An inference model for analyzing termination conditions of Evolutionary Algorithms

David Roche [a,1], Debora Gil [b] and Jesus Giraldo [a]

[a] *Laboratory of Systems Pharmacology and Bioinformatics, UAB, Bellaterra,Spain*
[b] *Computer Science Department and Computer Vision Center (CVC), UAB, Bellaterra,Spain*

Abstract

In real-world problems, it is mandatory to design a termination condition for Evolutionary Algorithms (EAs) ensuring stabilization close to the unknown optimum. Distribution-based quantities are good candidates as far as suitable parameters are used. A main limitation for application to real-world problems is that such parameters strongly depend on the topology of the objective function, as well as, the EA paradigm used.

We claim that the termination problem would be fully solved if we had a model measuring to what extent a distribution-based quantity asymptotically behaves like the solution accuracy. We present a regression-prediction model that relates any two given quantities and reports if they can be statistically swapped as termination conditions. Our framework is applied to two issues. First, exploring if the parameters involved in the computation of distribution-based quantities influence their asymptotic behavior. Second, to what extent existing distribution-based quantities can be asymptotically exchanged for the accuracy of the EA solution.

Keywords. Evolutionary Computation Convergence, Termination Conditions, Statistical Inference

Introduction

Evolutionary Algorithms (EAs) are a class of stochastic optimization methods that simulate the process of natural evolution [1]. EAs maintain a population of possible solutions that evolve according to rules of selection and other operators, such as recombination and mutation. By their ability to optimizing non-analytic multi-modal functions, EAs have been successfully applied to a wide range of real life problems [2,3,4].

As any iterative technique, EA requires a stop criterion. Unlike optimization methods evolving a single initial value (which rely on real analysis theory), by their stochastic nature, there is not a solid mathematical theory ensuring convergence of evolutionary methodologies in general [5,1]. The simplest (an most extended [5,6,7]) stopping criterion consists in reaching a number of iterations or function evaluations. This stopping

[1]Laboratory of Systems Pharmacology and Bioinformatics, UAB, Bellaterra,Spain; E-mail: david.roche@uab.es.

criterion is not useful by itself (the number of iterations that guarantee convergence significantly varies across problems [5]), though it can be necessary when used in addition with alternative criteria to ensure that the algorithm stops [8].

Existing approaches defining alternative termination conditions [9,10] address the definition of a quantity reflecting the amount of change between consecutive iterations and the condition that such quantity should fulfill. Two different termination conditions are considered in the literature. Terminate either if the measure of the amount of change is below a given threshold or in the case that such measure is below a threshold for a number of generations. Concerning the amount of change, among the quantities reported in the literature [8,5], distribution-based are the best suited. These quantities measure the distribution of a given percentage of the (best) evolving population. They compare to the accuracy of the solution (distance to the optimum) in terms of number of function executions, as far as suitable parameters are set [8].

Two different sets of parameters are involved in distribution-based termination conditions. On one hand, the parameters used for the termination condition itself: threshold and number of generations. On the other hand, those involved in the computation of the distribution quantity: percentage of the population used and whether individuals are selected among the best ones or randomly. A main limitation for application to real-world problems is that the above parameters strongly depend on the topology of the objective function and the EA paradigm used [8].

The goal of this work is twofold. Firstly, determining if the parameters involved in the computation of distribution-based quantities influence the behavior of the termination condition. Secondly, exploring if the asymptotic behavior of distribution-based quantities relates to the accuracy of the EA solution. We propose posing the termination problem in statistical inference terms and introduces a general regression-prediction model for determining if two quantities behave equally. We use our model to compare several types of distribution-based quantities reported in the literature [8]. Our experiments conclude that the distribution-based parameters (percentage and selection) do not influence their asymptotic behavior and indicate that the maximum distance to the best individual might be the best choice in terms of computational efficiency and capability of predicting EA accuracy.

This paper is organized as follows. Section 1 describes the existent distribution-based termination conditions for EA algorithms. In section 2 we give the statiscal inference framework. In section 3 we describe the experiments carried out and in section 4 we report the statiscal results. Finally section 5 concludes the paper.

1. Distribution-based Stopping Criteria

A termination condition is given by a quantity reflecting the amount of change between consecutive iterations and the conditon that such quantity should fullfill. Regarding the quantity reflecting the amount of change we have considered distributed-based quantities. Such quantities measure the distribution of a given percentage of the (best) evolving population. Following the literature [8], we have considered:

1. **Maximum Distance (MxD).** It is given by the maximum distance of the population to the best individual.

