Shape-constrained Symbolic Regression with NSGA-III

EuroCAST 2022 - Theory and Applications of Metaheuristic Algorithms

Christian Haider – University of Applied Sciences Upper Austria
23.02.2022

https://symreg.at

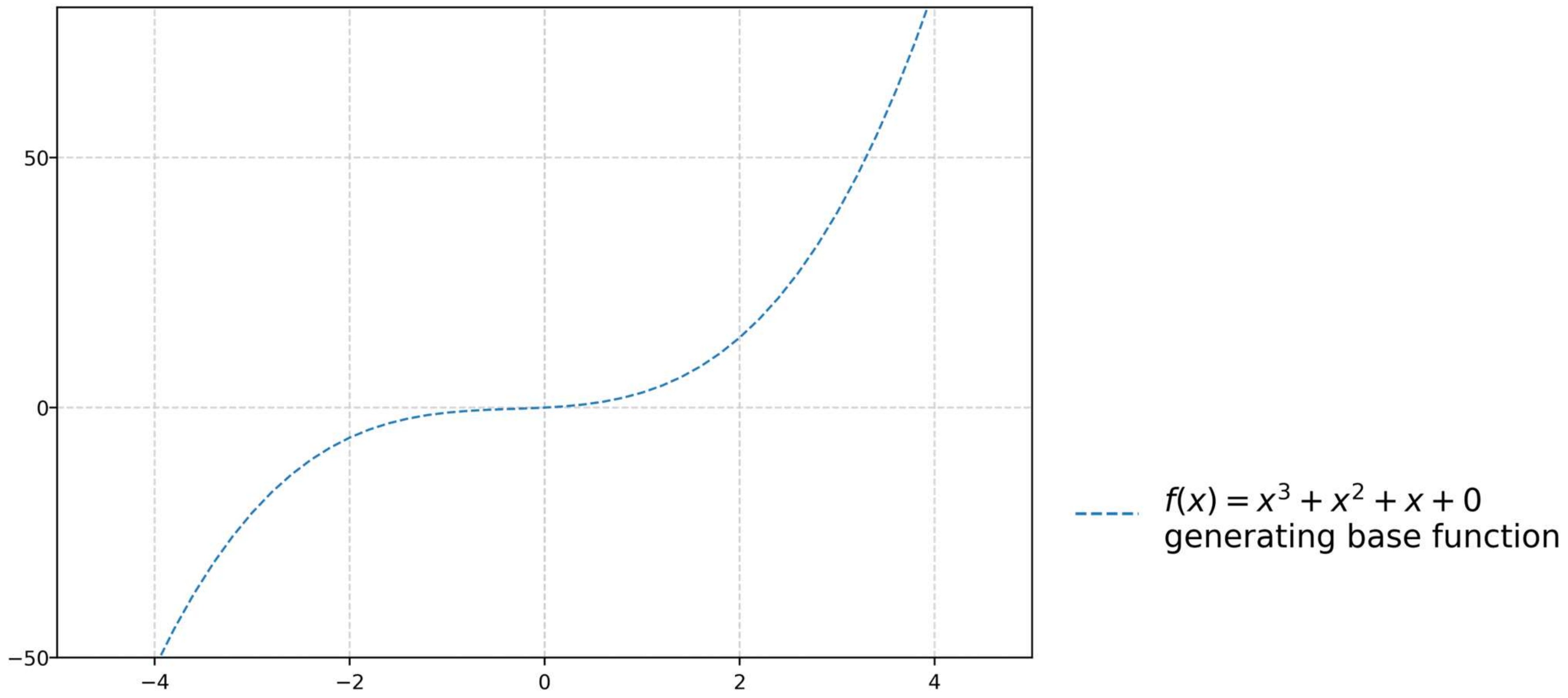
Agenda

- Motivation
- Shape-constrained symbolic regression
- Many-objective approach
- Results
- Conclusion

