Governing Equation Discovery with Relaxed Symmetry Constraints

Andrew Chen
Canyon Crest Academy
`andrewcsd66@gmail.com`

Jianke Yang
UCSD
`jiy065@ucsd.edu`

Rose Yu
UCSD
`roseyu@ucsd.edu`

Abstract

Existing methods for discovering governing equations from data often struggle with the vast search space of possible equations. Physical inductive biases such as symmetry are shown to reduce complexity and force symmetrical equations. State-of-the-art methods enforce symmetry by using *symmetry invariants* as relevant terms in symbolic regression. While effective, they assume perfect symmetry and fail to identify systems with symmetry-breaking effects. To solve this problem, we propose **Symmetry-Breaking Fine-Tuning (SBFT)** for genetic programming-based equation search, which aims to relax the symmetry constraints. Our method first searches with an emphasis toward invariants to recover a symmetric backbone, then fine-tunes those results with equal emphasis on invariants and raw variables to capture symmetry-breaking terms. On benchmark PDEs, SBFT recovers equations achieving median RMSE reduction of $\mathbf{85.56}\%$ and $\mathbf{67.83}\%$ relative to standard and invariant-based genetic programming, respectively, across all experiments.

1 Introduction

Discovering symbolic differential equations from data uncovers the fundamental dynamics behind complex systems. However, existing methods often struggle with the vast search space of equations, as the number of candidate expressions grows combinatorially with the number of variables and operators (Udrescu & Tegmark, 2020; Udrescu et al., 2020).

A common approach in equation discovery is to introduce inductive biases from physical laws. Specifically, various works leverage symmetry constraints to improve the accuracy and efficiency of equation discovery. However, applications of such methods have been limited. For example, Udrescu & Tegmark (2020) focus on algebraic expression discovery, and Otto et al. (2023); Yang et al. (2024) study Ordinary Differential Equation (ODE) systems. Extending these ideas to more complex systems, such as Partial Differential Equations (PDEs), has proven to be challenging. PDE-focused approaches often only encourage symmetry – using data augmentation (Brandstetter et al., 2022), regularization terms (Akhound-Sadegh et al., 2023), or self-supervised learning (Mialon et al., 2023) – but these "soft" constraints only guide the learning process, and lack the strength to ensure strict mathematical symmetry. Direct enforcement methods exist but typically handle only narrow classes of symmetries and systems (Wang et al., 2021; Otto et al., 2023; Yang et al., 2024). More recently, Yang et al. (2025) used symmetry invariants in symbolic regression for PDE discovery, leveraging the fact that any differential equation with a Lie group symmetry can be expressed entirely in terms of the group invariants. Treating invariants as "atomic entities" restricts the search space to symmetry-respecting equations, reducing search complexity and improving accuracy in both sparse regression and genetic programming (GP) on symmetric PDEs.

However, many real-world systems do not obey symmetry perfectly. Physical processes often contain *symmetry-breaking terms* arising from factors such as external forces. In these cases, invariant-based methods cannot fully capture the system's behavior, since the invariant framework excludes

39th Conference on Neural Information Processing Systems (NeurIPS 2025).

terms that explicitly violate the symmetry. To address these shortcomings, we propose **Symmetry-Breaking Fine-Tuning (SBFT)**, depicted in Figure 1. Our approach builds on invariant-based genetic programming for discovering governing equations, but extends it to handle systems with imperfect symmetries. SBFT first biases the search toward the symmetry-respecting invariants, building a symmetric backbone, then fine-tunes with equal weighting of invariants and raw variables, allowing symmetry-breaking terms to emerge.

