An analysis of iterated density estimation and sampling in the UMDA$_c$ algorithm

Jörn Grahl, Stefan Minner
Mannheim Business School
Dept. of Logistics
joern.grahl@bwl.uni-mannheim.de
minner@bwl.uni-mannheim.de

Franz Rothlauf
Mannheim Business School
Dept. of Information Systems
rothlauf@bwl.uni-mannheim.de

ABSTRACT

Of late, much progress has been made in developing Estimation of Distribution Algorithms (EDA), algorithms that use probabilistic modelling of high quality solutions to guide their search. While experimental results on EDA behaviour are widely available, theoretical results are still rare. This is especially the case for continuous EDA. In this article, we develop theory that predicts the effect of iterated density estimation and sampling used in the Univariate Marginal Distribution Algorithm in the continuous domain (UMDA$_c$). We use truncation selection and monotonous fitness functions. Monotonous functions are commonly used to model the algorithm behaviour far from the optimum. Our result includes formulae to predict population statistics in a specific generation as well as population statistics after convergence. We find that population statistics develop identically for monotonous functions. We show that if assuming monotonous fitness functions, the distance that UMDA$_c$ travels across the search space is bounded and solely relies on the percentage of selected individuals and not on the structure of the fitness landscape. This can be problematic if this distance is too small for the algorithm to find the optimum. Also, by wrongly setting the selection intensity, one might not be able to explore the whole search space.

1. INTRODUCTION

In recent years, much progress has been made in developing optimization techniques that use probabilistic models to guide the search for optimal or high quality solutions. These methods are called Estimation of Distribution Algorithms (EDA), Probabilistic Model Building Algorithms (PMBGA) or Iterated Density Estimation Algorithms (IDEAs). They constitute a rapidly growing, yet already established field of evolutionary computation.

EDA share the common feature that the joint distribution of promising solutions is estimated on the basis of an adequate probabilistic model. Then, new candidate solutions are generated by sampling from the model estimated. Genetic operators like crossover and mutation which are widely used in the field of traditional genetic algorithms (GA) are not used to generate new solutions. Rigorously applying this principle, EDA for the discrete domain (for example, see [9], [6], [2]), the continuous domain (for example [3], [1]) as well as the mixed discrete-continuous domain (see for example [8]) have been developed.

Their behaviour and area of applicability is most often demonstrated experimentally. But in contrast to the wide availability of experimental results, theoretical results on the behaviour of EDA are still rare. This is especially the case for real valued EDA.

Real valued EDA often use probabilistic models based on the Gaussian distribution to estimate the joint density of solutions and to sample candidate solutions. This is done for every generation as an iterative process.

Our work focuses on theory of iterated density estimation and sampling from gaussian distributions as used in the Univariate Marginal Distribution Algorithm in the continuous domain (UMDA$_c$). The UMDA$_c$ was first mentioned by [7]. The interested reader might want to consult the work by González, Lozano and Larrañaga ([4]). They analyze the UMDA$_c$ algorithm with tournament selection for linear and quadratic fitness functions.

In our paper, we present complementary work. We analyze the effect of iterated density estimation and sampling in UMDA$_c$ and present theory to predict its behaviour when the truncation selection scheme is used. We assume that the fitness function is monotonous. We focus on monotonous functions to model the effect of iterated density estimation and sampling when the algorithm is far from the optimum. We derive formulae that allow us to readily compute population statistics for a given generation t. Furthermore, we show how the population statistics converge, if t tends to infinity. We discuss why this behaviour can be problematic.

This article is structured as follows. In section 2, the UMDA$_c$ algorithm with truncation selection is introduced. Afterwards, we analyze the effect of the truncation selection scheme for use with monotonous fitness functions in section 3. Then, we derive analytical expressions in section 4. First, we show how population statistics change from generation t to generation $t+1$ (sections 4.3 and 4.4), then we analyze population statistics in generation t (section 4.5). Finally, we investigate convergence behaviour of UMDA$_c$ in section 4.6. Our results are discussed in section 4.7. In section 5, the paper ends with conclusions and a short outlook.

GECCO'05, June 25–29, 2005, Washington, DC, USA.
Copyright 2005 ACM 1-59593-010-8/05/0006 ...$5.00.

2. UMDA$_C$ WITH TRUNCATION SELECTION

In the following paragraphs, we describe the UMDA$_c$ algorithm with truncation selection.