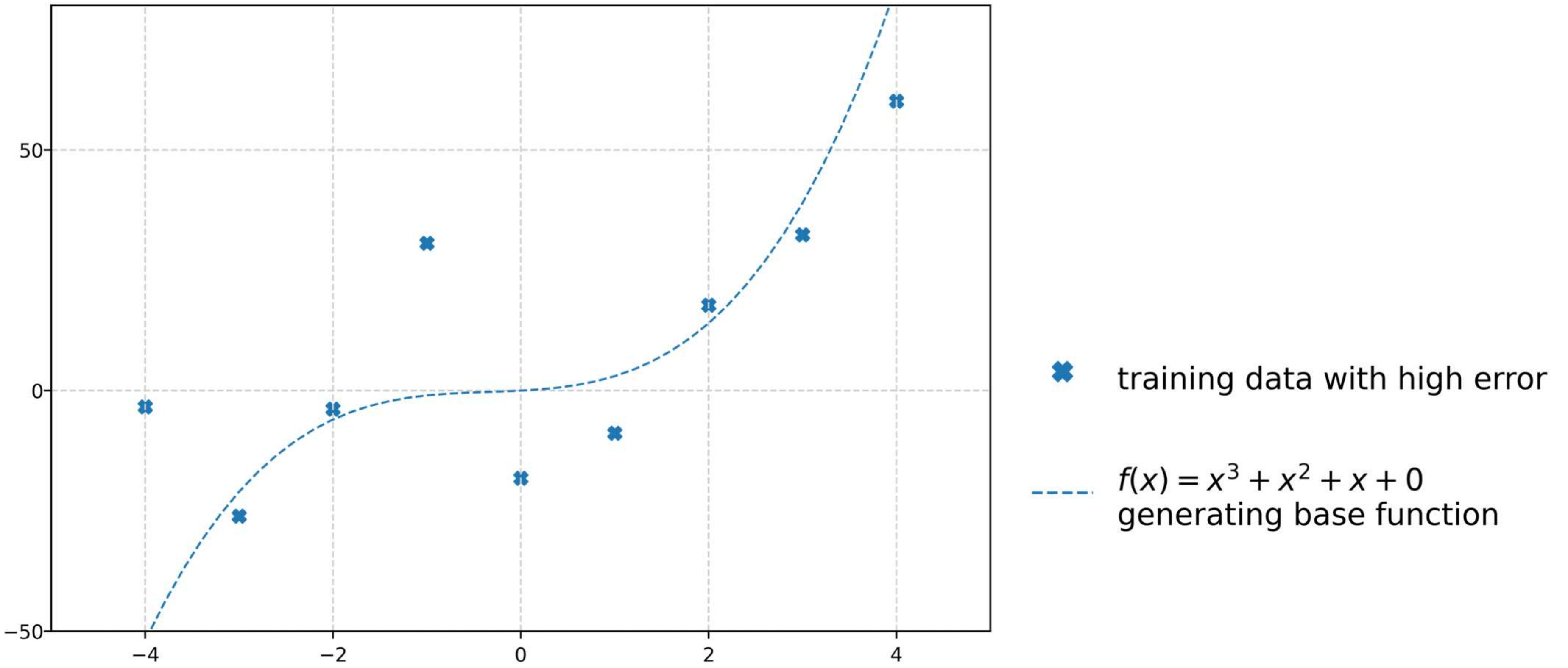
Motivation

- Data-based modeling approach
  - Allow to include knowledge about the system
- Restrict the function’s shape
  - Positivity/negativity of model output
  - Concavity/convexity or monotonicity of model’s output

