Grammatical Evolution: STE criterion in Symbolic Regression Task

R. Matousek, *Member, IAENG*

Abstract — **Grammatical evolution (GE) is one of the newest among computational methods (Ryan et al., 1998), (O'Neill and Ryan, 2001). Basically, it is a tool used to automatically generate Backus-Naur-Form (BNF) computer programmes. The method's evolution mechanism may be based on a standard genetic algorithm (GA). GE is very often used to solve the problem of a symbolic regression, determining a module's own parameters (as it is also the case of other optimization problems) as well as the module structure itself. A Sum Square Error (SSE) method is usually used as the testing criterion. In this paper, however, we will present the original method, which uses a Sum Epsilon Tube Error (STE) optimizing criterion. In addition, we will draw a possible parallel between the SSE and STE criteria describing the statistical properties of this new and promising minimizing method.**

Index Terms — **Grammatical Evolution, SSE, STE, Epsilon Tube, Laplace Distribution.**

I. Introduction

The general problem of optimizing a nonlinear model or a specific problem of nonlinear regression in statistics may be seen as typical problems that can successfully be approached using evolutional optimization techniques. A genetic-algorithm method, for example, may be very efficient in determining the parameters of models representing a multimodal or non-differentiable behaviour of the goal function. On the other hand, there are numerous mathematical methods that may, under certain conditions, be successful in looking for a minimum or maximum. The classical optimization methods may differ by the type of the problem to be solved (linear, nonlinear, integer programming, etc.) or by the actual algorithm used. However, in all such cases, the model structure for which optimal parameters are to be found is already known. But it is exactly the process of an adequate model structure design that may be of key importance in statistical regression and numerical approximation.

When a problem of regression is dealt with, it is usually split into two parts. First, a model is designed to fit the character of the empirically determined data as much as possible. Then its optimum parameters are computed for the whole to be statistically adequate. In the event of a pure approximation problem, the situation is the same: a model is designed (that is, its structure such as a polynomial $P(x)=ax^2+bx+c$) with its parameters being subsequently optimized (such as a, b, c) using a previously chosen minimization criterion, which is usually a Sum Square Error (SSE) one.

Manuscript received July 25, 2009. The results presented have been achieved using a subsidy of the research project MSM 0021630529 “Intelligent Systems in Automation“ and by the Grant Agency of the Czech Republic No.: 102/091668 “Control Algorithm Design by means of Evolutionary Approach”.

The author is with Brno University of Technology, Czech Republic. They are now with the Institute of Automation and Computer Science, Faculty of Mechanical Engineering, Technicka 2, 61669 Brno (e-mail: matousek@fme.vutbr.cz).

GE uses such data representation as to deal with the above problem in a comprehensive way, that is, combining the determination of a model structure and finding its optimum parameter in a single process. Such a process is usually called the problem of a symbolic regression. The GE optimisation algorithm usually uses a GA with the genotype-phenotype interpretation using a context-free Backus-Naur-Form based grammar [1, 2]. The problem solved is then represented by a BNF structure, which a priori determines its scope – an automatic computer program generating tool. Further we will focus on issues of symbolic regression of trigonometric and polynomial data, that is, data that can be described using a grammar containing trigonometric functions and/or powers.

II. Parameters used by GE

The advanced GE algorithm implemented was using a binary GA. The GA implemented the following operators: tournament selection, one- and two-point structural crossing-over, and structural mutations. Numerous modern methods of chromosome encryption and decryption were used [3, 4]. The following grammar G was used with respect to the class to be tested.

Table I: Grammar G which was used with respect of the approximation tasks

Approximation task	Grammar and example of generating function
Trigonometric	G = {+,-,*,/,sin,cos,variables,constants} notes: *unsigned integer constants* ∈ [0,15], *real constants* ∈ [0,1] $y = \sin(x) + \sin(2.5x)$
Polynomial	G = {+,-,*,/, variables, constants} notes: *unsigned integer constants* ∈ [0,15], *real constants* ∈ [0,1] $y = 3x^4 - 3x + 1$

III. STE – Sum Epsilon Tube Error

To keep further description as simple as possible, consider a data approximation problem ignoring the influence of a

ISBN:978-988-18210-2-7

random variable, that is, disregarding any errors of measurement. The least-square method is used to test the quality of a solution in an absolute majority of cases. Generally, the least-squares method is used to find a solution with the squared errors r summed over all control points being minimal. A control point will denote a point at which the approximation error is calculated using the data to be fitted, that is, the difference between the actual y value and the approximation $\hat{y}$, ($r = y - \hat{y}$).

