{true} = e^{-t}(p(x) + q(x))$ $p(x) = \frac{12sin(x)(1+cos(x))+4sin(3x)}{12}$ $q(x) = \frac{sin(4x)(1+cos(4x)}{4}$

Table 2: PDE-information

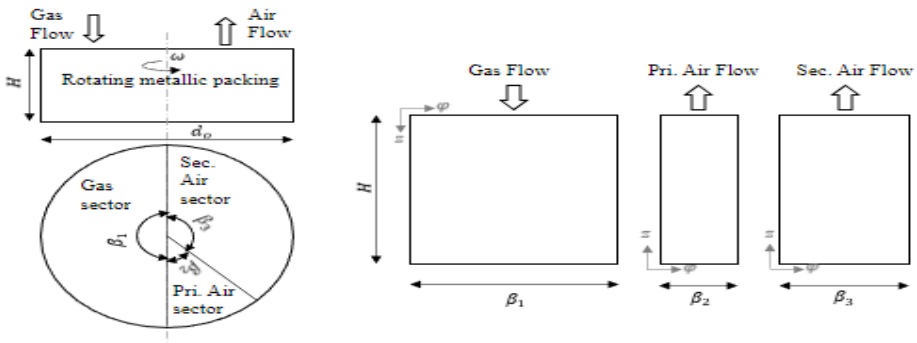

Figure 2: Air-Preheater schematic derived from [22]. The computational domain is divided into three parts, Gas, primary air and secondary air respectively.

$$\frac{\partial T_{m_j}[z=0,1]}{\partial z} = 0, \quad j = 1,2,3 \tag{8}$$

The boundary conditions are imposed by Gas inlet temperature ($T_{in,1}$), primary air inlet temperature ($T_{in,2}$), and secondary air inlet temperature ($T_{in,3}$) in Equation 4. Equations 5,6,7 impose continuity constraints on the metal temperature.

A.2 Experiments

We consider five systems of PDEs for our experiments, Diffusion equation [27], Navier-Stokes: Kovasznay flow [27], Navier-Stokes: Taylor Green Vortex equation [26], Diffusion Reaction equation, and two dimensional conjugate heat transfer in Air-Preheater[24]. Diffusion-Reaction PDEs are important for modeling chemical reactions [28] wherein there is a formation of new chemical products, and diffusion wherein there is a transfer of matter over a domain. Kovasznay flow is a two-dimensional steady-state Navier-Stokes equation with Reynold's Number of 20. Taylor-Green Vortex flow is a two-dimensional unsteady Navier-Stokes equation with viscosity $\nu = 0.01$. For both Kovasznay flow and Taylor-Green Vortex, we sample the boundary conditions from the ground-truth analytical solutions. Our final use case is that of Air-Preheaters (APH). APH is a heat exchanger deployed in thermal power plants to improve the thermal efficiency. Monitoring of internal temperature profiles of APH is important to avoid failures, which arises due to complex thermal and chemical phenomena. The reference solution of APH is derived using a Finite-Difference method

and doesn't have a ground-truth analytical solution. Inspection of internal temperature profiles can significantly benefit from symbolic representations in contrast to NNs due to improved interpretability. We describe the schematics of APH in A.1 and governing equations of other PDE systems in Table 2.

A.3 Training Schedule

PINNs: We consider a Neural Network with 3 hidden layers of 50 neurons each. We train using Adam optimizer for 60k epochs with multiplicative decay of 0.1 every 15k epochs, starting from a learning rate of $1e^{-3}$. We further finetune using the L-BFGS optimizer. In the examples of Diffusion, Diffusion Reaction, there are two input variables, x, t. We consider a computational domain of $[0, 1]$ **X** $[0, 1]$ with 10201 collocation points divided into 7500 training points and 2701 test points. In Kovasznay flow, we consider a 101x101 equally spaced grid domain to represent the XY plane, where X, Y $\in$ [-0.5,1.0] **X** [-0.5,1.5]. We consider 2601 collocation points and 320 boundary condition points, with 80 points for each face of the grid. Taylor-Green Vortex: the computational domain is defined for X,Y,T as [0,2] **X** [0,2] **X** [0,1], with a time-step of 0.1 in the time dimension. We sample 25k collocation points, 5k initial-boundary condition points throughout the domain. For evaluating the test-performance, across all PDE examples, we sample 10k points for every output variable of interest.