2. **Population Variability (Std).** It is the maximum standard deviation of the population.

Both quantities can be computed using all individuals or considering only a percentage p of the individuals. These percentage of the individuals can be randomly sampled over the whole population or selected among the best individuals. Given these definitions, for each distribution-based quantity, two different computational procedures can be considered:

1. Percentage p over randomly sampled individuals. It will be indicated by the suffix $Prop_p$.
2. Percentage p over the best individuals. It will be be indicated by the suffix $BestProp_p$. We note that, in this case, the population must be sorted before computing the measure.

We note that the combination of the distribution-based quantities with the two computational procedures described above give four generic alternative quantities: $MaxDProp_p$, $StdProp_p$, $MaxDBestProp_p$, $StdBestProp_p$. From now on, we will note by AltCrit any of these quantities.

Concerning termination conditions, we have the following two:

1. **Absolut Threshold**. The algorithm terminates if the distribution-based quantity is below a given threshold. This condition requires knowing the expected ranges of the distribution-based quantity in order to set the appropriate threshold. Besides it does not guarantee that EA has reached a local minimum.
2. **Stabilization**. The algorithm terminates if the distribution-based quantity is below a given threshold for a number of consecutive generations. This condition is equivalent to have differences between consecutive iterations below a given threshold. On one hand, the latter threshold is independent of the ranges of the distribution-based quantity. On the other hand, this condition ensures that the distribution-based quantity has stabilized and, thus, that EA does not evolve anymore since a local minimum has been attained.

A termination condition is suitable if it compares to the accuracy of the EA solution. The distance to the (known) function minimum is our gold-standard reference convergence criterion, given that is directly associated to the algorithm accuracy. This criterion can only be computed if the optimum of the test function is known and, thus, is useless in real-world problems. We compute it as the maximum distance to the function minimum of a certain percentage p of the individuals [8] and note it by $RefCrit$.

Our final goal is to control (predict) the values taken by $RefCrit$ from the values taken by the alternative measure $AltCrit$. Previous to the latter, we should analyze the influence that the different computational procedures may have on the distribution-based measures. In inference statistics, this can be achieved by relating quantities using a regression model.

2. Inference Model

Given a sampling of two random variables (x and y), the linear regression of y (response variable) over x (explicative variable) is formulated as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{1}$$

for x_i, y_i the sampling of x and y and ε_i an uncorrelated random error following a multivariate normal distribution, $N(0, \Sigma^2)$ of zero mean and variance $\Sigma^2 = \sigma^2 Id$.

The parameters of the regression model (1) are the regression coefficients $\beta = (\beta_0, \beta_1)$ and the error variance σ^2. The regression coefficients describe the way the two variables relate, while the variance indicates the accuracy of the model and, thus, measures to what extent x can predict y.

In the case of exploring equivalence among $AltCrit$, the inference model would be:

$$AltCrit1_i = \beta_0 + \beta_1 AltCrit2_i + \varepsilon_i \tag{2}$$

for $AltCrit1_i$, $AltCrit2_i$ the values of two different $AltCrit$ obtained at the i-th iteration. In the case of relating $AltCrit$ to $RefCrit$, our model would be:

$$RefCrit_i = \beta_0 + \beta_1 AltCrit_i + \varepsilon_i \tag{3}$$

for $RefCrit_i$, $AltCrit_i$ the values of $RefCrit$ and $AltCrit$ obtained at the i-th iteration.

For a sample of length N, the regression coefficients, $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1)$, are computed by Least Squares Estimation (LSE) as:

$$\widehat{\beta} = (X^T X)^{-1} X^T Y \tag{4}$$

for $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$, $Y = (y_1, \ldots, y_N)$ and ^{T} denoting the transpose of a matrix. The differences between estimated responses, $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, and observed responses y_i:

$$e_i = y_i - \widehat{y}_i \tag{5}$$

are called residuals. Their square sum provides an estimation of the error variance:

$$S_R = \widehat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \tag{6}$$

Due to a decrease in the population sparseness at advance stages of EA, the heterocedasticity and Gaussianity assumption are not satisfied. A monotonous increase in σ^2 is usually solved by taking logarithms in both variables [11].