$$SSE = \sum_i r_i^2 \tag{1}$$

However, in the event of a GE problem of symbolic regression, this criterion may not be sufficient. For this reason a new, Sum Epsilon-Tube Error (STE) evaluation criterion has been devised. For each value ε, this evaluation criterion may be calculated as follows:

$$STE_\varepsilon = \sum_i e_\varepsilon(r_i), \qquad e_\varepsilon(r_i) = \begin{cases} 0 & r_i \notin [-\varepsilon, \varepsilon] \\ 1 & r_i \in [-\varepsilon, \varepsilon] \end{cases} \tag{2}$$

where ε … is the radius of the epsilon tube,
i … is the index set of the control points,
e … is an objective function determining whether the approximated value $\hat{y}$ lies within the given epsilon tube.

The workings of this criterion may be seen as a tube of diameter epsilon that stretches along the entire approximation domain (see Fig. 1). The axis of such a tube is determined by the points corresponding to the approximated values. Such points may be called control points.

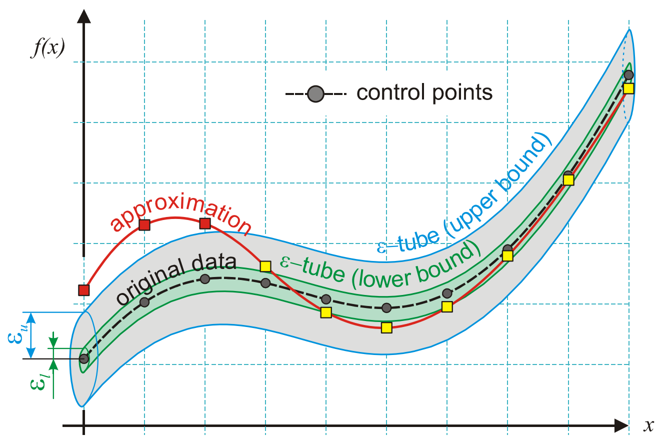

Fig.1: The principle of the Epsilon Tube

The actual evaluating function then checks whether an approximation point lies within or out of the tube (2). The epsilon value changes dynamically during the optimization process being set to the highest value at the beginning and reduced when the adaptation condition is met. The variable parameters of the STE method are listed in Table II. At the beginning of the evolution process, the algorithm sets the diameter of the epsilon tube. If condition (3), which indicates the minimum number of points that need to be contained in the epsilon cube, is met, the epsilon tube is adapted, that is, the current epsilon value is reduced. This value can either be reduced using an interest-rate model with the current epsilon value being reduced by a given percentage of the previous epsilon value or using a linear model with the current epsilon value being reduced by a constant value.

Table II: Parameters of the STE Minimization Method

Parameter	Value	Description
`steEpsStart`	Positive real number	Epsilon initial value. (ε-upper bound)
`steEpsStop`	Positive real number	Final epsilon value. (ε-lower bound)
`steEpsReduction`	Unsigned integer	Percentage of value decrease on adapting the value.
`steAdaptaion`	Unsigned integer	Percentage indicating the success rate (control points are in the ε tube) at which epsilon is adapted.
`steModel`	"Interest-Rate Model" "Linear Model"	Model of the computation of a new epsilon value if adaptation condition is met.
`steIteration`	Unsigned integer	This may replace the `steEpsStop` parameter, and then this parameter indicates the number of adaptations until the end of the computation.

At the beginning of the evolution process, the algorithm sets the diameter of the epsilon tube. If condition (3), which indicates the minimum number of points that need to be contained in the epsilon cube, is met, the epsilon tube is adapted, that is, the current epsilon value is reduced. This value can either be reduced using an interest-rate model with the current epsilon value being reduced by a given percentage of the previous epsilon value or using a linear model with the current epsilon value being reduced by a constant value. The condition for the adaptation of the epsilon tube to be finished may either be chosen as a limit epsilon value (the `steEpsStop` parameter) or be given by the number of adaptive changes (the `steIteration` parameter).

$$\frac{c}{STE_\varepsilon} \cdot 100 \geq steAdaptation \tag{3}$$

Where c is the number of control points.

IV. STE – Empirical Properties

Practical implementations of the symbolic regression problem use STE as an evaluating criterion. The residua obtained by applying STE to the approximation problems implemented have been statistically analyzed. It has been found out that these residua are governed by the Laplace distribution (4) with $\mu = 0$ and $b = 1/\mathrm{sqrt}(2)$ in our particular case. A random variable has a Laplace (μ, b) distribution if its probability density function is

$$f(x \mid \mu, b) = \frac{1}{2b}\exp\left(-\frac{|x-\mu|}{b}\right) = \frac{1}{2b}\begin{cases} \exp\left(-\frac{\mu - x}{b}\right) & \text{if } x < \mu \\ \exp\left(-\frac{x-\mu}{b}\right) & \text{if } x \geq \mu \end{cases} \tag{4}$$

ISBN:978-988-18210-2-7

Here, μ is a location parameter and b > 0 is a scale parameter. If $\mu = 0$ and $b = 1$, the positive half-line is exactly an exponential distribution scaled by ½.

