Gaussian EDA and Truncation Selection: Setting Limits for Sustainable Progress

Petr Pošík

Czech Technical University, Faculty of Electrical Engineering, Department of Cybernetics
Technická 2, 166 27 Prague 6, Czech Republic, Email: posik@labe.felk.cvut.cz

***Abstract*—In real-valued estimation-of-distribution algorithms, the Gaussian distribution is often used along with maximum likelihood (ML) estimation of its parameters. Such a process is highly prone to premature convergence. The simplest method for preventing premature convergence of gaussian distribution is to enlarge the maximum likelihood estimate of standard deviation σ by a constant factor k each generation. This paper surveys and broadens the theoretical models of the behaviour of this simple EDA on 1D problems and derives the limits for the constant k. The behaviour of this simple EDA with various values of k is analysed and the agreement of the model with the reality is confirmed.**

***Index Terms*—estimation of distribution algorithms, Gaussian distribution, truncated Gaussian distribution, truncation selection, premature convergence, variance enlargement**

I. Introduction

Evolutionary algorithms (EAs) are very popular optimization and search methods that are able to solve a broad range of optimization tasks, i.e. to find the optimum of a broad range of objective (fitness) functions. Estimation of distribution algorithms (EDAs) [1] are a class of EAs that (in contrast to EAs) do not use the crossover and mutation operators to create the offspring population. Instead, they build a probabilistic model describing the distribution of selected (i.e. promissing) individuals and create offspring by sampling from the model.

In EDAs working in the real domain (real-valued EDAs), the Gaussian distribution is often employed as the model of promissing individuals ([2], [3], [4], [5]). The distribution is often learned by maximum likelihood estimation (MLE). It was recognized by many authors that such a learning scheme makes the algorithm very prone to premature convergence (see e.g. [3], [6], [7], [8]).

There are many techniques that fight the premature convergence, usually by means of artificially enlarging the maximum likelihood estimate of variance of the learned distribution. In [6], the variance is kept on values greater than 1, while [7] uses self-adaptation of the variance. It is also true, that different approach should be used when the population is on the slope or in the valley of the fitness function: when the population is centered around optimum, there is actually no need to enlarge the variance, while enlarging the variance is essential on the slope. This approach is followed e.g. in [9] and [10] where the so called *adaptive variance scaling* is used when the population is on slope, while ordinary ML estimation is used around optimum.

This article aims at the simplest way for preventing premature convergence in the context of the simple Gaussian EDA displayed in Fig. 1. The variance is scaled by a constant factor k that does not change during evolution. Such an approach was used e.g. in [11] where also the minimal values of the 'amplification coefficient' were determined by search. In this paper, explicit mathematical values are developed not only for minimal, but also for the maximal values of k.

1) Set the initial values of parameters μ and σ.
2) Sample N new individuals from the distribution $\mathcal{N}(\mu, \sigma)$.
3) Evaluate them.
4) Select the τN best solutions.
5) Estimate new values of parameters μ and σ based on the selected individuals.
6) Enlarge the σ by a constant factor k.
7) If termination condition is not met, go to step 2.

Fig. 1. Simple Gaussian EDA analysed in this article.

The rest of the paper is organized as follows: in Sec. II, the ML estimates of parameters of truncated normal distribution are presented. These estimates are then used in Sec. III to develop models of the parameters of the Gaussian distribution—mean μ and standard deviation σ—on the slope and in the valley of the fitness function. Based on them the limits for the constant k are developed. Sec. IV then presents the behaviour of the model for various values of k and Sec. V evaluates the agreement between the model and the reality. Finally, Sec. VI summarizes and concludes the paper, and presents some pointers to future work.

II. Truncated Normal Distribution

Suppose we have a random variable X with the Gaussian distribution with the mean μ and standard deviation σ, $X \sim \mathcal{N}(\mu, \sigma)$. The probability distribution function (PDF) of the normal distribution is

$$\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \tag{1}$$

and the cumulative distribution function (CDF) is

$$\Phi_{\mu,\sigma}(x) = \int_{-\infty}^{x} \phi_{\mu,\sigma}(t)dt. \tag{2}$$

Note, that for the PDF and CDF of the standard normal distribution the notation ϕ and Φ (i.e. without subscripts) will be used throughout the paper.