UMDA$_c$ was introduced in [7] as a population-based technique for optimizing continuous functions. If we assume, that we want to maximize an n-dimensional function f by using the UMDA$_c$ algorithm, a single solution is represented as a vector $\boldsymbol{x} \in \mathbb{R}^n$. Using UMDA$_c$, we seek to find the vector $\boldsymbol{x}^\star$ that maximizes f.

A first population of candidate solutions is sampled uniformly from the set of feasible solutions. The fitness of each individual is evaluated using f. Now we use truncation selection of the best $\alpha \cdot 100\%$ individuals. This means, given a population size of Z individuals, we select the $\alpha \cdot Z$ best individuals. Selection, pushes the population towards promising regions of the search space.

From these selected individuals the following probabilistic model is estimated. In UMDA$_c$ it is assumed, that the joint distribution of the selected individuals follows a n-dimensional normal distribution that factorizes over n univariate normals. That is, the covariances between all x_i and x_j are 0 for all $i \neq j$. Thus, in generation t, the n variables $X_{1\ldots n}$ follow a univariate normal distribution with mean μ_i^t and standard deviation σ_i^t:

$$\begin{aligned} X_i &\sim N(\mu_i^t, \sigma_i^t) \\ P(X_i = x_i) &= \phi_{\mu_i^t, \sigma_i^t} \\ &= \frac{1}{\sqrt{2\pi}\sigma_i^t} \exp\left\{ -\frac{(x_i - \mu_i^t)^2}{2(\sigma_i^t)^2} \right\} \end{aligned} \tag{1}$$

The parameters μ_i^t and σ_i^t are estimated from the selected individuals by using the well known maximum likelihood estimators for moments of the univariate normal distribution.

Now, new individuals are generated by sampling from the estimated joint density. These individuals completely replace the old population. The algorithm continues with truncation selection again. This process is iterated until a termination criterion is met. For further details on UMDA$_c$, we refer to [7].

Summarizing, the algorithm tries to find the optimum by iterative density estimation and sampling from Gaussian distributions. In UMDA$_c$, the essential parameters are μ_i^t and σ_i^t. We are interested in how these parameters change over time. Therefore, in the following section we make some assumptions on the structure of f.

3. MONOTONOUS FITNESS FUNCTIONS AND TRUNCATION SELECTION

In the previous section, the UMDA$_c$ algorithm was introduced. We saw that a fitness function f needs to be specified for evaluating the population. In many work on parameter optimization, highly complex fitness functions are used as benchmark problems. These functions exhibit deceptive structures, multiple optima, and other features that makes optimizing them a stiff test.

In this article, we make simplified assumptions on the structure of f. In particular, we focus on the behaviour of UDMAc on monotonous functions. This set of functions includes, but is not limited to linear ones. This can be seen as a way to model the structure of the search space far away from the optimum. A nice side effect is reduced complexity of our analytical analysis.

In the following we will introduce monotonous fitness functions. Furthermore, We will show an interesting result that occurs when combining truncation selection with monotonous fitness functions. This result will be of some importance later on.

Let S be a population of individuals. Let x_j and $x_k \in S$ be two distinct individuals of the population and let $g_i : \mathbb{R} \rightarrow \mathbb{R}$ be a fitness function defined over the $i - th$ gene of the individuals. Then,

$$\begin{aligned} &g_i \text{ is increasing if } x_j \leq x_k \text{ implies that} \\ &\quad g_i(x_j) \leq g_i(x_k)\ \forall x_j, x_k \in S \\ &g_i \text{ is decreasing if } x_j \leq x_k \text{ implies that} \\ &\quad g_i(x_j) \geq g_i(x_k)\ \forall x_j, x_k \in S. \end{aligned} \tag{2}$$

We consider a fitness landscape monotonous if the fitness function f is either increasing or decreasing. Note that the class of monotonous functions includes, but is not limited to linear functions.

Assume that a population P of individuals is given. We use l different increasing functions $f_{1\ldots l}$ to evaluate this population P. After each evaluation process, we use truncation selection of the best $\alpha \cdot 100\%$ of the individuals and call the l sets of selected individuals $M_{1\ldots l}$. It is a simple, yet interesting fact that all sets $M_{1\ldots l}$ have to be identical. Note that the fitness of the selected individuals may of course be different. For our analysis it is more important that the selected individuals are identical. Note that if all fitness functions are decreasing, this fact is true as well.