From now on, the values of $RefCrit$ and $AltCrit$ will be assumed to be in logarithmic scale for the inference model:

$$log(y_i) = \beta_0 + \beta_1 log(x_i) + \varepsilon_i \tag{7}$$

We note that, by taking exponentials, the regression model in the original scale is polynomial with multiplicative errors:

$$y_i = e^{\beta_0} x_i^{\beta_1} e^{\varepsilon_i} \tag{8}$$

Previous to any kind of inference, it is mandatory to verify that the estimated parameters make sense. That is, whether it really exists a linear relation between x and y. By the Gauss-Markov theorem, such linear relation can be statistically checked using the following T-test [12]

$$TM: \; H_0 : \beta_1 = 0 \, , H_1 : \beta_1 \neq 0 \tag{9}$$

where a $p - value$ close to zero (below α) ensures the validity of the linear model with a confidence $(1 - \alpha)100\%$.

We note that if the slope $\beta_1 \approx 1$, then both quantities stabilize at the same time and, thus,they are equivalent under the stabilization termination condition. This requirement can be statistically checked using the following unilateral T-tests:

$$\begin{array}{l} TP_1 : H_0 : \beta_1 - 1 \geq \epsilon \quad , \; H_1 : \beta_1 - 1 < \epsilon \\ TP_2 : H_0 : \beta_1 - 1 \leq -\epsilon \; , \; H_1 : \beta_1 - 1 > -\epsilon \end{array} \tag{10}$$

a p value close to zero (below α) for both tests ensures that $|\beta_1 - 1| \leq \epsilon$ with a confidence $(1 - \alpha)100\%$. The minimum ϵ ensuring rejection of TP_1 and TP_2 is given by $|1 - max(|CI(\beta_1)|)|$ for $CI(\beta_1)$ the $(1-\alpha)$-confidence interval for the regression slope. The higher ϵ we have, the least stabilization equivalence.

3. Experimental Settings

In order to compare the different distribution-based quantities, we have considered six well-known test functions [13] having a minimum at zero:

1. Esphere:

$$f_1(x) = \sum_{i=1}^{n} x_i^2$$

2. Rosenbrock:

$$f_2(x) = \sum_{i=1}^{n-1} [100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2]$$

3. Rastrigin:

$$f_3(x) = \sum_{i=1}^{n} [x_i^2 - 10\cos(2\pi x_i + 10)]$$

4. Griewangk:

$$f_4(x) = \sum_{i=1}^{n} x_i^2 - \prod_{i=1}^{n} \cos(\frac{x_i}{\sqrt{i}}) + 1$$

5. Ackley:

$$f_5(x) = 20 + e - 20e^{-0.2\sqrt{\frac{1}{2}\sum_{i=1}^{n} x_i^2}} - e^{\frac{1}{2}\sum_{i=1}^{n} cos(2\pi x_i)}$$

6. Easom:

$$f_7(x) = -\cos x_1 . \cos x_2 . \\ \exp(-((x_1 - \pi)^2 + (x_1 - \pi)^2))$$

We have used a Differential Evolution (DE) technique for the minimization task. Differential evolution is a real parameter encoding evolutionary algorithm for global optimization over continuous spaces [14,15]. In this paper, we use the 3-parameter DE1 scheme [14] for solving DE. For a real search space of dimension D, the population is randomly initialized with ND vectors (for ND the first algorithm parameter). Each vector v in the population is evolved by mutation and recombination operators. Given a mutation rate $F \in [0, 2]$ (second parameter of the algorithm), the mutation operator produces a new vector vm by adding a vector difference of two randomly chosen population vectors $v1$ and $v2$ to another randomly chosen vector $v3$:

$$vm = v1 + F(v2 - v3) \tag{11}$$

For the recombination step, a new vector vf is created from the mutation vector by means of a combination rate CR (third parameter of the algorithm) as follows:

$$vf_i = \begin{cases} vm_i & \text{if} \quad r_i < CR \text{ or } i = k \\ v_i & \text{otherwise} \end{cases} \tag{12}$$

for vf_i the i-th component of vf and $r_i \in [0, 1]$ a random number and k a random number uniformly distributed in $[1, D]$. Finally a selection operator is applied. The vector vf and the initial vector v are compared and the vector that better fits the objective function is selected and remains in the next population. This process is iteratively repeated until a stopping criterion is reached. Following the literature [15], we have chosen the following values for DE parameters: D=2, ND=20, F=0.9, CR=0.5. For each test function, we have executed 100 trails of the algorithm during 10.000 iterations each one.