Suppose further, that we truncate the distribution, i.e. we allow the values of X to come only from interval $\langle x_1, x_2\rangle$. The expected value of such a distribution can be written as

$$E(X|X \in \langle x_1, x_2\rangle) = \mu + \sigma \cdot d(z_1, z_2), \tag{3}$$

where $z_i = \frac{x_i - \mu}{\sigma}$ and

$$d(z_1, z_2) = -\frac{\phi(z_2) - \phi(z_1)}{\Phi(z_2) - \Phi(z_1)}. \tag{4}$$

The variance of the truncated distribution can be written as

$$\mathrm{Var}(X|X \in \langle x_1, x_2\rangle) = \sigma^2 \cdot c(z_1, z_2) \tag{5}$$

where

$$c(z_1, z_2) = 1 - \frac{z_2 \cdot \phi(z_2) - z_1 \cdot \phi(z_1)}{\Phi(z_2) - \Phi(z_1)} - d(z_1, z_2)^2 \tag{6}$$

The derivation of the above equations can be found in literature and is included in the appendix for completeness.

III. Truncation Selection

Truncation selection is one of the standard selection schemes used in EAs. It has one parameter, so-called *selection proportion* τ, $\tau \in (0, 1)$, which states how large portion of the population is selected. The selection scheme then chooses $\tau \cdot N$ best individuals from the old population to become parents (in case of EAs), or to serve as the sample to which a probabilistic model is fitted (in case of EDAs).

The coefficients c and d (Eqs. 6 and 4) have a direct meaning to the dynamics of the distribution parameters: d is the distance between two consecutive estimates of μ measured in the standard deviations, while c is the ratio of the two consecutive estimates of σ^2.

The effect of the truncation selection on the population distribution is different depending on the fitness function values. There are basically two bounding cases:

- The population is on the slope of the fitness function.
- The population is in the valley of the fitness function.

A. On the Slope of Fitness Function

First, suppose the population is on the slope of the fitness function. The best $\tau \cdot N$ individuals are at the right- (or on the left-) hand side of the distribution, i.e.

$$z_1 = \Phi^{-1}(1 - \tau), \text{ and} \tag{7}$$

$$z_2 \to \infty, \tag{8}$$

where Φ^{-1} is the inverse cumulative distribution function of the Gaussian distribution. We can thus define

$$d_{\mathrm{slope}}(\tau) = d\left(\Phi^{-1}(1 - \tau), \infty\right), \text{ and} \tag{9}$$

$$c_{\mathrm{slope}}(\tau) = c\left(\Phi^{-1}(1 - \tau), \infty\right). \tag{10}$$

These equations were already presented in previous works (e.g. in [8]), although in a bit different form.

B. In the Valley of the Fitness Function

The other extreme is the case, when the population is centered around the optimum. In that case

$$z_1 = \Phi^{-1}\left(\frac{1 - \tau}{2}\right), \text{ and} \tag{11}$$

$$z_2 = \Phi^{-1}\left(\frac{1 + \tau}{2}\right). \tag{12}$$

We can again define

$$d_{\mathrm{valley}}(\tau) = d\left(\Phi^{-1}\left(\frac{1 - \tau}{2}\right), \Phi^{-1}\left(\frac{1 + \tau}{2}\right)\right), \text{ and} \tag{13}$$

$$c_{\mathrm{valley}}(\tau) = c\left(\Phi^{-1}\left(\frac{1 - \tau}{2}\right), \Phi^{-1}\left(\frac{1 + \tau}{2}\right)\right). \tag{14}$$

Also these equations were presented in previous work ([12]). In this article, however, all the equations are derived using the framework of general truncated normal distribution.

C. Comparison of the Dynamics

We can now compare the functions $c(\tau)$ and $d(\tau)$ on the slope and in the valley graphically. Figure 2 shows the dependence of the value d on τ for the two cases, i.e. if the population is on the slope or in the valley. We can see that when the EDA population already resides on the optimum, it does not make any progress in the sense of shifting the population mean (coefficient $d = 0$ for all values of τ), it does not need to. On the other hand, when on slope, the largest step can be made when only the one best individual is selected ($d(\tau)$ is maximal when τ is minimal, i.e. $\tau = \frac{1}{N}$). When all individuals are selected ($\tau = 1$), no progress is made ($d(1) = 0$). However, when using very low τ, the variance would shrink very quickly ($\lim_{\tau \to 0} c(\tau) = 0$), as can be seen in Fig. 3. On the slope the variance of the τN best individuals is greater than the variance of the τN best individuals in the valley.