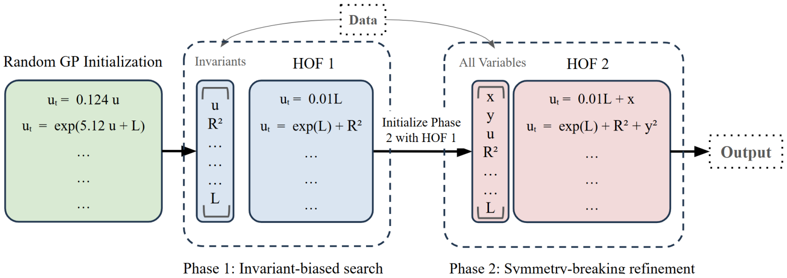

Figure 1: Proposed Symmetry-Breaking Fine-Tuning pipeline.

2 Methodology

Problem Definition Differential equation discovery seeks to find the governing mathematical laws of a system directly from observational data. Given measurements of system variables and their derivatives over time or space, the goal is to recover a symbolic equation that accurately captures the underlying dynamics of the system. Such methods aim to discover symbolic PDEs in the form $F(\mathbf{x}, u^{(n)}) = 0$, where $\mathbf{x}$ denotes the independent variables, $u^{(n)}$ consists of the dependent variable u and all of its up-to-nth order partial derivatives.

2.1 Invariant-Based Equation Discovery

We first briefly review the approach in (Yang et al., 2025), which uses symmetry invariants to embed physical principles directly into symbolic regression. The core idea is that any differential equation that admits a symmetry of a Lie group can be rewritten entirely in terms of the invariants of that group: quantities formed from the system's variables and derivatives that remain unchanged under the transformations of the symmetry.

By treating these invariants as "atomic entities", two things are ensured: 1. Every equation proposed using invariants automatically respects the specified symmetry; 2. The search space is dramatically reduced since symmetry violating equations are not considered. In practice, the approach works as follows: the infinitesimal generators of the symmetry group are used to construct the differential invariants up to the desired order. These invariants are then evaluated on the data set and used as the feature library for symbolic regression. Genetic programming or sparse regression algorithms can then search over this restricted library to propose candidate equations.

Yang et al. (2025) has shown that this approach consistently produces more accurate results for systems with exact symmetry. In a strictly symmetric setting, invariants lead to faster convergence, cleaner equations, and lower reconstruction error than using raw variables alone. However, this reliance on exact symmetry fails in systems with symmetry-breaking effects, motivating our approach.

2.2 Symmetry-Breaking Fine-Tuning

In this work, we focus on genetic programming as the base algorithm for equation discovery, given its broader and more flexible search space. We introduce the Symmetry-Breaking Fine-Tuning (SBFT) pipeline, which discovers PDEs in systems with imperfect symmetry by operating on a combined variable library of differential invariants of the prescribed symmetry group and the original jet variables. This unified representation preserves the symmetric structure of the system while admitting potential symmetry-breaking terms. As in standard symbolic regression, one variable is chosen as the left-hand side; to avoid trivial reconstructions (i.e. an invariant is reconstructed directly from raw variables), we remove a specified subset of variables depending on the system from the available library. For example, if the LHS target is invariant $L = u_{xx} + u_{yy}$, then at least one of its defining variables u_{xx} or u_{yy} is removed. The algorithm is described in full in Algorithm 1.

Algorithm 1 Symmetry-Breaking Fine-Tuning

```
Require: PDE order $n$, dataset $\{\mathbf{z}^i = (\mathbf{x}^i, (u^{(n)})^i) \in M^{(n)}\}_{i=1}^{N_D}$, base SR algorithm $\mathcal{S}$ (e.g., PySR), symmetry generators $\mathcal{B} = \{\mathbf{v}_a\}$
Compute invariants $\eta^1, \dots, \eta^K$ from $\mathcal{B}$ up to order $n$
Form the combined variable library $\Phi = \{\text{regular variables}\} \cup \{\text{invariants}\}$ and dataset $X_\Phi$
Remove any variables in the "delete list" to avoid trivial solutions
for each target equation (and specified LHS) do
    Set current LHS variable $y$ (target) and RHS feature matrix $X_\Phi^{(-y)}$
    Phase 1: Invariant-biased search
    Run $\mathcal{S}$ on $(X_\Phi^{(-y)}, y)$ with variable complexities: $c(v) = 1$ if $v$ is invariant, else $\alpha$ (default $\alpha = 5$). Save all discovered candidate equations as symmetric backbone: this set of candidates is called the Phase 1 Hall of Fame (HOF).
    Phase 2: Symmetry-breaking refinement
    Set all variable complexities to 1 (equal weighting) and warm-start $\mathcal{S}$ with Phase 1 HOF
    Run $\mathcal{S}$ to refine candidates, allowing regular variables to form symmetry-breaking terms, which are not constrained by symmetry. Save best-scoring equation
end for
Return final best-scoring equation(s) from Phase 2 for all targets
```

