# Evolving Differential Equations with Developmental Linear Genetic Programming and Epigenetic Hill Climbing

William La Cava University of Massachusetts Amherst, MA 01003 wlacava@umass.edu Lee Spector Hampshire College Amherst, MA 01002 Ispector@hampshire.edu Kourosh Danai University of Massachusetts Amherst, MA 01003 danai@ecs.umass.edu

### ABSTRACT

This paper describes a method of solving the symbolic regression problem using developmental linear genetic programming (DLGP) with an epigenetic hill climber (EHC). We propose the EHC for optimizing the epigenetic properties of the genotype. The epigenetic characteristics are then inherited through coevolution with the population. Results reveal that the EHC improves performance through maintenance of smaller expressed program sizes. For some problems it produces more successful runs while remaining essentially cost-neutral with respect to number of fitness evaluations.

#### **Categories and Subject Descriptors**

I.2.2 [Artificial Intelligence]: Automatic Programming— Program synthesis; J.2 [Computer Applications]: Physical Science and Engineering—Engineering

#### **General Terms**

Algorithms

#### Keywords

Genetic programming; epigenetics; differential equations; system identification

### 1. LINEAR GP WITH EPIGENESIS

Genetic programming (GP) has become a popular method for searching equation space and finding nonlinear differential equation solutions [3, 2, 4]. GP traditionally represents individuals using tree structures. Here we use a syntaxfree linear representation to conduct GP. This encoding is meant to maximize flexibility of the representation, simplify genetic operations, and allow for easy implementation of epigenetics. The term epigenetics is used broadly to refer to the

*GECCO'14*, July 12–16, 2014, Vancouver, BC, Canada. ACM 978-1-4503-2881-4/14/07. http://dx.doi.org/10.1145/2598394.2598491 . ways in which gene expression is developed and inherited [5]. Methylation [6] silences clusters of genetic material, allowing separation of genotype from the phenotype and the creation of non-coding segments (i.e. introns). Moderate levels of introns can improve EA performance by reducing the destructive effect of crossover operations while maintaining blocks of effective code [7, 1, 8]. We implemented two characteristics of methylation: 1) dependence on environmental factors by use of the EHC, and 2) inheritability by coevolution of epilines with their corresponding genotypes.

Equations are represented by linear genotypes, as demonstrated in Figure 1. In addition to having a genotype composed of a list of instructions, the EHC creates a binary array of equivalent length in each individual, referred to as an epiline. During genotype to phenotype conversion, only instructions from the list with a true value in the corresponding epiline are executed. Each generation, the population undergoes one iteration of EHC to optimize the epiline.

The results are summarized in Table 1. Using DLGP alone found exact solutions in 25 out of 30 runs (83.33%) for both the mass spring damper and van der Pol oscillator. After adding epigenetic co-evolution and the EHC, the program succeeded for 100% of the runs for both cases. This improvement in success rate was achieved without a significant increase in computational effort, as shown in Table 1. While the total genotypic size remains similar to DLGP, with EHC activated the effective size of the programs (the number of executed instructions) was 33% shorter for the mass spring damper and 28% shorter for the van der Pol oscillator.

Using the Pagie-1 problem as a benchmark, we varied the average initial percent of genes that were turned on and the switching rate, denoted by "i(percent)" and by "t(percent)" in Figure 2, and ran 50 trials of each setting. There was not a statistically significant improvement in Mean Best Error, but we saw improvements in effective program size and beneficial genetics (genetic operations that result in fitter offspring). This agrees with previous literature showing the beneficial effects that introns can have [1, 7]. In all of these cases the effective program sizes throughout the run were significantly lower than without the hill climber, indicating that it may be a good way to combat the problem of code growth and contribute to equation parsimony.

#### 2. ADDITIONAL AUTHORS

Additional authors: Matthew Lackner (University of Massachusetts Amherst, email: lackner@ecs.umass.edu).