Differentiable Program Architecture: Table 3 consists of the information on the maximum depth of the architecture and operators of interest for every PDE system. The weights of the DPA are initialized using the Glorot-uniform optimizer which is used to perform SR. We use an Adam Optimizer for 100k epochs with multiplicative decay of 0.1 every 25k epochs with an initial learning rate of $1e^{-2}$. We further apply early stopping if the generalization on datapoints generated by PINNs doesn't improve in the last 5k epochs. For training the unpruned-DPA, we apply L1-regularization over all weights with L1-coefficient of $1e^{-5}$. All experiments were conducted on Nvidia P100 GPU with 16 GB GPU Memory and 1.32 GHz GPU Memory clock using Pytorch framework.

SymbolicGPT: The training methedology of SymbolicGPT is adapted from the original paper [8]. The hyperparameter specifications are as follows: numEpochs: 20, embeddingSize: 512, number of points:10k, blockSize: 200, testBlockSize: 400, batchSize: 128, variable-embedding: False.

AI-Feynman: The training methodology of AI-Feynman is adapted from the original paper [12]. We consider the following hyperparameter sets for our experiments: 'bftt':{60,120}, 'epochs':{300,400,500}, 'op':{'7ops.txt','10ops.txt','14ops.txt','19ops.txt'}, 'polynomial degree':{3,4,5}.

DSR: The training methodology of Deep Symbolic Regression is adapted from the original paper [13]. We consider the following hyperparameter sets for our experiments: 'seed':{1,2,3,4,5}, 'function set':{'add', 'sub', 'mul', 'div', 'sin', 'cos', 'exp', 'log', 'const'}, batch size:1000, learning rate: $5e^{-4}$, entropy coefficient: 0.005, risk factor: 0.05.

		Depth	Unpruned	Pruned	Operators
Diffusion	u	2	157	15	[x,t,sin,exp,+,$*$]
Kovasznay	u	2	343	12	[x,y,sin,exp,+,$*$,log,$pow2$,$pow3$]
	v	2	343	13	[x,y,sin,exp,+,$*$,log,$pow2$,$pow3$]
	p	2	343	9	[x,y,sin,exp,+,$*$,log,$pow2$,$pow3$]
Taylor-Green	u	3	3097	20	[x,y,t,sin,exp,+,$*$,log,$pow2$,$pow3$]
	v	3	3097	33	[x,y,t,sin,exp,+,$*$,log,$pow2$,$pow3$]
	p	3	3097	35	[x,y,t,sin,exp,+,$*$,log,$pow2$,$pow3$]
Diffusion Reaction	u	3	3097	63	[x,t,sin,exp,+,$*$,log,$pow2$,$pow3$]
Air-Preheater	T_{fg}	2	157	6	[θ,z,sin,exp,+,$*$]
	T_{mg}	2	157	15	[θ,z,sin,exp,+,$*$]
	T_{fa_1}	2	157	15	[θ,z,sin,exp,+,$*$]
	T_{ma_1}	2	157	9	[θ,z,sin,exp,+,$*$]
	T_{fa_2}	2	157	8	[θ,z,sin,exp,+,$*$]
	T_{ma_2}	2	157	12	[θ,z,sin,exp,+,$*$]

Table 3: Differentiable Program Architecture characteristics. Unpruned and pruned refers to # parameters in the resulting DPA respectively.