Motivation

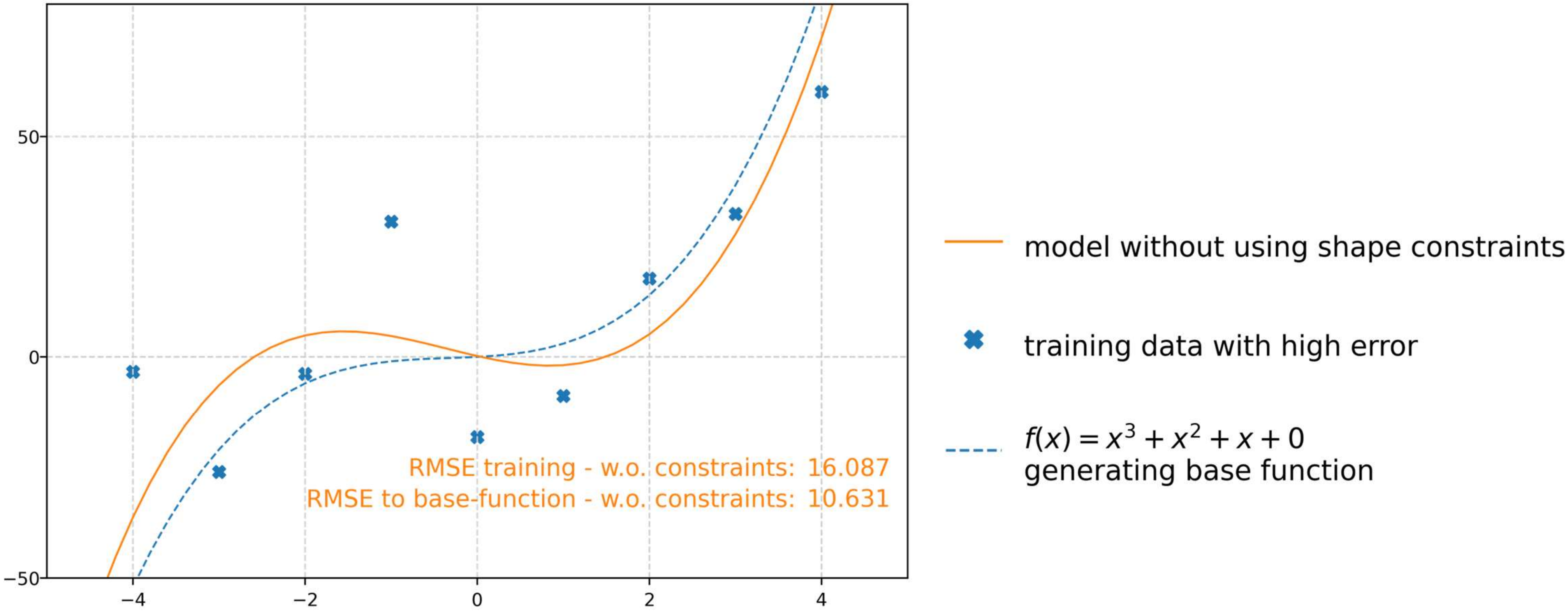


Motivation



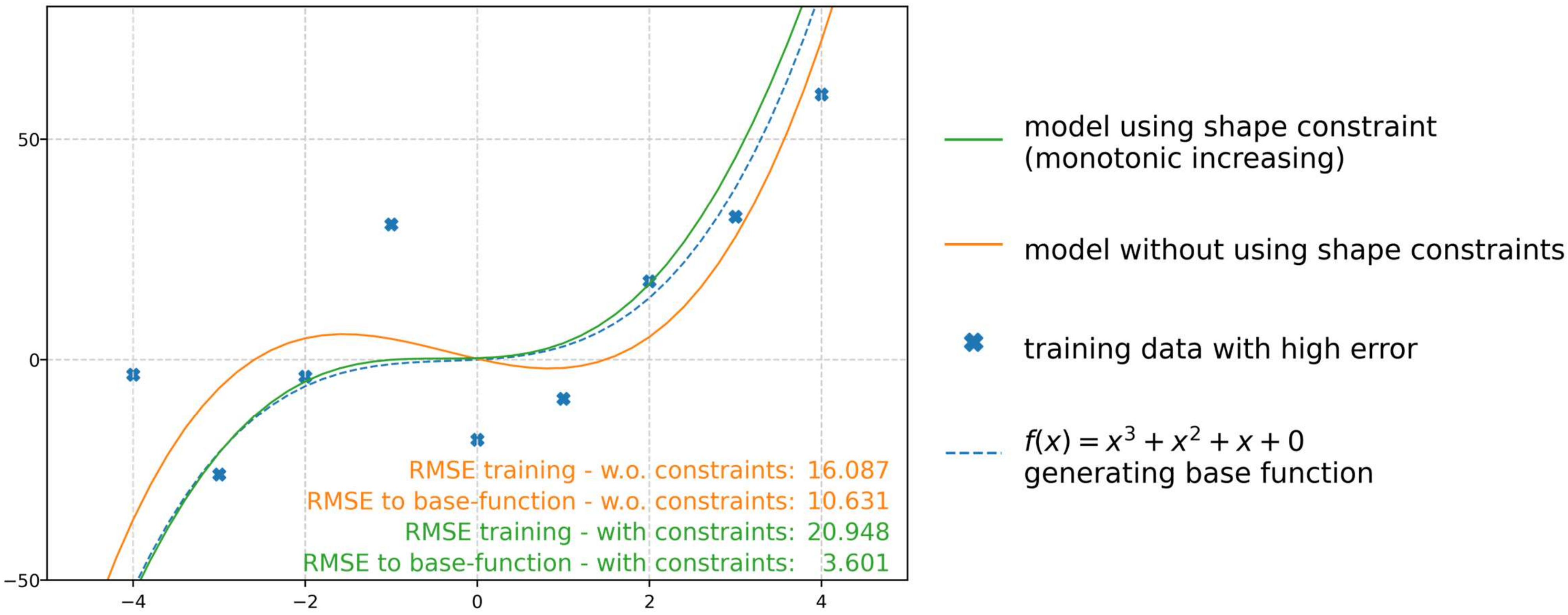
50
0
−50
−4
−2
0
2
4
training data with high error
$f(x) = x^3 + x^2 + x + 0$
generating base function