<sup>\*</sup>corresponding author

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage, and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

Table 1: Performance Comparisons. No evaluation differences are significant to less than .05 using the two-tailed t-test  $p_t$  or the non-parametric ranked t-test [9],  $p_r$ . All size differences are  $(p_t < .001, p_r < .001)$ .

Problem	Trials	Method	Success Rate	Evaluations	Mean Effective Size
$\ddot{x} = -1/2(0.75\dot{x} + 3x - F)$	30	DLGP	83.33%	8.83E05	145.11
	30	DLGP+EHC	100%	10.77 E05	96.65
$\ddot{x} = -1.5(x^2 + 1)\dot{x} - x$	30	DLGP	83.33%	9.20E05	140.06
	30	DLGP+EHC	100%	9.27 E05	101.34
$\dot{x} = 3x - 2xy - x^2$	50	DLGP	100%	8.9202 E05	29.63
	50	DLGP+EHC	100%	8.0602 E05	24.69
$\dot{y} = 2y - xy - y^2$	50	DLGP	100%	8.1724 E05	30.3595
	50	DLGP+EHC	100%	8.6847 E05	25.1297

Epiline	Genotype	Stack	Tree
1	X	х	
0	У	х	<u> </u>
1	3	x 3	$\bigtriangledown$
1	-	(x-3)	
0	4	(x-3)	- $(*)$
1	z	(x-3) z	$(\vec{x})$ $(\vec{3})$ $(\vec{z})$ $(\vec{4})$
1	4	(x-3) z 4	
0	+	(x-3) z 4	(2) $(4)$
1	*	(x-3) (z*4)	
1	/	((x-3)/(z*4)	))
1	+	((x-3)/(z*4))	)) } Phenotype

Figure 1: Epigenetics added to the original encoding of  $f = x + \frac{y-3}{4(z+4)}$ , which in this example results in  $f = \frac{x-3}{(z+4)}$ .



Figure 2: Beneficial crossover percentages for the Pagie-1 problem with various settings of the EHC.

## 3. ACKNOWLEDGEMENTS

This work is partially supported by the NSF-sponsored IGERT: Offshore Wind Energy Engineering, Environmental Science, and Policy (Grant Number 1068864), as well as Grant No. 1017817. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the National Science Foundation.

#### 4. **REFERENCES**

- [1] Markus Brameier and Wolfgang Banzhaf. *Linear Genetic Programming*, volume Vol. 1. Springer, 1 edition, 2007.
- [2] Hongqing Cao, Lishan Kang, Yuping Chen, and Jingxian Yu. Evolutionary modeling of systems of ordinary differential equations with genetic programming. *Genetic Programming and Evolvable Machines*, 1(4):309–337, 2000.
- [3] Gary J. Gray, David J. Murray-Smith, Yun Li, Ken C. Sharman, and Thomas Weinbrenner. Nonlinear model structure identification using genetic programming. *Control Engineering Practice*, 6(11):1341–1352, 1998.
- [4] Hitoshi Iba. Inference of differential equation models by genetic programming. *Information Sciences*, 178(23):4453-4468, 2008. Special Section: Genetic and Evolutionary Computing.
- [5] Eva Jablonka and Marion J. Lamb. The changing concept of epigenetics. Annals of the New York Academy of Sciences, 981(1):82–96, 2002.
- [6] Peter A. Jones and Daiya Takai. The role of DNA methylation in mammalian epigenetics. *Science*, 293(5532):1068–1070, August 2001. PMID: 11498573.
- [7] Peter Nordin, Frank Francone, and Wolfgang Banzhaf. Explicitly defined introns and destructive crossover in genetic programming. In Justinian P. Rosca, editor, *Proceedings of the Workshop on Genetic Programming: From Theory to Real-World Applications*, pages 6–22, Tahoe City, California, USA, 1995.
- [8] I Tanev and K Yuta. Epigenetic programming: Genetic programming incorporating epigenetic learning through modification of histones. *Information Sciences*, 178(23):4469–4481, December 2008.
- [9] Mark Wineberg and Steffen Christensen. An introduction to statistics for EC experimental analysis, 2004.