UMDA$_c$ uses density estimation and sampling to generate new candidate solutions. In the density estimation process, the fitness of the individuals is not considered. Density estimation solely relies on the genotypes of the selected individuals, which is the x. As the parameters μ_i^t and σ_i^t are estimated from the x, they are identical for all $f_{1\ldots l}$.

This fact simplifies our further analysis. We can now state that the UMDA$_c$ will behave the same for all increasing fitness functions (and for all decreasing functions). Thus, we can base our analysis on the simplest monotonous function that is the linear one. Yet, we know that our results are valid for all monotonous functions.

4. UMDA$_C$ FOR MONOTONOUS FITNESS FUNCTIONS

In the following paragraphs, we model the behaviour of UMDA$_c$ with truncation selection on fitness functions of the type

$$f(\boldsymbol{x}) = \sum_{i=1}^{n} g_i(x_i), \tag{3}$$

where $g_i(x_i)$ is an increasing function. Note, that the case of decreasing functions does not provide additional insight. Thus, we focus our analysis on increasing functions.

The specific structure of the fitness function allows us a decomposition. UMDA$_c$ factorizes over n univariate normals. The fitness function consists of a sum of n univariate monotonous functions.

Thus, we can reduce our analysis to the analysis of one single $g_i(x_i)$. We develop mathematical expressions that model how μ_i^t and σ_i^t change over time. More specifically, we are interested in μ_i^{t+1} and σ_i^{t+1} given the corresponding parameter values at generation t.

Furthermore, we analyze population statistics for a specific generation and the limit behaviour of UMDA$_c$. This means we investigate the convergence of μ_i^t and σ_i^t for $t \to \infty$.

4.1 Notation

The following symbols are used in our analysis:

$\phi(x)$	Standard normal density at value x
$\phi_{\mu,\sigma}(x)$	Normal density with mean μ and standard deviation σ at value x
$\Phi(x)$	Cumulative standard normal density, $x \cdot 100\%$ quantile
$\Phi_{\mu,\sigma}(x)$	Cumulative density of normal distribution with mean μ and standard deviation σ, $x \cdot 100\%$ quantile
$\Phi^{-1}(x)$	Quantile function of standard normal density
$\Phi^{-1}_{\mu,\sigma}(x)$	Quantile function of normal distribution with mean μ and standard deviation σ

4.2 Monotonous fitness functions and truncation selection

We analyze the truncation selection step in presence of fitness functions of type (3).

Due to the structure of UMDA$_c$'s probabilistic model and the structure of $f(\boldsymbol{x})$, we can decompose the fitness function and analyze the behaviour of each μ_i and σ_i independently.

As we have seen in section 3, we can simplify our approach even further and replace all $g_i(x_i)$ by linear functions of the form $y_i(x_i) = a_i \cdot x_i + b_i$ and study how truncation selection influences the population statistics.

In UMDA$_c$, new candidate solutions are generated by sampling new individuals x_i from a normal distribution with mean μ_i^t and variance $(\sigma_i^t)^2$. The fitness y_i is obtained from a linear function $y_i = a_i \cdot x_i + b_i$. As the x are realizations of a random variable, the fitness also is a random variable. The distribution of the fitness Y can be expressed in terms of a normal random variable with mean μ_f and variance σ_f^2:

$$\begin{aligned} \mu_f &= a_i \cdot \mu_i + b_i \\ \sigma_f^2 &= a_i^2 \cdot (\sigma_i^t)^2 \end{aligned} \tag{4}$$

By truncation selection, the best $\alpha \cdot 100\%$ of the individuals are selected. Thus, all individuals with a fitness larger than a fitness minimum of y_m are selected. Their probabilities sum up to α. Since the fitness is normally distributed, y_m can be obtained from the quantile function of the normal distribution as follows:

$$\begin{aligned} y_m &= \Phi^{-1}_{\mu_f,\sigma_f}(1-\alpha) \\ &= \Phi^{-1}(1-\alpha) \cdot a_i\sigma_i^t + a_i\mu_i^t + b_i \end{aligned}$$

Statistically, the fitness distribution is truncated from below at y_m. We refer to y_m as the fitness truncation point. Now, we are interested in the individual x_m that corresponds to the fitness value y_m. All individuals $x_i > x_m$ are selected. We call x_m the corresponding population truncation point. This is illustrated graphically in figure 1. The fitness truncation point y_m can be transformed into the population truncation x_m as follows.