Motivation



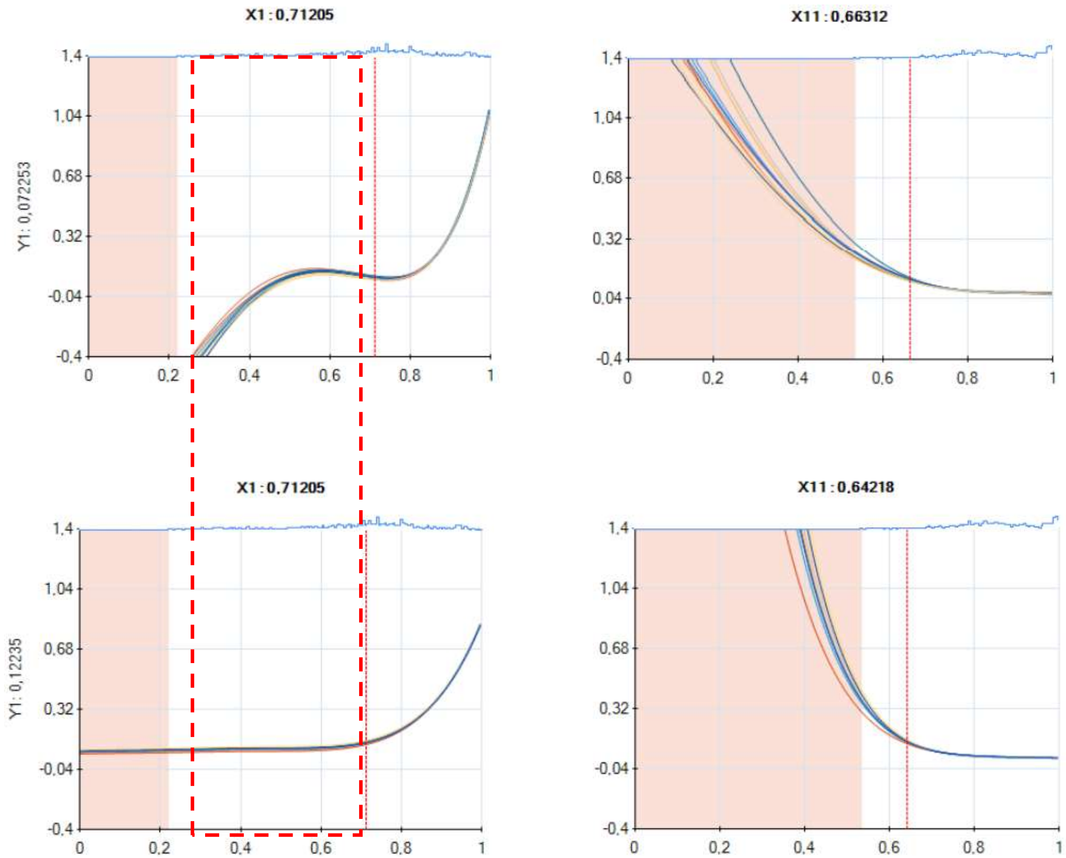
50
0
−50
−4
−2
0
2
4
RMSE training - w.o. constraints: 16.087
RMSE to base-function - w.o. constraints: 10.631
model without using shape constraints
training data with high error
$f(x) = x^3 + x^2 + x + 0$
generating base function