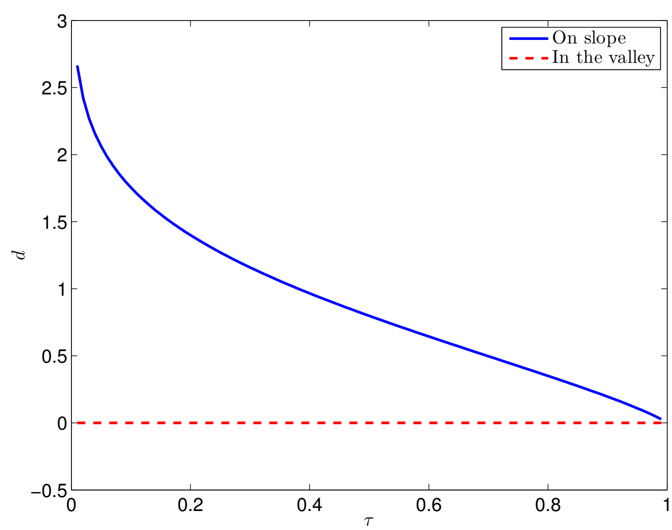

Fig. 2. Coefficient d: how large step can EDA make in one iteration for various values of τ on slope and in the valley.

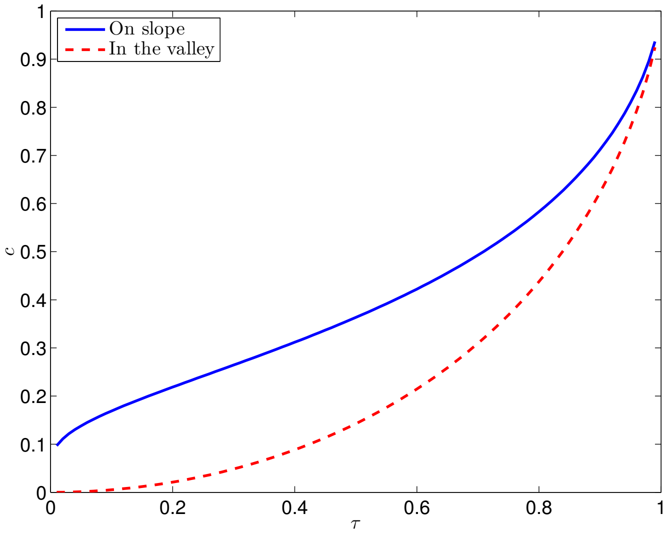

Fig. 3. Coefficient c: how much the variance shrinks in one iteration of EDA for various values of τ on slope and in the valley.

D. Preventing Premature Convergence

As already said in the introduction, when ML estimates of mean and standard deviation are used, the premature convergence can occur even on the slope of the fitness function since (see [8])

$$\lim_{t\to\infty} \mu^t = \mu^0 + \sigma^0 \cdot d(\tau) \cdot \frac{1}{1-\sqrt{c(\tau)}}, \text{ and} \tag{15}$$

$$\lim_{t\to\infty} \sigma^t = 0, \tag{16}$$

i.e. the distance the population can travel is bounded and is given only by the initial values of the parameters, μ^0 and σ^0, and by the selection proportion τ.

The simplest method of preventing premature convergence is to enlarge the estimated standard deviation σ by a constant factor k. Thus

$$\sigma_{t+1} = k \cdot \sigma_t \cdot \sqrt{c} \tag{17}$$

In order to prevent the premature convergence on the slope, the ratio of the consecutive standard deviations should be at least 1, i.e.

$$\frac{\sigma_{t+1}}{\sigma_t} = k \cdot \sqrt{c_{\text{slope}}} \geq 1 \tag{18}$$

$$k \geq \frac{1}{\sqrt{c_{\text{slope}}}} \tag{19}$$

On the other hand, to be able to exploit the optimum, the model must be allowed to converge in the valley. The ratio of the two consecutive standard deviations should be lower than 1, i.e.

$$\frac{\sigma_{t+1}}{\sigma_t} = k \cdot \sqrt{c_{\text{valley}}} < 1 \tag{20}$$

$$k < \frac{1}{\sqrt{c_{\text{valley}}}} \tag{21}$$

Joining these two conditions together gives us the limits for the constant k:

$$\frac{1}{\sqrt{c_{\text{slope}}}} \leq k < \frac{1}{\sqrt{c_{\text{valley}}}} \tag{22}$$

The limits are depicted on Fig. 4, the same values are shown in tabular form in Table I.