3 Experiments

In this section, we evaluate the performance of our proposed SBFT pipeline. We present the core results, demonstrating that SBFT consistently discovers equations with lower error than standard genetic programming and a pure symmetry-invariant GP approach.

3.1 Data Generation

We consider the following systems of PDEs with symmetry-breaking and invariants obtained from Yang et al. (2025):

Darcy Flow. We consider the steady state 2D Darcy flow equation from Takamoto et al. (2022) with spatially varying viscosity $a(x, y)$ and a constant force term:

$$-\nabla(a(x, y)\nabla u(x, y)) = 1 \tag{1}$$

This equation admits an SO(2) rotation symmetry $\mathbf{v} = y\partial_x - x\partial_y$. In our experiment, we use the following complete set of 2nd-order rotational invariants calculated from Yang et al. (2025): $\{\frac{1}{2}(x^2 + y^2), u, xu_y - yu_x, xu_x + yu_y, u_{xx} + u_{yy}, u_{xx}^2 + 2u_{xy}^2 + u_{yy}^2, x^2u_{xx} + y^2u_{yy} + 2xyu_{xy}\}$.

We consider the following asymmetric viscosity terms: $a_1(x, y) = e^{-4(x^2+y^2)+0.1x}$, and $a_2(x, y) = 1.1 - 2(x^2 + y^2) + 0.2x$.

To prevent the trivial case where SBFT simply recovers an invariant as equal to its defining expression in terms of the regular variables (e.g., $\zeta_2 = xu_x + yu_y$), yielding no genuine discovery, we remove certain variables depending on the dataset.

For Darcy Flow, we remove $\{u_{xx}, u_{yy}\}$ for a_1 and $\{y, u_y\}$ for a_2.

Reaction-Diffusion. We consider the following system of PDEs from Champion et al. (2019):

$$\begin{aligned} u_t &= d_1\nabla^2 u + (1 - A^2)u + A^2 v \\ v_t &= d_2\nabla^2 v - A^2 u + (1 - A^2)v \end{aligned} \tag{2}$$

where $A^2 = u^2 + v^2$ and $d_1 = d_2 = 0.1$.

The symmetry invariants are calculated from Yang et al. (2025). The system exhibits rotational symmetry in the phase space: $\mathbf{v} = u\partial_v - v\partial_u$. The ordinary invariants are $\{t, x, y, u^2 + v^2\}$. The higher-order invariants are $\{\mathbf{u} \cdot \mathbf{u}_\mu, \mathbf{u}^\perp \cdot \mathbf{u}_\mu\}$, where $\mathbf{u} = (u, v)^T$ and μ is any multi-index of t, x and y. Expanding these inner products explicitly yields:

$$\mathbf{u} \cdot \mathbf{u}_\mu = u\, u_\mu + v\, v_\mu, \qquad \mathbf{u}^\perp \cdot \mathbf{u}_\mu = -v\, u_\mu + u\, v_\mu,$$

We introduce two symmetry-breaking variants: (i) unequal diffusivities, $d_2 = d_1 + \epsilon$ with $\epsilon = 0.03$; (ii) external forcing, adding $-\epsilon v$ and $-\epsilon u$ with $\epsilon = 0.15$ to the RHS of u_t and v_t respectively. For SBFT, we remove $\{u_t, v_t\}$ in both cases.