$$\begin{aligned} x_m &= y^{-1}(y_m) \\ &= \frac{(y_m - b_i)}{a_i} \\ &= \frac{\Phi^{-1}(1-\alpha) \cdot a_i\sigma_i^t + a_i\mu_i^t + b_i - b_i}{a_i} \qquad a_i \neq 0 \\ &= \Phi^{-1}_{\mu_i^t,\sigma_i^t}(1-\alpha) \end{aligned} \tag{5}$$

It is interesting to see that obviously the truncation point is independent from a_i and $b_i (a_i > 0)$. Put differently, no matter which linear function with $a_i > 0$ we choose, the population truncation point remains the same. Thus, the effect of selection is independent from a_i and b_i, for all $a_i > 0$. No matter how these parameters are chosen, the same individuals are selected.

As we have seen in this section, the selected individuals are identical for all linear functions $y_i(x_i)$, where $a_i > 0$. Furthermore, the population truncation point solely relies on statistical parameters of the population. Thus, the selected individuals can be obtained from the population statistics directly without taking a look at the fitness landscape.

4.3 Change of μ_i^t to μ_i^{t+1}

We have seen that the effect of truncation selection is identical for all linear functions y_i with positive gradient. We have also shown, that for predicting the behaviour of UMDA$_c$ we do not need to model the distribution of the fitness. The change in population statistics can be obtained from the population statistics directly. In this section, we derive mathematical expressions for the change of the population mean from generation t to generation $t+1$.

We model selection by truncation of the normally distributed population density. Assume that a population is distributed with mean μ_i^t and standard deviation σ_i^t. The fitnesses of the individuals are calculated and the best $\alpha \cdot 100\%$ individuals are selected. This equals a left-truncation of the normal distribution in x_m.

To do this, we first refer to results from econometric literature on the truncated normal distribution (see [5], appendix). We start with presenting the moments of a doubly truncated normal distribution. A doubly truncated normal distribution with mean μ and variance σ^2, where $x_a < x < x_b$ can be modeled as a conditional density where $x \in A = [x_a, x_b]$ and $-\infty < x_a < x_b < +\infty$.

$$f(X|X \in A) = \frac{\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)}{\Phi\left(\frac{x_b-\mu}{\sigma}\right) - \Phi\left(\frac{x_a-\mu}{\sigma}\right)} \tag{6}$$

The moment generating function of this distribution is:

$$\begin{aligned} m(t) &= E(e^{tX}|X \in A) \\ &= e^{\mu t + \sigma^2 t^2/2} \cdot \frac{\Phi\left(\frac{x_b-\mu}{\sigma} - \sigma t\right) - \Phi\left(\frac{x_a-\mu}{\sigma} - \sigma t\right)}{\Phi\left(\frac{x_b-\mu}{\sigma}\right) - \Phi\left(\frac{x_a-\mu}{\sigma}\right)} \end{aligned} \tag{7}$$

From the moment generating function, we can derive the statistical moments of the distribution. We are interested in the mean. It can be computed as follows:

$$\begin{aligned} E(X|X \in A) &= m'(t)|_{t=0} \\ &= \mu - \sigma \cdot \frac{\phi\left(\frac{x_b-\mu}{\sigma}\right) - \phi\left(\frac{x_a-\mu}{\sigma}\right)}{\Phi\left(\frac{x_b-\mu}{\sigma}\right) - \Phi\left(\frac{x_a-\mu}{\sigma}\right)} \end{aligned} \tag{8}$$