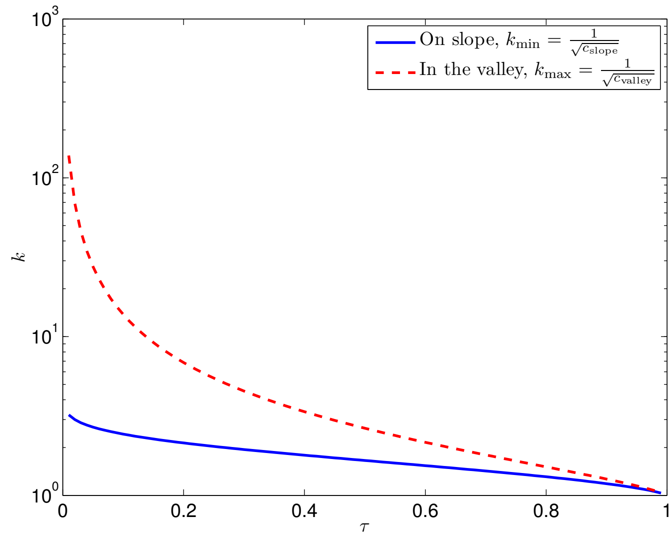

Fig. 4. Minimum and maximum values for setting the enlarging parameter k for various values of τ.

TABLE I
THE LIMITS FOR THE STANDARD DEVIATION ENLARGMENT CONSTANT k FOR VARIOUS VALUES OF THE SELECTION PROPORTION τ

τ	0.01	0.10	0.30	0.50	0.70	0.90	0.99
$k_{\min}$	3.213	2.432	1.944	1.659	1.424	1.185	1.033
$k_{\max}$	138.195	13.798	4.540	2.648	1.797	1.267	1.040

E. EDA on Symmetric Unimodal Function

Suppose we would like to model the whole evolution of the simple Gaussian EDA on symetric unimodal function. In order to use the apparatus for the truncated normal distribution, we would need to express the selection limits x_1 and x_2 in terms of the current μ and σ and in terms of known position of the optimum x_0.

Now we have models of our simple EDA behavior in two special cases: (1) when μ is very far from x_0, we can model this situation as the slope and set one of the interval limits to ∞; (2) when $\mu = x_0$, we know that the interval should include $\frac{\tau \cdot N}{2}$ best individuals below and over the optimum x_0. However, we still need a model describing the behavior in the general case in which the following equation must hold:

$$\tau = \int_{x_0-l}^{x_0+l} \phi_{\mu,\sigma}(x)dx, \tag{23}$$

where the τN best individuals are contained in the interval $(x_0 - l, x_0 + l)$. We need to get the value of l to be able to compute the μ and σ of the truncated normal distribution, i.e. the parameters of the sampling distribution for the next generation. In [11], a simple line search algorithm is suggested to find the value of l. In this paper, we shall adopt that approach but we suggest how to make the life of the line search easier by setting a reasonable bounds on the value of l.

First, suppose that $\mu > x_0$. Based on the relations among the distributions $\mathcal{N}(x_0, \sigma)$ and $\mathcal{N}(\mu, \sigma)$, the following inequalities can be stated:

$$\begin{aligned} \Phi^{-1}_{x_0,\sigma}\left(\frac{1+\tau}{2}\right) <& x_0 + l < \Phi^{-1}_{\mu,\sigma}\left(\frac{1+\tau}{2}\right) \\ & x_0 - l < \Phi^{-1}_{x_0,\sigma}\left(\frac{1-\tau}{2}\right) \\ \Phi^{-1}_{\mu,\sigma}(\tau) <& x_0 + l \end{aligned}$$

From these inequalities the following limits on the value l can be derived:

$$\left.\begin{matrix} \Phi^{-1}_{x_0,\sigma}\left(\frac{1+\tau}{2}\right) - x_0 \\ x_0 - \Phi^{-1}_{x_0,\sigma}\left(\frac{1-\tau}{2}\right) \\ \Phi^{-1}_{\mu,\sigma}(\tau) - x_0 \end{matrix}\right\} < l < \Phi^{-1}_{\mu,\sigma}\left(\frac{1+\tau}{2}\right) - x_0 \qquad (24)$$

Similarly, when $\mu < x_0$, the following inequalities can be stated:

$$\begin{aligned} \Phi^{-1}_{\mu,\sigma}\left(\frac{1-\tau}{2}\right) <& x_0 - l < \Phi^{-1}_{x_0,\sigma}\left(\frac{1-\tau}{2}\right) \\ \Phi^{-1}_{x_0,\sigma}\left(\frac{1+\tau}{2}\right) <& x_0 + l \\ & x_0 - l < \Phi^{-1}_{\mu,\sigma}(1-\tau) \end{aligned}$$