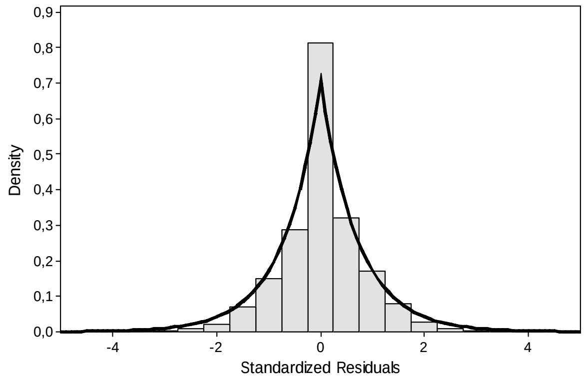

Fig 2: Histogram of the standardized residua (from the experiment) and the probability density function of the Laplace distribution.

An intuitive approach to approximation shows better results for the STE method mainly because there are no big differences between the approximating functions (see Fig. 1, Fig. 3). The reason for this is obvious: very distant points influence the result of approximation a great deal for SSE while, for STE, the weight of such points are relatively insignificant.

Next, the main advantages and disadvantages are summarized of the minimization methods using the SSE and STE criteria.

SSE (advantages, disadvantages):

- + a classical and well researched method both in statistical regression and numerical approximation,
- + a metric is defined,
- + a N($\mu = 0$, σ^2) residua error distribution may be assumed with descriptive statistics for this distribution being available,
- – more time-consuming,
- – being used excessively, this method hardly provides an incentive for users to gain new insights [5].

STE (advantages, disadvantages):

- + less time consuming,
- + more intuitive and readable,
- + when using a GE-based symbolic regression, the Laplace distribution may be assumed (which is one of the original results of this paper),
- + better results are generated for the problem class,
- – mathematical processing is worse with no metric defined
- – the method being new, it is not so well described

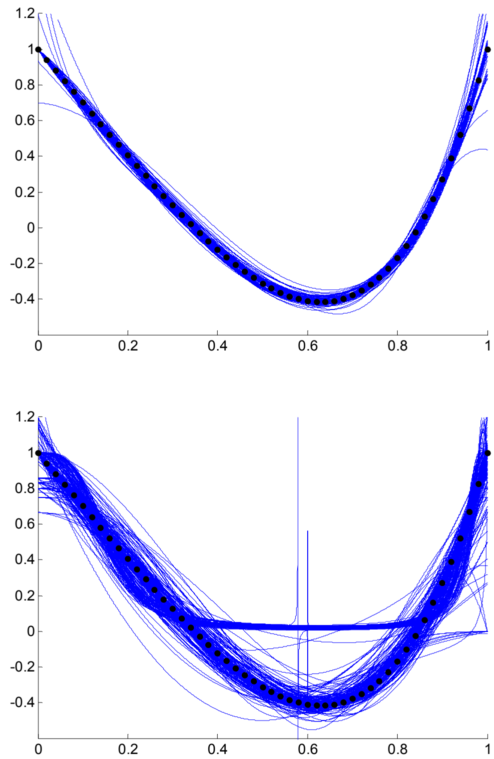

Fig 3: The final approximations (symbolic regression) of polynomial data (see Tab. I). STE method was used (above), respectively SEE method was used (down).

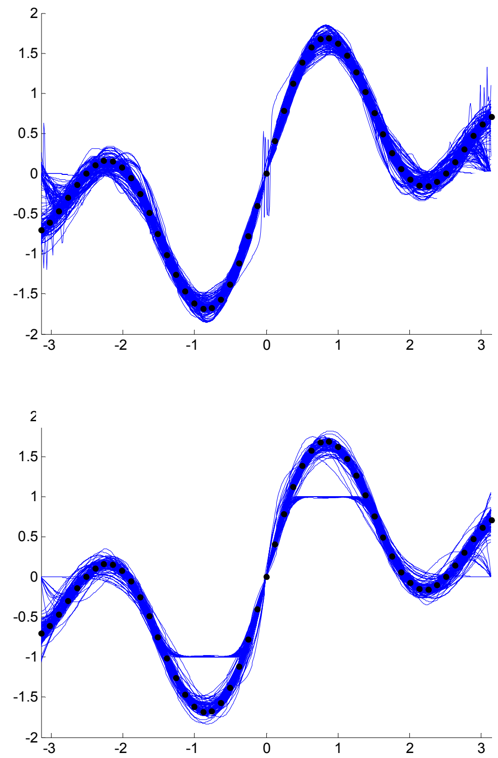

Fig 4: The final approximations (symbolic regression) of trigonometric data (see Tab. I). STE method was used (above), respectively SEE method was used (down).