Two different experiments have been carried out:

1. **Computation of Distribution-based Quantities.** A first compulsory step before exploring if any $AltCrit$ can substitute $RefCrit$ is checking if the computation of distribution-based quantities using different parameter settings yields equivalent quantities. In the case such equivalency held, it would imply that **MxD** and **Std** can be computed directly from EA current population. Otherwise, EA output should be adapted in order to conform to the most adequate proportion and selection strategy.
2. **Capability for Substituting the Reference Criterion.** Under the assumption that any parameters can be used for computing $AltCrit$, we can compare it to $RefCrit$ using the natural setting for DE: a percentage ($p = 30\%$) of the whole population.

For both experiments, we have assessed the validity of the linear model in logarithmic scale (given by S_R and TM test), as well as, its stabilization equivalence (given by the ϵ rejecting TP_i, $i = 1, 2$).

4. Experiments and Results

4.1. Computation of Distribution-based Quantities

Tables 1 and 2 report the estimation of the model parameters (the regression coefficients $\widehat{\beta}_0$, $\widehat{\beta}_1$ and the residual variance S_R), the p-value of the model verification T-test and the stabilization equivalence tolerance ϵ. Table 1 shows results for the comparison between best and randomly selected individuals for the proportion $p = 60\%$ and the two distribution-based quantities. Table 2 shows results for the comparison across proportions for **MxD**.

For all cases, there is a clear linear relation between the different quantities (with p close to the working precision). We can assume that $\widehat{\beta}_1$ is close to 1 with a tolerance $\epsilon \approx 10^{-2}$ for dependency on the best individuals (table 1) and $\epsilon \approx 10^{-1}$ for dependency on proportions (table 2). Therefore, all computational strategies are equivalent for the stabilization termination condition. Besides, the sign of the constant term $\widehat{\beta}_0$ gives the following (expected) inequalities. For any percentage p, we have:

$$\mathbf{MxDBestProp}_p < \mathbf{MxDProp}_p \text{ and } \mathbf{StdBestProp}_p < \mathbf{StdProp}_\mathbf{p}$$

and for increasing proportions:

$$\mathbf{MxDProp_{30}} < \mathbf{MxDProp_{60}} \sim \mathbf{MxDProp_{100}}$$

The above inequalities indicate that best individuals are a more compact cluster and that the estimation of the population dispersion increases with the number of samples considered.

	MxD					**Std**				
	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ
Esphere	$\leq 10^{-30}$	0.99	0.75	0.11	0.00002	$\leq 10^{-30}$	1.00	0.62	0.07	0.00005
Rosenbrock	$\leq 10^{-30}$	0.99	0.51	0.14	0.00274	$\leq 10^{-30}$	0.99	0.43	0.09	0.00259
Rastrigin	$\leq 10^{-30}$	1.03	0.80	0.10	0.04029	$\leq 10^{-30}$	1.03	0.74	0.08	0.03576
Griewangk	$\leq 10^{-30}$	1.03	0.62	0.33	0.03243	$\leq 10^{-30}$	1.02	0.59	0.27	0.02988
Ackley	$\leq 10^{-30}$	1.02	0.99	0.10	0.02530	$\leq 10^{-30}$	1.02	0.84	0.07	0.02094
Easom	$\leq 10^{-30}$	1.05	1.06	0.05	0.05647	$\leq 10^{-30}$	1.04	0.90	0.03	0.04537

Table 1. Comparison between best or random selection for a percentage $p = 60\%$ of the population.

4.2. Capability for Substituting the Reference Criterion

Given that the setting used for the computation of alternative quantities is irrelevant for termination by stabilization, we have computed **MxD** and **Std** for a uniform sampling of $p = 30\%$ of the population. Table 3 reports the estimation of the model parameters (the regression coefficients $\widehat{\beta}_0$, $\widehat{\beta}_1$ and the residual variance S_R), the p-value of the model verification T-test and the stabilization equivalence tolerance ϵ. We report values for each test function (rows) and alternative quantity (columns). For all cases, there is a clear lin-

	Proportion p=30% versus p=60%					Proportion p=30% versus p=100%				
	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ
Esphere	$\leq 10^{-30}$	0.99	-0.5	6.32	0.00081	$\leq 10^{-30}$	1.00	-0.6	5.23	0.00344
Rosenbrock	$\leq 10^{-30}$	0.97	-0.4	4.85	0.01825	$\leq 10^{-30}$	0.96	-0.4	5.36	0.03187
Rastrigin	$\leq 10^{-30}$	0.99	-0.2	0.48	0.00804	$\leq 10^{-30}$	0.98	-0.6	0.42	0.02287
Griewangk	$\leq 10^{-30}$	0.95	-1.0	3.36	0.04759	$\leq 10^{-30}$	0.80	-3.4	11.75	0.19316
Ackley	$\leq 10^{-30}$	0.99	-0.1	0.63	0.00212	$\leq 10^{-30}$	0.99	-0.2	0.53	0.00113
Easom	$\leq 10^{-30}$	0.99	-0.3	0.40	0.01058	$\leq 10^{-30}$	0.99	-0.3	0.48	0.00611

Table 2. Comparison across percentages for individuals randomly chosen in the computation of **MxD**.