Again, the following limits on the value l can be derived:

$$\left.\begin{matrix} x_0 - \Phi^{-1}_{x_0,\sigma}\left(\frac{1-\tau}{2}\right) \\ \Phi^{-1}_{x_0,\sigma}\left(\frac{1+\tau}{2}\right) - x_0 \\ x_0 - \Phi^{-1}_{\mu,\sigma}(1-\tau) \end{matrix}\right\} < l < x_0 - \Phi^{-1}_{\mu,\sigma}\left(\frac{1-\tau}{2}\right) \qquad (25)$$

Now, we can search for the value of l using simple line search heuristic (e.g. the `fminbnd` function from MATLAB) using the limits derived above.[1]

IV. Simulated Evolution

Let us look at the evolution traces for various values of the selection proportion τ and various values of the variance enlargement constant k. The fitness function is symmetric around zero ($f = x^2$ was used), and the settings of the initial distribution are $\mu_0 = 100$, $\sigma_0 = 1$. This way, the algorithm starts on the slope, but then it should change its behaviour in the valley.

Figure 5 shows the evolution of the values of μ and σ in 100 generations for various values of selection proportion τ and enlarging constant k. For all values of k and τ, a similar pattern can be observed. In the first phase of evolution, parameter μ first approaches the optimum x_0 slowly (the population is on slope), but with increasing speed (when $k > k_{\min}$) since the standard deviation σ gets larger. Then the population reaches the neighborhood of the optimum x_0 and the second phase starts: μ quickly approaches the optimum x_0, and when $k < k_{\max}$ is used, standard deviation σ starts to get smaller allowing the algorithm to focus the search in the neighborhood of the optimum x_0.

If the setting $k < k_{\min}$ is used, the premature convergence on the slope occurs, and the algorithm is not able to locate the position of optimum x_0 (unless it is initialized with values from its neighborhood). On the other hand, when using $k > k_{\max}$, the algorithm locates the neighborhood of the optimum quickly, but the sequence of the standard deviations diverges, i.e. gets larger. What cannot be seen in Fig. 5 is the fact that the enlarging standard deviation prevents the algorithm from finding more precise solutions when finite population is used.

In [11], the value of $k_{\min}$ was called 'the optimal amplifying parameter', however, as can be seen in Fig. 5, in cases when the population is located far away from the optimum, we can get better results by choosing $k > k_{\min}$. The neighborhood of the optimum can be found faster in that case.

As the model suggests, with infinite population size and appropriate enlarging constant k it is profittable to use small values of selection proportion τ—the convergence is quick since the EDA can make large steps on the slope.

V. Experimental Evaluation

To evaluate the agreement of the theoretical model with the real-world behaviour of the algorithm, several experiments were conducted. The population size was 10,000 and 25 independent runs were carried out. The average values of μ and σ during evolution are displayed in Fig. 6. As can be seen, the model predicts the behavior accurately.

VI. Summary and Conclusions

This paper summarized and enriched the field of mathematical modelling of a simple EDA based on Gaussian distribution and truncation selection assuming infinite population size. Specifically, the paper aimed at the simplest way of preventing premature convergence arising when maximum likelihood estimation is used, i.e. on enlarging the ML estimate of standard deviation σ by a constant factor k. Based on the models of the EDA behaviour on 1D monotonous functions and 1D unimodal symmetric functions, the limits for the value of k were developed. The limits correspond to the requirements that the variance of the distribution should be

- sufficiently large to allow the algorithm to traverse the slope of the fitness function with at least nondecreasing velocity, and
- sufficiently low to allow the algorithm to exploit the neighborhood of the optimum once it is found.

The optimal value of k depends on the initial distance of the distribution from the optimum x_0 (if the distance is large, large value of k is profitable to traverse the slope quickly) and on the desired accuracy level (if high accuracy is required, small values of k are profitable to quickly focus to the optimum).

[1] We can evaluate all the lower limits and use the maximum of them as the actual lower limit.