Motivation



50
0
−50
−4
−2
0
2
4
RMSE training - w.o. constraints: 16.087
RMSE to base-function - w.o. constraints: 10.631
RMSE training - with constraints: 20.948
RMSE to base-function - with constraints: 3.601
model using shape constraint (monotonic increasing)
model without using shape constraints
training data with high error
$f(x) = x^3 + x^2 + x + 0$ generating base function

Motivation



- First model
  - Violates positivity and monotonicity constraint over x_1
- Second model
  - Fulfills all constraints

Constraints:

$$M_{y1}(\boldsymbol{x}) \geq 0$$
$$\frac{\partial}{\partial x_1} M_{y1}(\boldsymbol{x}) \geq 0$$
$$\frac{\partial}{\partial x_{11}} M_{y1}(\boldsymbol{x}) \leq 0$$

Shape-constrained Symbolic Regression

- Model is a function mapping real-valued inputs to real-valued outputs
  - Constraints refer to the shape of the function
- Constraints are expressed by partial derivatives of the model

- Mathematical formulation of shape-constrained regression:

$$
\begin{aligned}
f^*(x) &= \underset{f(x)\in M}{\operatorname{argmin}} L(f(x), y), && x \in \Omega \\
&\text{subject to shape constraints } c_i(X_i), && X_i \subseteq \Omega
\end{aligned}
$$

Shape-constrained Symbolic Regression Constraints

Property	Mathematical formulation
Non-negativity	$f(x) \geq 0$
Non-positivity	$f(x) \leq 0$
Image inside a boundary	$l \leq f(x) \leq u$
Monotonically non-decreasing	$\frac{\partial}{\partial x_i} f(x) \geq 0$
Monotonically non-increasing	$\frac{\partial}{\partial x_i} f(x) \leq 0$
Convexity	$\frac{\partial^2}{\partial x_i^2} f(x) \geq 0$
Concavity	$\frac{\partial^2}{\partial x_i^2} f(x) \leq 0$

Shape-constrained Symbolic Regression Constraints

- Shape-constraints need approximation methods for evaluation
- E.g.: non-decreasing monotonicity constraint
  – Partial derivative of the model w.r.t the variable needs to be equal or greater than zero for the whole input domain ➔ minimum value of partial derivative in the given domain has to be found
  – If model is non-linear ➔ non-linear optimization problem (NP-hard)
  – Therefore approximation methods are useful
- One option: interval arithmetic

Interval Arithmetic

- Allows to apply arithmetic operations on intervals
  - Variables are bounded or unbounded intervals
- Intervals are described as $[a, b]$ a set of values $\{x \mid a \leq x \leq b\}$
- For a multivariate function $f(x)$ an exact interval can be estimated when:
  - Only a single occurrence of every variable x_i is given
  - The function is monotonic w.r.t any multiple occurring variable
- Small extra computational cost for evaluation compared to e.g. sampling
- Biggest disadvantage is the potential high overestimation if none of the exact interval estimation criteria is met
  - Occurring cause of dependency problem of multiple occurring variables

Shape-constrained Symbolic Regression Many-objectives

- Performs well on instances with many objectives
  - ≥ 3 Objectives
- *“The performance of NSGA-III has been compared with several versions of a recently proposed MOEA/D procedure. Although different MOEA/Ds have shown their working on different problems, no single version is able to solve all problems efficiently. Having solved all problems well by the proposed NSGA-III procedure, there is another advantage of it that is worth mentioning here. Unlike MOEA/D versions, NSGA-III procedure does not require any additional parameter to be set.”*[1]

[1]Showed promising results in *K. Deb and H. Jain, "An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach, Part I: Solving Problems With Box Constraints," in IEEE Transactions on Evolutionary Computation, vol. 18, no. 4, pp. 577-601, Aug. 2014, doi: 10.1109/TEVC.2013.2281535.*

Research Question

Can the many-objective algorithm NSGA-III help to improve the quality/speed of shape-constrained symbolic regression?