2D Heat. Consider the heat equation given by $u_t = \alpha(u_{xx} + u_{yy}) + f(x, y)$. In our experiments, we use $\alpha = 0.01$ and two types of source terms: $f(x, y) = \exp(-x^2)$ and $f(x, y) = \exp(-x^2 - 0.9y^2)$. We will use the same set of invariants as Darcy Flow. This data was generated by an explicit finite-difference solver, with an initial condition of $\exp(-50(x - 0.5)^2 + (y - 0.5)^2)$. No variable deletion from SBFT was necessary for this dataset.

3.2 Experimental Setup

We compare the results of SBFT with two key baselines: **Standard Genetic Programming** searches equations using raw variables and their derivatives(e.g. $u, u_x, u_{xx}, \dots$) without any built-in knowledge of symmetry; **Invariant-Based Genetic Programming** uses a special set of symmetry invariants such that any equation built from these invariants is guaranteed to respect the system's symmetry. We selected PySR and invariant-based PySR as baselines as they share the same genetic programming framework, isolating the effect of symmetry treatment (Cranmer, 2023; Yang et al., 2025). In contrast, alternative approaches such as AI Feynman rely on different pipelines and additional priors, while SINDy operates over a different equation space and is therefore not a directly comparable baseline (Udrescu & Tegmark, 2020). Specific implementation details are shown below:

Table 1: Hyperparameters passed to PySR Cranmer (2023) for each dataset, consistent across trials.

Dataset	**Populations**	**Max Equation Size**	**Iterations**	**Population Size**	**Fraction Replaced**
Darcy Flow	127	25	200	64	0.1
Reaction-Diffusion	127	25	400	64	0.5
Heat 2D	127	25	50	27	0.00036

Experiments are run on symmetry-breaking datasets, reporting the mean Root Mean Squared Error (RMSE) over three independent trials, with and without normalization, trained on 10,000 samples. We apply normalization by scaling each variable using z-score standardization before passing the data to PySR, and then denormalizing the predicted outputs back to the original scale. All methods use identical hyperparameters and randomized seeds, with variable weighting $\alpha = 5$. To ensure fair comparison, SBFT's two stages each run for half the baseline iteration count.

3.3 Experimental Results

Under identical experimental settings, SBFT consistently achieves lower RMSE than the baselines, which translates to more practically useful governing equations. On average, SBFT discovered equations with $\mathbf{85.56}\%$ and $\mathbf{67.83}\%$ lower median RMSE than regular and invariant mode respectively across all datasets. Table 2 reports the full results: normalization improves performance consistently, and aside from a single outlier, SBFT markedly outperforms all baselines.

Table 2: Experimental results across normalization schemes, measured by RMSE (rounded).

	Regular		**Invariant**		**SBFT**	
	No Norm.	**Norm.**	**No Norm.**	**Norm.**	**No Norm.**	**Norm.**
Darcy Flow: $a_1(x, y)$	1.0×10^{-2}	6.9×10^{-3}	1.8×10^{-3}	3.0×10^{-2}	$\mathbf{7.7 \times 10^{-4}}$	7.9×10^{-4}
Darcy Flow: $a_2(x, y)$	1.8×10^{-2}	1.0×10^{-2}	$\mathbf{1.3 \times 10^{-4}}$	3.1×10^{-2}	8.9×10^{-4}	2.1×10^{-3}
Reac-Diff: Unequal Diff.	3.2×10^{-2}	4.7×10^{-2}	2.2×10^{-2}	2.0×10^{-2}	1.4×10^{-2}	$\mathbf{7.7 \times 10^{-3}}$
Reac-Diff: External Forcing	4.0×10^{-2}	5.6×10^{-2}	6.0×10^{-2}	7.3×10^{-2}	1.6×10^{-2}	$\mathbf{6.9 \times 10^{-3}}$
Heat 2D: $f(x) = \exp(-x^2)$	2.1×10^{-1}	7.8×10^{-2}	1.6×10^{-1}	5.8×10^{-2}	7.8×10^{-2}	$\mathbf{2.4 \times 10^{-3}}$
Heat 2D: $f(x) = \exp(-x^2 - 0.9y^2)$	1.3×10^{-1}	4.1×10^{-2}	1.8×10^{-1}	2.3×10^{-2}	$\mathbf{1.6 \times 10^{-2}}$	1.6×10^{-2}