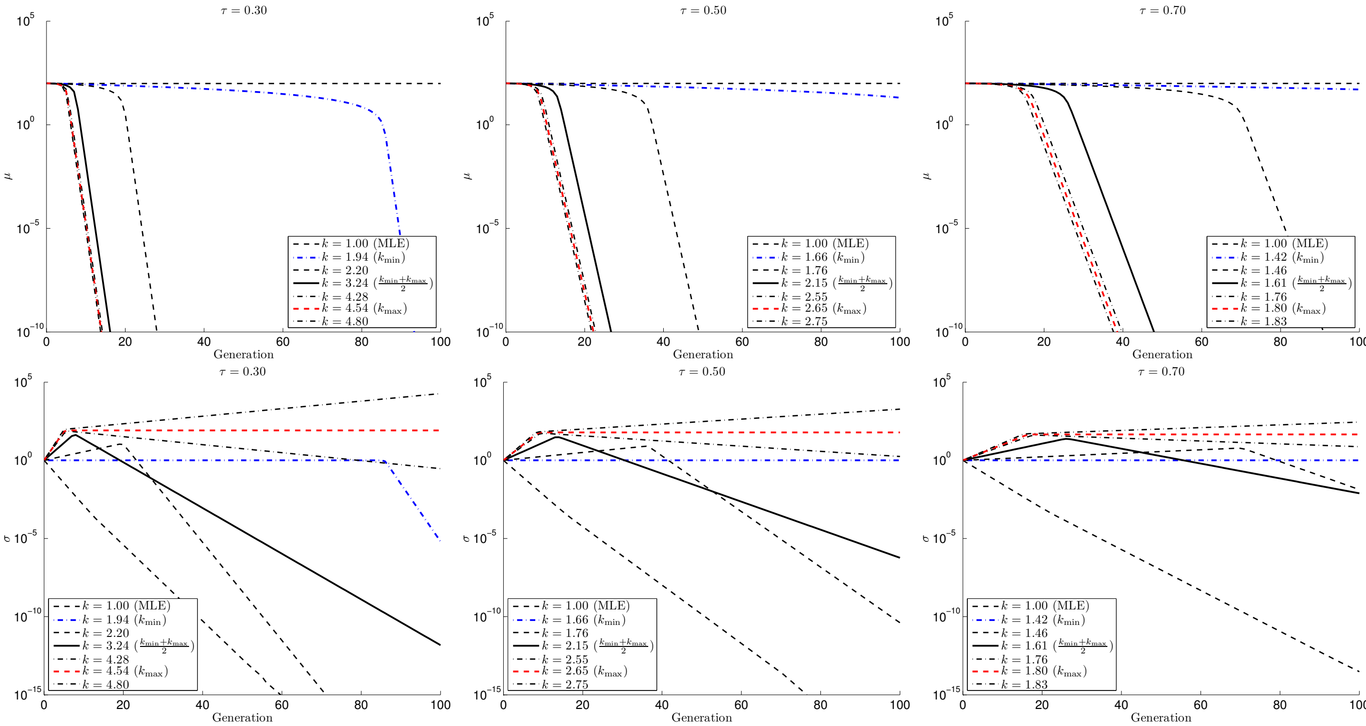

Fig. 5. Evolution of the mean μ (top row) and standard deviation σ (bottom row) for various values of the enlargement constant k with selection proportions $\tau = 0.3$ (left column), $\tau = 0.5$ (middle column), and $\tau = 0.7$ (right column).

As a rule of thumb, one can use the average of the lower and upper limit for k for the particular value of τ.

It should be noted, however, that the limits for k are based on a model with the assumption of infinite population size. The effect of using finite and small populations must be studied and understood yet. Similarly, this model was derived only for the 1D fitness function. Although we can use it for high-dimensional search spaces as a rule of thumb, the real behaviour of Gaussian EDA will be different. Again, this remains as a topic for future research.

Nevertheless, having the bounds for the coefficient k is very profitable from the experimenter's point of view. These bounds could be also used in various adaptive variance scaling schemes as safeguards.

Acknowledgment

The author is supported by the Grant Agency of the Czech Republic with the grant no. 102/08/P094 entitled "Machine learning methods for solution construction in evolutionary algorithms".

Appendix

A. Expected Value of the Truncated Normal Distribution

The probability density function of the standard normal distribution is

$$\phi(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}. \tag{26}$$

Note, that

$$\phi'(z) = -z \cdot \phi(z) \tag{27}$$

The probability density function of the standard normal distribution truncated at z_1 from below and at z_2 from top is

$$\phi_T(z) = \begin{cases} \frac{\phi(z)}{\Phi(z_2)-\Phi(z_1)}, & \text{iff } z \in \langle z_1, z_2 \rangle \\ 0 & \text{otherwise} \end{cases} \tag{28}$$

Expected value of the truncated standard normal distribution is:

$$\begin{aligned}
&E(Z|Z \in \langle z_1, z_2 \rangle) = \\
&= \int_{-\infty}^{\infty} z \cdot \phi_T(z) dz = \\
&= \frac{1}{\Phi(z_2) - \Phi(z_1)} \int_{z_1}^{z_2} z \cdot \phi(z) dz \left| \begin{array}{ll} t & = \phi(z) \\ dt & = -z \cdot \phi(z) dz \end{array} \right. = \\
&= -\frac{1}{\Phi(z_2) - \Phi(z_1)} \int_{\phi(z_1)}^{\phi(z_2)} dt = -\frac{\phi(z_2) - \phi(z_1)}{\Phi(z_2) - \Phi(z_1)}
\end{aligned} \tag{29}$$

For the random variable X which is created by linear transform of variable Z, $X = \mu + \sigma Z$ we can write

$$E(X|X \in \langle x_1, x_2 \rangle) = \mu + \sigma E(Z|Z \in \langle z_1, z_2 \rangle), \tag{30}$$

where $z_i = \frac{x_i - \mu}{\sigma}$.