Methodology

- Compare NSGA-II to NSGA-III
- Both algorithms are implemented in HeuristicLab
- The same number of evaluations is used in both algorithms

Shape-constrained Symbolic Regression Many-objective

1+n objective approach:

- Data-base loss function ➔ minimizing NMSE
- Constraint loss functions ➔ minimize violations:

$$P_i = P_i^{\inf} + P_i^{\sup}$$
$$P_i^{\inf} = \left|\min\left(\inf\left(f_i(x)\right) - \inf(c_i), 0\right)\right|$$
$$P_i^{\sup} = \left|\max\left(\sup\left(f_i(x)\right) - \sup(c_i), 0\right)\right|$$

Problem Instances

Instance	Expression
I.6.20	$\exp\left(\frac{-\left(\frac{\theta}{\sigma}\right)^2}{2}\right)\frac{1}{\sqrt{2\pi}\sigma}$
I.9.18	$\frac{G\,m1\,m2}{(x2-x1)^2+(y2-y1)^2+(z2-z1)^2}$
I.30.5	$\operatorname{asin}\left(\frac{lambd}{nd}\right)$
I.32.17	$\frac{1}{2}\epsilon c\,Ef^2\,\frac{8\pi r^2}{3}\,\frac{\omega^4}{(\omega^2-\omega_0{}^2)^2}$
I.41.16	$\frac{h\,\omega^3}{\pi^2\,c^2\left(\exp\left(\frac{h\omega}{kbT}\right)-1\right)}$
I.48.20	$\frac{m\,c^2}{\sqrt{1-\frac{v^2}{c^2}}}$
II.35.21	$n_{rho}\,mom\,\tanh\left(\frac{momB}{kbT}\right)$
III.9.52	$\frac{p_d\,Ef\,t}{h}\sin\left(\frac{(\omega-\omega_0)t}{2}\right)^2$
III.10.19	$mom\,\sqrt{Bx^2+By^2+Bz^2}$
Pagie-1	$\frac{1}{1+x^{-4}}+\frac{1}{1+y^{-4}}$