V. Probabilistic Mapping of SSE to STE

This part shows how the results obtained by a SSE-based minimization method may be used to derive a result corresponding to the values of a STE-based minimization criterion. In other words, a procedure will be shown for using SSE to obtain a corresponding STE. The procedure is based on the following:

- It is assumed that, when applying the least-squares method, the residua are normally distributed with N(μ = 0, σ^2). Here, the standard deviation may be arbitrary, but a reduction to the standardized normal distribution N(0, 1) is possible.
- In order to reach a sufficient statistical significance, the experiment simulated 10,000 instances of solution to an approximation problem (nRUN). This problem was discretized using 100 control points.
- Different sizes eps ={2,1,0.5,0.25,0.125} of epsilon tubes were chosen to simulate possible adaptations of the epsilon tube by (3).
- The frequencies were calculated of the values lying within the given epsilon tubes.

This simulation and the subsequent statistical analysis were implemented using the Minitab software package. Table III provides partial information on the procedure used.

Table III: A fragment of a table of the simulated residua values with standardized normal distribution and the corresponding $e_\varepsilon(r_i)$ values.

r	i	nRUN	$\varepsilon=2$	$\varepsilon=1$	$\varepsilon=0.5$	$\varepsilon=0.25$	$\varepsilon=0.125$
0.48593	1	1	1	1	1	0	0
-0.07727	2	1	1	1	1	1	1
1.84247	3	1	1	0	0	0	0
-1.29301	4	1	1	0	0	0	0
0.34326	5	1	1	1	1	0	0
…	…	…	…	…	…	…	…
0.02298	100	10 000	1	1	1	1	1

where: r … residua°($X \sim N(0,1)$),
i … control point index ($i \in$ [1,100] in our particular case),
nRUN …given implementation (nRUN $\in$ [1,10000] in our particular case).

Next, for each residuum value, the number of cases will be determined of these residua lying in the interval given by a particular epsilon tube. This evaluation criterion denoted by STE (Sum epsilon-Tube Error) is defined by (2) and, for the particular test values, corresponds to (5):

$$STE_\varepsilon = \sum_{i=1}^{100} e_\varepsilon(r_i), \qquad \varepsilon = \{2,1,0.5,0.25,0.125\} \quad (5)$$

By a statistical analysis, it has been determined that STE behaves as a random variable with a Binomial distribution

$$K \sim Bi(n, p_\varepsilon), \text{ where } p_\varepsilon = P\left(X \sim N(0,1) \in [-\varepsilon, \varepsilon]\right) \quad (6)$$

For the epsilons chosen, the STE characteristic has a binomial distribution given by (6) where the probabilities p_ε are calculated from Table IV. For the case of $\varepsilon = 2$, Figure 5 shows the probability calculation.

Table IV: Table of probabilities $p_\varepsilon = P\left(N(0,1) \in [-\varepsilon, \varepsilon]\right)$

ε	$P\left(N(0,1) < -\varepsilon\right)$	$P\left(N(0,1) < \varepsilon\right)$	p_ε
2	0.022750	0.977250	0.954500
1	0.158655	0.841345	0.682689
0.5	0.308538	0.691462	0.382925
0.25	0.401294	0.598706	0.197413
0.125	0.450262	0.549738	0.099476

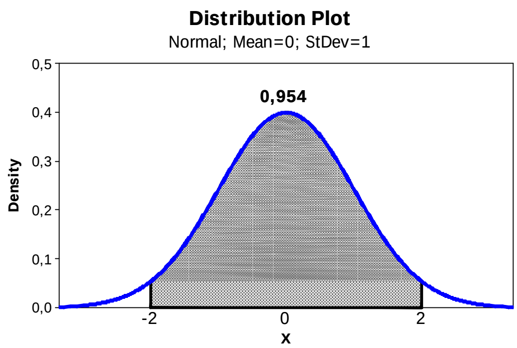

Fig 5: Calculating probability p_ε with $\varepsilon = 2$.

The figure below shows the probability functions for epsilon tubes with value k indicating, with a given probability, the number of control points for which the $e(r)$ function takes on a value of one, that is, the number of control points contained in a given epsilon tube.