A.4 Generated Symbolic expressions

Diffusion	$u = (1.51x - 2.04sin(-2.51x + 0.20t))(2.62x + 0.32t)$ $+3.67sin(1.30sin(0.37t + 1.63) + 3.13x) - 3.53$
Kovasznay Flow	$u = 1.01 + 0.99(sin(6.28y - 1.57)e^{-1.81x})$ $v = sin(6.28y - 3.14)(0.29 - 0.54x + 0.46x^2 - 0.28x^3 + 0.12x^4)$ $p = -2x^4 + 4.38x^3 - 3.75x^2 + 1.81x + 0.02$
Taylor-Green Vortex	$u = (1.93x^2 - 0.13xt - 0.03t - 0.26x - 0.36)(4.13y - 4.12)*$ $(-0.72sin(1.71y) + xy - 0.56y^2 + 0.06y) + 0.50$ $v = 0.21sin(-0.33xt + 0.34yt + 3.41x - 3.43y) + (0.74x + 0.54y + 0.13t - 1.09)$ $*(-1.03x - 0.93y - 0.31)(0.31x - 0.69sin(1.69x + 1.87y - 2.11))$ $p = 0.75sin(h_1) - 0.29h_3 - 0.17(sin((-0.91x + 0.32t + 0.61) + 3.30))$ $h_3 = sin(2.07(-0.18t + 1.84)(-1.64x - 0.44t + 0.41))$ $h_1 = 0.25t - 1.61sin(1.89y + 0.08) + (1.30y - 0.03t)(0.18y + 0.23t - 1.41) + 0.33$
Diffusion Reaction	$u = 0.30log(0.64log(-0.44x^2 - 1.9x - 0.95)) - 0.07h_2^2 + h_3$ $h_2 = 0.84 - 0.25sin(0.95x - 0.33) + 0.46h_8^2 - 0.43h_9^3$ $h_3 = 0.19(-0.26x - 0.12y - 0.15e^{-0.66-0.22y})^3$ $h_8 = 0.87x - 0.47e^{0.83x+0.64} + 0.83(-0.99x + 0.33y - 0.25)$ $(0.49x + 0.02y + 0.71) - 0.07(0.07x - 0.30y - 0.04)^3$ $h_9 = 0.17sin(0.22x - 0.07y - 0.28) - 0.19(0.67x - 0.14y - 0.07)^2$ $-0.09(0.50x + 0.72)^3 + 0.22x - 0.07y - 0.02$
APH	$T_{fg} = 0.99sin(0.22z + 0.83e^{-0.21z+0.61\theta+0.66})$ $T_{mg} = (0.52sin(1.56\theta) + 1.18)(0.83e^{-0.44\theta} + 0.18z\theta - 0.23\theta^2 - 0.27\theta)$ $T_{fa_1} = (-0.11z + 0.77\theta - 1.59 + 0.27(z\theta))(0.03z + 0.22\theta - 0.60 + 0.37\theta^2)$ $T_{ma_1} = (-1.14sin(0.07z + 0.88\theta - 1.06))(0.06z + 0.69\theta + 0.97)$ $T_{fa_2} = 0.86sin(0.34z + 1.26\theta + 0.55)(-0.10z - 1.02\theta)(0.11z + 1.10\theta) + 0.96$ $T_{ma_2} = (0.13z + 1.30sin(1.08\theta) + 1.25)(0.08z - 0.68\theta + 0.77)$