B. Variance of the Truncated Normal Distribution

For the variance of the truncated normal distribution we can write:

$$\begin{aligned}
&\mathrm{Var}(Z|Z \in \langle z_1, z_2 \rangle) = \\
&= E(Z^2|Z \in \langle z_1, z_2 \rangle) - E^2(Z|Z \in \langle z_1, z_2 \rangle)
\end{aligned} \tag{31}$$

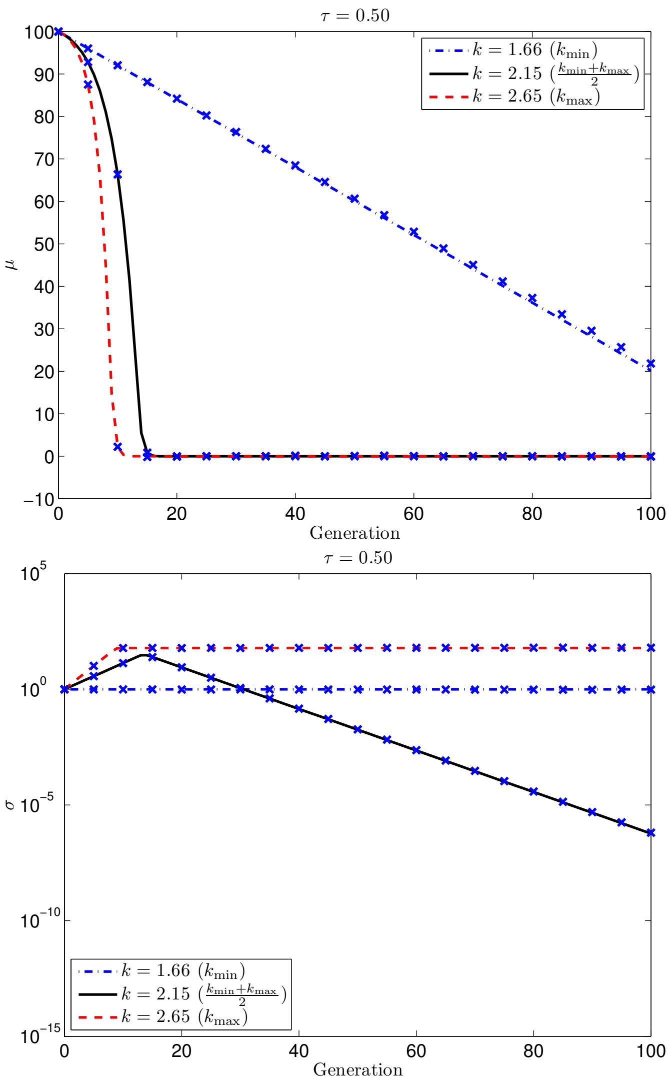

Fig. 6. Agreement of the model with the reality. Lines: values predicted by theoretical model. Crosses: experimental values.

Thus we need to compute the second moment of the truncated normal distribution:

$$\begin{aligned}
&E(Z^2|Z \in \langle z_1, z_2\rangle)) = \\
&= \int_{-\infty}^{\infty} z^2 \cdot \phi_T(z) dz = \\
&= \frac{1}{\Phi(z_2) - \Phi(z_1)} \int_{z_1}^{z_2} z^2 \cdot \phi(z) dz \left| \begin{array}{ll} u & = z \\ v' & = z \cdot \phi(z) \\ u' & = 1 \\ v & = -\phi(z) \end{array} \right. = \\
&= [-z \cdot \phi(z)]_{z_1}^{z_2} + \int_{z_1}^{z_2} \phi(z) dz = \\
&= -\frac{z_2 \cdot \phi(z_2) - z_1 \cdot \phi(z_1)}{\Phi(z_2) - \Phi(z_1)} + \frac{\Phi(z_2) - \Phi(z_1)}{\Phi(z_2) - \Phi(z_1)} = \\
&= 1 - \frac{z_2 \cdot \phi(z_2) - z_1 \cdot \phi(z_1)}{\Phi(z_2) - \Phi(z_1)} \qquad (32)
\end{aligned}$$