Problem Instances

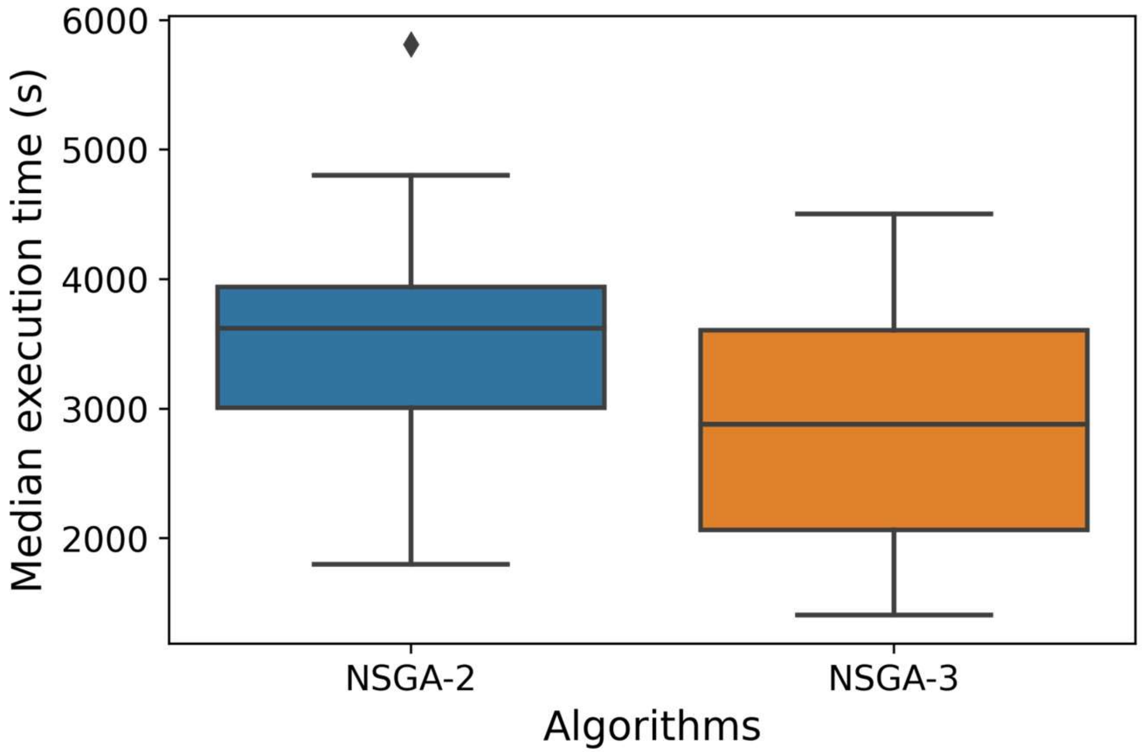
Instance	Input space	Constraints
I.6.20	$(\sigma, \theta) \in [1..3]^2$	$([0..\infty], 0, -1)$
I.9.18	$(x1, y1, z1, m1, m2, G, x2, y2, z2) \in [3..4]^3 \times [1..2]^6$	$([0..\infty], -1, -1, -1, 1, 1, 1, 1, 1, 1)$
I.30.5	$(lambd, n, d) \in [1..5]^2 \times [2..5]$	$([0..\infty], 1, -1, -1)$
I.32.17	$(\epsilon, c, Ef, r, \omega, \omega_0) \in [1..2]^5 \times [3..5]$	$([0..\infty], 1, 1, 1, 1, 1, -1)$
I.41.16	$(\omega, T, h, kb, c) \in [1..5]^5$	$([0..\infty], 0, 1, -1, 1, -1)$
I.48.20	$(m, v, c) \in [1..5] \times [1..2] \times [3..20]$	$([0..\infty], 1, 1, 1)$
II.35.21	$(n_{rho}, mom, B, kb, T) \in [1..5]^5$	$([0..\infty], 1, 1, 1, -1, -1)$
III.9.52	$(p_d, Ef, t, h, \omega, \omega_0) \in [1..3]^4 \times [1..5]^2$	$([0..\infty], 1, 1, 0, -1, 0, 0)$
III.10.19	$(mom, Bx, By, Bz) \in [1..5]^4$	$([0..\infty], 1, 1, 1, 1)$
Pagie-1	$(x, y) \in [-5..5]^2$	$([0..2], -1(x < 0), 1(x > 0), -1(y < 0), 1(y > 0))$

Algorithm Configurations

Parameters	Value
Function set	$+, \times, -, AQ(x,y) = x\sqrt{1+x^2}, \log, \exp, \sin, \tanh$
Terminal set	parameters
Max. tree length	50 nodes
Tree initialization	Probabilistic tree creator (PTC2)
Max. evaluated solutions	500000
Population size	1000 individuals
Selection	Tournament selection with group size 5 (NSGA2) Random selection (NSGA3)
Mutation rate	15%

Results: Median Runtime [sec]

	NSGA2	NSGA3
I.6.20	1798.02	**1407.67**
I.30.5	3621.99	**3604.91**
I.41.16	3858.61	**2879.05**
I.48.20	2825.38	**1647.43**
II.32.17	5812.10	**4504.23**
II.35.21	3217.67	**3045.41**
III.9.52	3009.62	**2064.16**
III.10.19	3939.14	**2254.29**
Pagie-1	4800.77	**4105.86**



Results: Median Test Error (NMSE in %)

	NSGA2	NSGA3
I.6.20	20.88	**19.14**
I.30.5	7.32	**6.24**
I.41.16	18.50	**15.21**
I.48.20	24.19	**22.58**
II.32.17	7.17	**6.38**
II.35.21	14.60	**14.54**
III.9.52	89.03	**89.00**
III.10.19	11.30	**10.62**
Pagie-1	46.41	**40.71**

$$\mathrm{NMSE}(y,\hat{y}) = \frac{100}{\mathrm{var}(y)N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2$$

Summary

- NSGA-III decreased runtime of experiments by about 60%
- Equal or slightly improved model quality

Outlook:

- Further tests on more instances
- Hyperparameter optimization for both algorithms

Shape-constrained Symbolic Regression with NSGA-III

EuroCAST 2022 - Theory and Applications of Metaheuristic Algorithms

Christian Haider – University of Applied Sciences Upper Austria

23.02.2022

https://symreg.at