Table 4: Expressions obtained after pruning DPA

	AI-Feynman	Symbolic GPT	DSR
u	$1.01exp(-0.99t) * cos(0.99x - 1.56)$	$1.22x - log(x - \theta)$	$sin(cos(sin(log(x)))$
u	$1.24cos(x)$	$exp(0.43x - 0.71y) + 2.23$	$sin(sin(x)) + tan(y)$
v	$sin(x) - 1.14cos(y)$	1.52	$2.57tan(y)$
p	$0.51 - 0.48e^{\lambda x}$	$log(y - exp(1.13x) + 0.47)$	$0.52x - 1.62y$
u	$0.75tan(1.61sin(y - z) - 0.33cos(t)$	1.1704	1.6321
v	4.43	$1.82 + 1.10y - 0.43t$	$x^3 - 1.22cos(t) - 1.15$
p	$2.24x - (1.19log(x) - 0.47t - 0.22)^3$	$0.95x - 0.18y + 1.44$	$0.08log(1.87)$
u	$1.27exp(x - t)(cos(x) - sin(cos(t)))$	$3.33e^{-t}(1.14x - x^2)$	$sin(sin(x))cos(cos(t))$
T_{fg}	$1.45log(2z\theta + (sin(z - \theta)^3)$	$\theta - log(0.96z + 0.34)$	$sin(sin(sin(\theta)))$
T_{mg}	$2.67exp(1.33cos(\theta - 5.53)) - 3.41$	$z - \theta^2$	3.32
T_{fa_1}	$tan(0.77z) - \theta * cos(z) - 1.05$	$3.22log(1.19z) - 0.46$	$cos(cos(cos(\theta)))$
T_{ma_1}	$1.22z^2$	$1.17\theta - 0.95$	$z - \theta + cos(cos(1.32))$
T_{fa_2}	$1.64log(2.32\theta) - 0.19z^3$	$2.43log(\theta)log(z) + 3.21$	$cos(cos(1.33sin(z) - \theta))$
T_{ma_1}	$z - \theta$	$1.52log(\theta z) - 0.96$	$1.67\theta - z$