	$\mathbf{MxDProp}_{30}$					$\mathbf{StdProp}_{30}$				
	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ	TM	$\widehat{\beta}_1$	$\widehat{\beta}_0$	S_R	ϵ
Esfere	$\leq 10^{-32}$	1.000	-0.183	0.05	0.000	$\leq 10^{-32}$	1.000	0.685	0.03	0.000
Rosenbrock	$\leq 10^{-32}$	1.002	-0.122	0.09	0.002	$\leq 10^{-32}$	1.003	0.761	0.05	0.003
Rastrigin	$\leq 10^{-32}$	1.006	-0.207	0.06	0.006	$\leq 10^{-32}$	1.007	0.686	0.03	0.006
Griewangk	$\leq 10^{-32}$	1.009	-0.1714	0.02	0.007	$\leq 10^{-32}$	1.008	0.7021	0.01	0.007
Ackley	$\leq 10^{-32}$	1.004	-0.148	0.06	0.004	$\leq 10^{-32}$	1.004	0.725	0.03	0.004
Easom	$\leq 10^{-32}$	1.011	-0.143	0.04	0.011	$\leq 10^{-32}$	1.011	0.766	0.03	0.011

Table 3. Comparison to $RefCrit$.

ear relation between accuracy and the alternative quantities (with p close to the working precision). Besides the goodness-of-fit is excellent, given that S_R is extremely small compared to the variable ranges.

Concerning the relation between the two variables, it is worth noticing two aspects. Firstly, we observe that the estimated slope $\widehat{\beta}_1$ is close to 1 for all cases with tolerance $\epsilon \leq 0.11$ for both quantities **MxD** and **Std**. This implies that the relation in logarithmic scale is a translation of the identity and the regression model in the original scale is also linear. Secondly, the constant coefficients $\widehat{\beta}_0$ are sorted as follows:

$$\widehat{\beta}_0(\mathbf{MxDProp}) \leq 0 \leq \widehat{\beta}_0(\mathbf{StdProp})$$

The above commented points indicate that there might be the following tendency:

$$\mathbf{StdProp} \leq RefCrit \leq \mathbf{MxDProp}$$

This already suggests that the value of maximum distances itself might guarantee an upper bound for the EA accuracy.

5. Conclusions and Future work

A main challenge in Evolutionary Algorithms (EAs) is determining a termination condition ensuring stabilization close to the optimum in real-world applications. Although distribution-based conditions are the best suited, a major concern is setting appropiate parameters for their computation. In this context, this paper addresses EA termination

condition in terms of statistical inference and contributes in two issues. First, it explores if the parameters involved in the computation of distribution-based quantities influence the behavior of the termination condition. Second, it reports a preliminary study on the relation between such termination conditions and the accuracy of the EA solution.

According to our experiments on several known test functions, the following conclusions can be derived. On one hand, there is a high correlation among the different computational procedures for distribution-based quantities and they behave equally with respect their stabilization. Therefore, we conclude that in order to use a quantity behaving as the solution accuracy we can choose, among the various computational procedures, the most convenient for the particular EA paradigm we are running. On the other hand, there is a strong linear relation between distribution-based quantities and the distance to the optimum. From our analysis, we conclude that quantities based on maximum distances have the highest concordance to EA accuracy, in the sense that they guarantee an upper bound for the accuracy. Thus, they are the best candidates for terminating EA in real-world problems.

We consider that there are some issues that should be further developed. The test functions used are a small set of benchmarking data sets (we cover two out of the five categories described in [16]) and only 2-D problems have been solved. Enlarging the test function data set including groups of functions with specific key features [16] is work currently under development. However, the functions used include three properties (multimodality, global structure and scalability) reported in a recent study [17] to have a high influence in the performance of EA's. Regarding size, although it definitely influences convergence rate (more iterations of EA are required [16]), this is independent of the relationship between *RefCrit* and *AltCrit*. Thus, size is not a limitation for the prediction model, which links convergence rate with population stability.