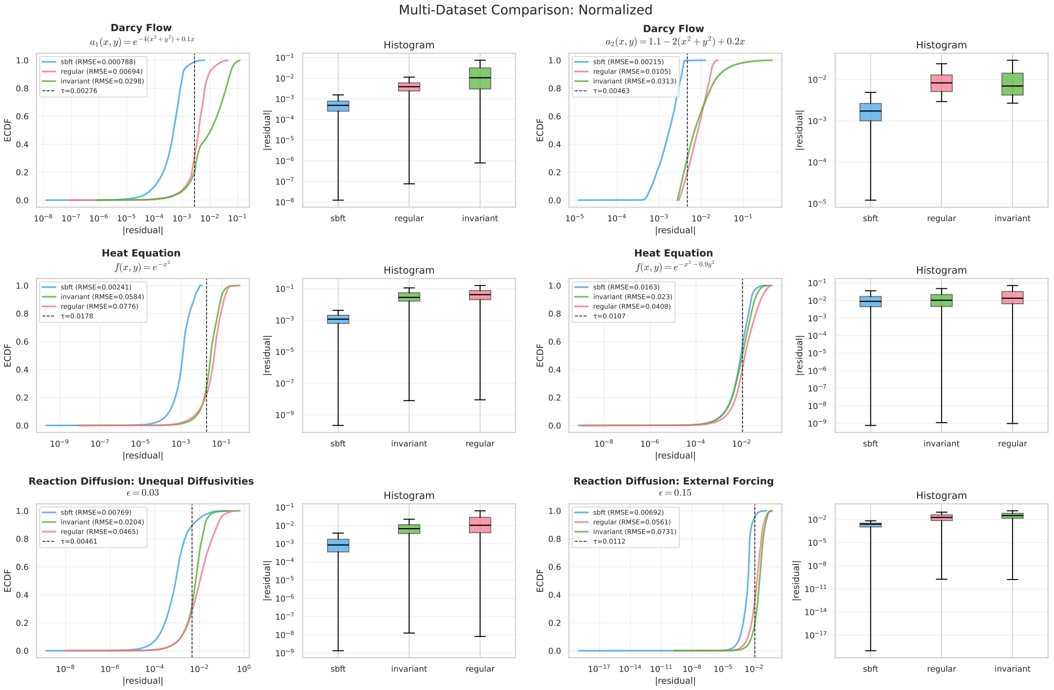

Figure 2: SBFT vs baselines across various symmetry-breaking datasets, with normalization.

Although SBFT does not exactly recover the ground truth equation, it makes substantial progress toward the correct form. Baseline methods, by contrast, often fail to accommodate the symmetry-breaking term and produce expressions that deviate markedly from the target structure. Some equation samples are provided in Table 3 below:

Table 3: Equation samples for **Darcy Flow** and **Heat 2D** datasets (without normalization).