Table 5: Symbolic Expressions generated by benchmark methods

A.5 Pruned Vs Unpruned: Diffusion Equation

$U_{unpruned} = -1.76e^{-6}x + 0.30t - 0.78sin(h_{20}) - 0.43e^{h_{21}} - 4.84e^{-5}(h_{22} + h_{23}) + 1.04h_{24}h_{25} + 3.45e^{-5}$
$h_{20} = -0.57x - 6.45e^{-5}t - 0.54sin(h_{10}) - 0.47e^{h_{11}} + 1.28(h_{12} + h_{13}) + 1.05e^{-4}h_{14}h_{15} + 0.15$
$h_{21} = -1.16e^{-5}x - 2.18e^{-5}t + 7.28e^{-5}sin(h_{16}) - 1.85e^{-5}e^{h_{17}} - 1.77e^{-4}(h_{18} + h_{19}) - 2.12e^{-5}h_{110}h_{111} + 0.07$
$h_{22} = -1.16e^{-5}x - 2.18e^{-5}t + 7.28e^{-5}sin(h_{112}) - 1.85e^{-5}e^{h_{113}} - 1.77e^{-4}(h_{114} + h_{115}) - 2.12e^{-5}h_{116}h_{117} - 6.65e^{-5}$
$h_{23} = 7.42e^{-5}x + 1.63e^{-5}t - 1.60e^{-4}sin(h_{118}) - 8.29e^{-5}eh^{119} - 6.21e^{-5}(h_{120} + h_{121}) + 1.59e^{-4}h_{122}h_{123} + 0.0002$
$h_{24} = 6.04e^{-1}x - 1.95e^{-1}t - 8.08e^{-1}sin(h_{124}) + 7.38e^{-5}e^{h_{125}} - 2.00e^{-1}(h_{126} + h_{127}) - 6.44e^{-5}h_{128}h_{129} + 0.20$
$h_{25} = 3.39e^{-2}x - 1.44e^{-5}t - 8.03e^{-5}sin(h_{130}) + 3.85e^{-1}e^{h^{131}} - 3.38e^{-5}(h_{132} + h_{133}) + 1.14h_{134}h_{135} + 0.29$
$h_{10} = -0.04x + 1.32t + 0.60,$
$h_{11} = 4.65e^{-1}x + 6.25e^{-6}t + 0.5031$
$h_{12} = -7.07e^{-1}x + 1.47e^{-5}t + 8.60e^{-6},$
$h_{13} = -1.02x + 1.49e^{-5}t + 1.23e^{-5}$
$h_{14} = -0.0001x - 0.0001t - 0.0002,$
$h_{15} = -1.25e^{-4}x + 5.77e^{-5}t - 0.0002$
$h_{16} = -4.31e^{-5} - 1.34e^{-4} - 8.35e^{-5},$
$h_{17} = -1.64e^{-4}x + 2.97e^{-5}t - 7.37e^{-5}$
$h_{18} = 4.88e^{-6}x + 7.52e^{-5}t - 8.31e^{-5},$
$h_{19} = -9.6106e^{-5}x - 9.97e^{-5}t + 1.88e^{-5}$
$h_{110} = 5.38e^{-5}x - 6.86e^{-5}t + 0.0001,$
$h_{111} = -2.50e^{-4}x + 9.12e^{-5}t - 0.0001$
$h_{112} = 0.0001x + 0.0001t - 5.28e^{-5},$
$h_{113} = 1.23e^{-5}x + 6.02e^{-5}t + 2.20e^{-5}$
$h_{114} = 0.0002x - 0.0002t + 0.0003,$
$h_{115} = 3.81e^{-5}x + 7.75e^{-5}t + 8.39e^{-5}$
$h_{116} = 3.08e^{-4}x + 1.16e^{-5}t - 0.0003,$
$h_{117} = 6.94e^{-5}x - 1.38e^{-4}t - 4.28e^{-5}$
$h_{118} = 9.81e^{-5}x - 6.65e^{-5}t + 3.90e^{-5},$
$h_{119} = 9.98e^{-5}x - 1.34e^{-4}t + 8.06e^{-5}$
$h_{120} = -4.02e^{-6}x + 5.20e^{-5}t - 0.0003,$
$h_{121} = -4.28e^{-5}x - 1.22e^{-4}t - 7.00e^{-6}$
$h_{122} = -0.0002x + 0.0001t + 9.46e^{-5},$
$h_{123} = 4.80e^{-5}x - 4.22e^{-5}t + 3.66e^{-5}$
$h_{124} = -2.06x + 0.65t + 0.53,$
$h_{125} = 5.84e^{-5}x + 1.66e^{-4}t - 1.68e^{-5}$
$h_{126} = -2.03e^{-1}x + 1.86e^{-5}t - 2.52e^{-5},$
$h_{127} = -4.98e^{-6}x + 1.8e^{-5}t + 1.07e^{-5}$
$h_{128} = -3.32e^{-5}x - 8.32e^{-5}t - 0.0001,$
$h_{129} = 2.89e^{-5}x + 6.21e^{-5}t + 0.0001$
$h_{130} = -1.42e^{-4}x + 1.01e^{-7}t + 6.28e^{-5},$
$h_{131} = 3.57e^{-5}x - 5.78e^{-2}t + 0.37$
$h_{132} = -2.34e^{-4}x + 1.66e^{-6}t - 6.33e^{-5},$
$h_{133} = 0.0001x + 0.0002 - 0.0002$
$h_{134} = 0.81x - 0.03t - 0.65,$
$h_{135} = -1.23x + 0.11t + 1.03e^{-5}$

$U_{pruned} = (1.51x - 2.04sin(-2.51x + 0.20t))(2.62x + 0.32t) + 3.67sin(1.30sin(0.37t + 1.63) + 3.13x) - 3.53$

A.6 PRUNING ALGORITHM VISUALIZATION

As an example, we take a DPA of depth 2 with sin,exp,log as operators, and x,y,c as leaf nodes.

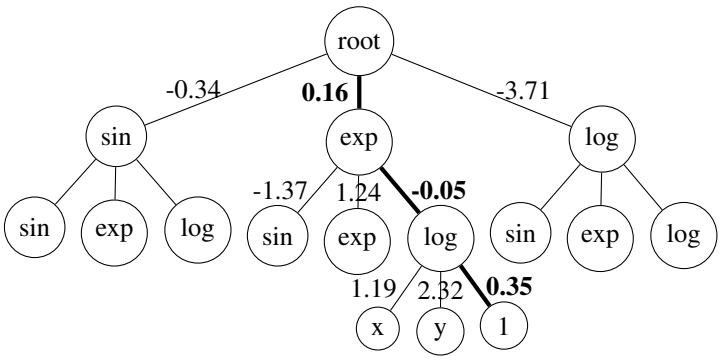

Figure 3: Starting from the root, we recursively select the node with minimum magnitude till we encounter the leaf. The path followed is root→exp→log→1. Now, we prune the leaf and assume after fine-tuning, the loss of newer architecture is at par with the original. Thus, we accept the prune.