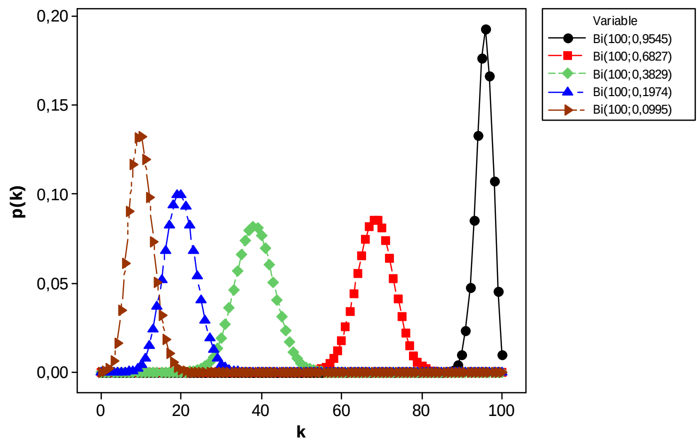

Fig 6: STE probability density functions for epsilon tubes with ε = {2,1,0.5,0.25,0.125}.

The correctness of the calculation was proved empirically. The STE values simulated (10 000 runs) were compared with distribution (6). The following figure (Fig. 7) displays an empirical probability density function obtained from Table III by (5) compared with the probability density function as calculated by (6) for $\varepsilon = 1$.

ISBN:978-988-18210-2-7

It can be seen in Figure 7 that the empirical distribution is close the calculated one, which was verified using a goodness-of-fit test.

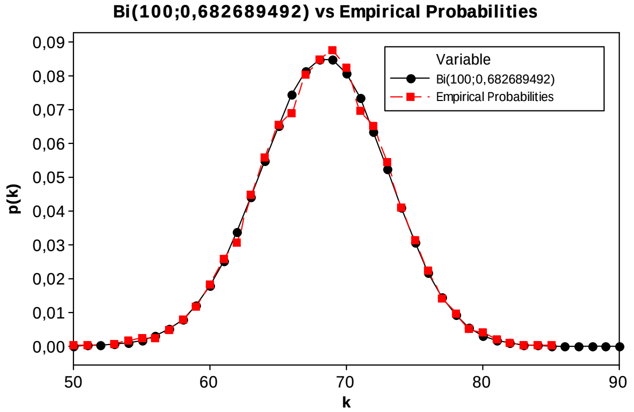

Fig 7: Comparing empirical and calculate probability density functions of occurrences within the epsilon tube for $\varepsilon = 1$.

VI. CONCLUSION

- A new STE evaluation criterion was introduced using epsilon tubes. Based on the experiments conducted, this criterion seems to be more appropriate for the symbolic regression problem as compared with the SSE method.
- Using statistical tools, a method was shown of transforming the results obtained by SSE minimization into results obtained by STE.
- Using a particular class implemented using approximations obtained by an STE minimization, the empirically obtained residua were found to be Laplace distributed.
- The properties of such residua and transformation of the STE-based minimization criterion into an SSE-based one will be the subject of further research.

REFERENCES

[1] O'Neill, M., Ryan, C.: Grammatical Evolution: Evolutionary Automatic Programming in an Arbitrary Language, *Kluwer Academic Publisher*, London, USA, ISBN 1-4020-7444-1 (2003)

[2] Ryan, C., O'Neill, M., Collins, J.J.: Grammatical Evolution: Solving Trigonometric Identities, In *proceedings of MENDEL '98*, Brno, Czech Republic, pp. 111-119, ISBN 80-214-1199-6 (1998)

[3] Osmera, P., Matousek, R., Popelka, O., Panacek, T.: Parallel Grammatical Evolution for Circuit Optimization, In *proceedings of MENDEL 2008*, Brno, Czech Republic, pp. 206-213, ISBN 978-80-214-3675-6 (2008)

[4] Matousek, R.: GAHC: Hybrid Genetic Algorithm. In *the Springer book series (eds:. Ao, S.L., Rieger, B., Chen, S.S.) Lecture Notes in Electrical Engineering: Advances in Computational Algorithms and Data Analysis*, Volume 14, ISSN 1876-1100, ISBN 978-1-4020-8918-3, pp. 549-562, Springer Netherlands (2008)

[5] Popela, P.: Numerical Techniques and Available Software, *Chapter 8 in Part II, J. Dupa_cov_a, J. Hurt, J. St_ep_an: Stochastic Modeling in Economics and Finance*, pp.206-227, Kluwer Academic Publishers