Method	**Equation**
Darcy Flow $a_1(x,y)$ (Rounded to one significant figure)	
Ground Truth (Regular)	$0 = u_{xx} + u_{yy} - 8(xu_x + yu_y) + 0.1u_x + \exp\big(-4(x^2+y^2)^2 + 0.1x\big)$
Ground Truth (Invariant)	$\boldsymbol{0 = L - 8\zeta_2 + 0.1u_x + \exp\big(-4R^2 + 0.1x\big)}$
Regular	$0 = 1.0\,uu_x\exp(-0.2u_{yy}) + 1.0\,u_x\exp(u_{xx} - 0.2u_{yy}) + 0.3u_x\exp(-0.2u_{yy}) + 0.6x$
Invariant	$0 = 0.03L + 0.5R^2\exp(1.4R) + 1.0\,u\exp(1.4R) - 0.5\zeta_2 - 0.1$
SBFT	$\boldsymbol{0 = 0.1L + 0.1u_x - 1.0\zeta_2 + 0.1\exp(4.0R - 0.1x)}$
Heat 2D $f(x) = \exp(-x^2)$ (Rounded to one sig. figure)	
Ground Truth (Regular)	$0 = u_t - 0.01u_{xx} - 0.01u_{yy} + \exp(-x^2)$
Ground Truth (Invariant)	$\boldsymbol{0 = u_t - 0.01L + \exp(-x^2)}$
Regular	$0 = 0.7u_t + 0.5\exp(x) - 1.0$
Invariant	$0 = -Ru^2 + u_t - \exp(0.02L + 0.005\theta_3 + \zeta_2)$
SBFT	$\boldsymbol{0 = -0.01L + u_t - 0.8\exp(-x - x^2)}$

4 Conclusion

In this work, we introduce the Symmetry-Breaking Fine-Tuning pipeline, a two-stage framework that extends invariant-based discovery to systems with imperfect symmetries. On benchmark PDEs, SBFT consistently outperforms standard and invariant-based GP under identical settings, discovering lower-error equations with closer form that are more useful in practice. While exact discovery remains challenging, SBFT makes meaningful progress, recovering governing equations that align more closely with underlying dynamics than baseline methods.

References

Akhound-Sadegh, T., Perreault-Levasseur, L., Brandstetter, J., Welling, M., and Ravanbakhsh, S. Lie point symmetry and physics informed networks. *arXiv preprint arXiv:2311.04293*, 2023.

Brandstetter, J., Welling, M., and Worrall, D. E. Lie point symmetry data augmentation for neural pde solvers. In *International Conference on Machine Learning*, pp. 2241–2256. PMLR, 2022.

Champion, K., Lusch, B., Kutz, J. N., and Brunton, S. L. Data-driven discovery of coordinates and governing equations. *Proceedings of the National Academy of Sciences*, 116(45):22445–22451, 2019. doi: 10.1073/pnas.1906995116.

Cranmer, M. Interpretable machine learning for science with pysr and symbolicregression.jl, 2023.

Mialon, G., Garrido, Q., Lawrence, H., Rehman, D., LeCun, Y., and Kiani, B. Self-supervised learning with lie symmetries for partial differential equations. *Advances in Neural Information Processing Systems*, 36:28973–29004, 2023.

Otto, S. E., Zolman, N., Kutz, J. N., and Brunton, S. L. A unified framework to enforce, discover, and promote symmetry in machine learning, 2023.

Takamoto, M., Praditia, T., Leiteritz, R., MacKinlay, D., Alesiani, F., Pflüger, D., and Niepert, M. Pdebench: An extensive benchmark for scientific machine learning. *Advances in Neural Information Processing Systems*, 35:1596–1611, 2022.

Udrescu, S.-M. and Tegmark, M. Ai feynman: A physics-inspired method for symbolic regression. *Science Advances*, 6(16):eaay2631, 2020.

Udrescu, S.-M., Tan, A., Feng, J., Neto, O., Wu, T., and Tegmark, M. Ai feynman 2.0: Pareto-optimal symbolic regression exploiting graph modularity. *Advances in Neural Information Processing Systems*, 33:4860–4871, 2020.

Wang, R., Walters, R., and Yu, R. Incorporating symmetry into deep dynamics models for improved generalization. In *International Conference on Learning Representations*, 2021.

Yang, J., Rao, W., Dehmamy, N., Walters, R., and Yu, R. Symmetry-informed governing equation discovery. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Yang, J., Bhat, M., Hu, B., Cao, Y., Dehmamy, N., Walters, R., and Yu, R. Discovering symbolic differential equations with symmetry invariants. *arXiv preprint arXiv:2505.12083*, 2025.