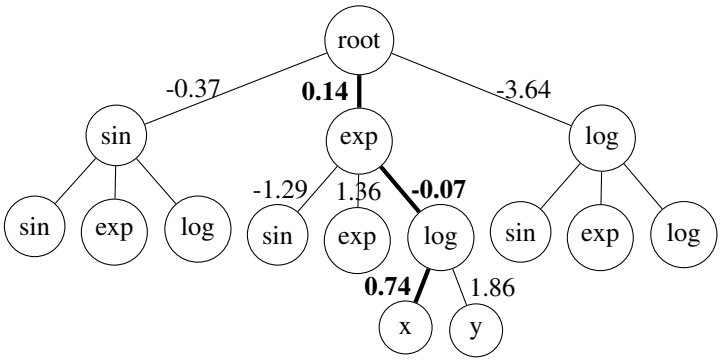

Figure 4: Currently, we are at node root→exp→log. The node x is selected and tested for pruning. Let's assume the pruning is accepted. The resulting tree is as follows:

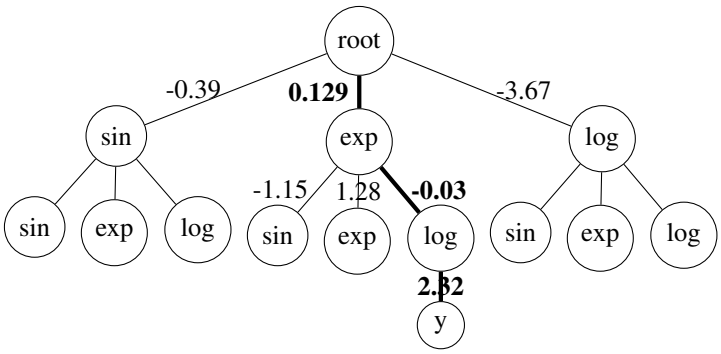

Figure 5: Now the node root→exp→log→y is tested. Let's assume fine-tuned weight performance is worse, hence it's not prunable. Resulting DPA:

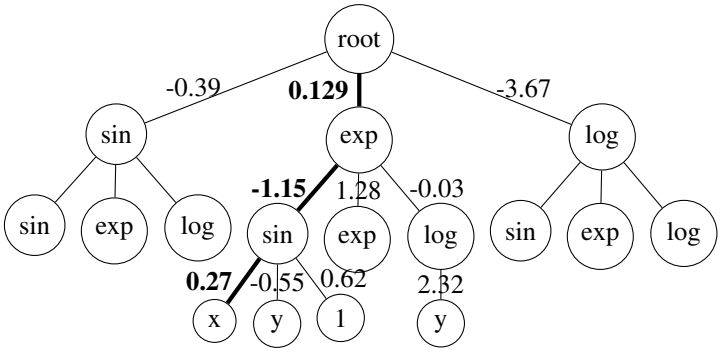

Figure 6: Let's assume root→exp→log→y is not prunable, the algorithm recurses back to root→exp→log. The prune is rejected again, and recurses to node root→exp. The nodes selected following the algorithm are sin→x, and the algorithm continues recursively, in a depth-first manner. Let's assume after all the nodes are visited, the DPA obtained is:

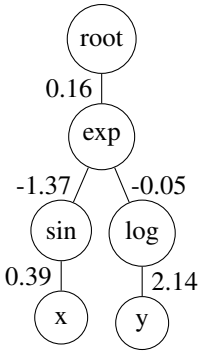

Figure 7: The final symbolic expression is $0.16exp(-1.37sin(0.39x) - 0.05log(2.14y))$.