Substituting 29 and 32 into 31 we get

$$\begin{aligned}
&\mathrm{Var}(Z|Z \in \langle z_1, z_2\rangle) = \\
&= 1 - \frac{z_2 \cdot \phi(z_2) - z_1 \cdot \phi(z_1)}{\Phi(z_2) - \Phi(z_1)} - \left(\frac{\phi(z_2) - \phi(z_1)}{\Phi(z_2) - \Phi(z_1)}\right)^2 \qquad (33)
\end{aligned}$$

Again, for random variable X, $X = \mu + \sigma Z$, we can write

$$\mathrm{Var}(X|X \in \langle x_1, x_2\rangle) = \sigma^2 \cdot \mathrm{Var}(Z|Z \in \langle z_1, z_2\rangle), \qquad (34)$$

where $z_i = \frac{x_i - \mu}{\sigma}$.

References

[1] P. Larrañaga and J. A. Lozano, Eds., *Estimation of Distribution Algorithms*, ser. GENA. Kluwer Academic Publishers, 2002.

[2] P. Larrañaga, J. A. Lozano, and E. Bengoetxea, "Estimation of distribution algorithms based on multivariate normal distributions and gaussian networks," Dept. of Computer Science and Artificial Intelligence, University of Basque Country, Tech. Rep. KZZA-IK-1-01, 2001.

[3] P. A. N. Bosman and D. Thierens, "Expanding from discrete to continuous estimation of distribution algorithms: The IDEA," in *PPSN VI: Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*. London, UK: Springer-Verlag, 2000, pp. 767–776. [Online]. Available: http://portal.acm.org/citation.cfm?id=668936

[4] S. Rudlof and M. Köppen, "Stochastic hill climbing by vectors of normal distributions," in *First Online Workshop on Soft Computing*, Nagoya, Japan, 1996.

[5] C. W. Ahn, R. S. Ramakrishna, and D. E. Goldberg, "Real-coded bayesian optimization algorithm: Bringing the strength of BOA into the continuous world," in *Proceedings of the Genetic and Evolutionary Computation Conference GECCO 2004*, K. Deb, Ed. Springer-Verlag, Berlin, 2004, pp. 840–851. [Online]. Available: http://www.springerlink.com/content/fxmhuudye73tnh04

[6] B. Yuan and M. Gallagher, "On the importance of diversity maintenance in estimation of distribution algorithms," in *Proceedings of the Genetic and Evolutionary Computation Conference GECCO-2005*, H. G. Beyer and U. M. O'Reilly, Eds., vol. 1. New York, NY, USA: ACM Press, 2005, pp. 719–726. [Online]. Available: http://portal.acm.org/citation.cfm?id=1068009.1068129

[7] J. Ocenasek, S. Kern, N. Hansen, and P. Koumoutsakos, "A mixed bayesian optimization algorithm with variance adaptation," in *Parallel Problem Solving from Nature - PPSN VIII*, X. Yao, Ed. Springer-Verlag, Berlin, 2004, pp. 352–361.

[8] J. Grahl, S. Minner, and F. Rothlauf, "Behaviour of UMDAc with truncation selection on monotonous functions," in *IEEE Congress on Evolutionary Computation, CEC 2005*, vol. 3, 2005, pp. 2553–2559.

[9] J. Grahl, P. A. N. Bosman, and F. Rothlauf, "The correlation-triggered adaptive variance scaling IDEA," in *Proceedings of the 8th annual conference on Genetic and Evolutionary Computation Conference - GECCO 2006*. New York, NY, USA: ACM Press, 2006, pp. 397–404.

[10] P. A. N. Bosman, J. Grahl, and F. Rothlauf, "SDR: A better trigger for adaptive variance scaling in normal EDAs," in *GECCO '07: Proceedings of the 9th annual conference on Genetic and Evolutionary Computation*. New York, NY, USA: ACM Press, 2007, pp. 492–499.

[11] B. Yuan and M. Gallagher, "A mathematical modelling technique for the analysis of the dynamics of a simple continuous EDA," in *IEEE Congress on Evolutionary Computation, CEC 2006*. Vancouver, Canada: IEEE Press, 2006, pp. 1585–1591.

[12] J. Grahl, P. A. N. Bosman, and S. Minner, "Convergence phases, variance trajectories, and runtime analysis of continuous EDAs," in *GECCO '07: Proceedings of the 9th annual conference on Genetic and evolutionary computation*. New York, NY, USA: ACM Press, 2007, pp. 516–522.