In this study we have restricted to DE algorithm. We are currently enlarging EA methods in order to cover existing EA paradigms: genetic algorithms [18], evolutionary strategies [19], particle Swarm optimization [20], among others. Nevertheless, we do not expect any significant changes in our conclusions since DE already presents the main features of EA [21].

Finally, in our experimental setting test functions have been studied separately. We consider that the influence of the test function should be taken into account, so that the inference can be done independently of the function features. This will be studied by using generalized regression models including random effects [22] modelling the impact of the test function group.

Acknowledgements. This work was supported by the Spanish projects PI071188, TIN2009-13618 and CONSOLIDER-INGENIO 2010 (CSD2007-00018). The 2nd author has been supported by The Ramon y Cajal Program.

We would like to thank Llorenç Badiella from the Statistics Service at the Universitat Autonoma de Barcelona for his advise.

References

[1] Bäck, T., Hammel, U., Schwefel, H.: Evolutionary computation: Comments on the history and current state. IEEE Transactions on Evolutionary Computation **1** (1997) 3–17

[2] Ravikumar, V., Panigrahi, B.: Comparative study of evolutionary computing methods for parameter estimation of quality signals. International Journal of Applied Evolutionary Computation **1 (2)** (2010) 28–59
[3] Rizki, M., Zmuda, M., Tamburino, L.: Evolving pattern recognition systems. IEEE Transactions on Evolutionary Computation **6 (6)** (2002) 594–609
[4] Cagnoni, S., Lutton, E., Olague, G.: Genetic and Evolutionary Computation for Image Processing and Analysis. Hindawi Publishing Corporation (2008)
[5] Safe, M., Carballido, J., Ponzoni, I.: On stopping criteria for genetic algorithms. In: 17th BSAI, LNAI 3117. (2004) 405–413
[6] Price, K., Storm, R., Lampinen, J.: Differential Evolution: A Practical Approach to Global Optimization. Springer (2005)
[7] Tagetiren, M., Suganthan, P.: A multi-populated differential evolution algorithm for solving contrained optimization problem. In: IEEE Congres on Evolutionary Computation. (2006) 340–347
[8] Zielinski, K., Laur, R.: Stopping criteria for differential evolution in constrained single-objective optimization. In: Advances in Differential Evololution, vol. 143. (2010) 111–138
[9] Zielinski, K., Laur, R.: Stopping criteria for a constrained single objective particle swarm optimization algorithm. Informatica 31(1) (2007) 51–59
[10] K. Zielinski, D.P., Laur, R.: Stopping criteria for single-objective optimization. In: Third International Conference on Computational Intelligence, Robotics and Autonomous Systems, Singapore. (2005)
[11] Arnold, S.: The theory of linear models and multivariate observations. Wiley (1997)
[12] Newbold, P., Carlson, W., Thorne, B.: Statistics for Business and Economics. 6th edn. Pearson Education (2007)
[13] Digalakis, J., Margaritis, K.: An experimental study of benchmarking functions for genetic algorithms. International Journal of Computer Mathematics **79(4)** (2002) 403–416
[14] Storn, R., Price, K.: Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization **11** (1997) 341–359
[15] Das, S., Konar, A.: An improved differential evolution scheme for noisy optimization problems. In: LNCS vol. 3776. (2005) 417–421
[16] Hansen, N., Auger, A., Ros, R., Finck, S., Posík, P.: Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009. In: GECCO 2010. (2010)
[17] Mersmann, O., Preuss, M., Trautmann, H.: Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. In: PPSN XI, Part I, LNCS 6238. (2010) 73–82
[18] Goldberg, D., Richardson, J.: Genetic algorithms with sharing for multimodal function optimization. In: Second International Conference on Genetic Algorithms and their Application. (1987) 41–49
[19] Beyer, H.G., Schwefel, H.P.: Evolution strategies: A comprehensive introduction. Journal Natural Computing **1(1)** (2002) 3–52
[20] Barrera, J., Coello, C.A.C. In: A Review of Particle Swarm Optimization Methods used for Multimodal Optimization. Springer-Verlag (2009) 9–37
[21] Ronkkonen, J.: Continuous Multimodal Global Optimization with Diferential Evolution Based Methods. PhD thesis, Lappeenranta University of Technology (2009)
[22] Lee, Y., Nelder, J., Pawitan, Y.: Generalized linear models with random effects. Chapman and Hall/CRC Boca Raton (2006)