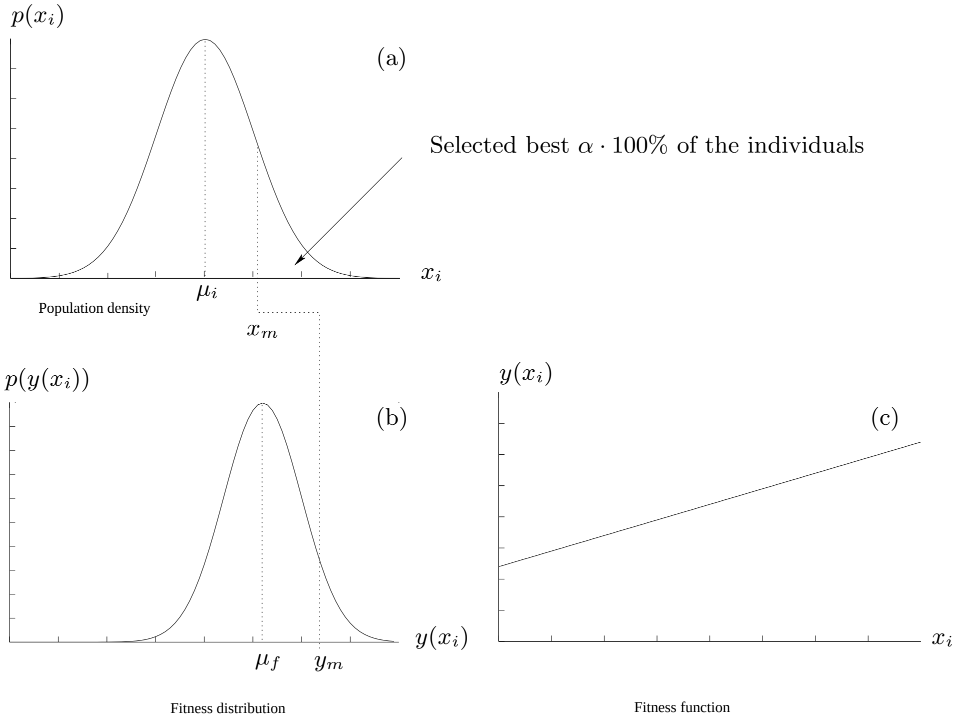

Figure 1: Impact of truncation selection of the best $\alpha \cdot 100\%$ individuals on the fitness distribution (b) and the population density (a). We assume an increasing fitness function (c). The fitness distribution is truncated from the left in y_m. The population distribution is truncated from the left in the corresponding point x_m.

We are not interested in the mean of a doubly truncated normal distribution, but in the mean of a left-truncated normal distribution. Thus, we now let x_b tend to infinity. This results in:

$$E(X|X > x_a) = \mu + \sigma \cdot \frac{\phi(\frac{x_a-\mu}{\sigma})}{\Phi(\frac{x_a-\mu}{\sigma})} \qquad (9)$$

From section 4.2 we know that $x_a = x_m = \Phi^{-1}_{\mu_i^t,\sigma_i^t}(1-\alpha)$. Inserting and rearranging leads to:

$$\begin{aligned} \mu_i^{t+1} &= E(X|X > x_m) \\ &= \mu_i^t + \sigma_i^t \cdot \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} \\ &= \mu_i^t + \sigma_i^t \cdot d(\alpha), \qquad \text{where } d(\alpha) = \frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} \end{aligned} \qquad (10)$$

Note that the mean of the population after applying truncation selection can now be easily computed. The factor $d(\alpha)$ is illustrated in figure 2. It can be seen that for $\alpha \to 1$ the factor $d(\alpha)$ converges to 0 leaving the mean of the population unchanged in $t+1$.

4.4 Change of σ_i^t to σ_i^{t+1}

Again, we model truncation selection by truncation of the normally distributed population density. Therefore, we again make use of the moment generating function as in section 4.3. We let $x_b \to \infty$. Finally, we get:

$$\begin{aligned} Var(X|X > x_m) &= E(X^2|X > x_m) - E(X|X > x_m)^2 \\ &= \sigma^2 \cdot \left\{ 1 + \frac{\frac{x_m-\mu}{\sigma} \cdot \phi\left(\frac{x_m-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x_m-\mu}{\sigma}\right)} \right. \\ &\qquad \left. - \left[\frac{\phi\left(\frac{x_m-\mu}{\sigma}\right)}{1 - \Phi\left(\frac{x_m-\mu}{\sigma}\right)} \right]^2 \right\} \end{aligned} \qquad (11)$$

We use this equation in the context of our model by assigning appropriate indices, inserting x_m, simplifying, and rearranging. This leads us to:

$$\begin{aligned} (\sigma_i^{t+1})^2 &= (\sigma_i^t)^2 \cdot \left\{ 1 + \frac{\Phi^{-1}(1-\alpha)\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} - \right. \\ &\qquad \left. \left[\frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} \right]^2 \right\} = (\sigma_i^t)^2 \cdot c(\alpha) \\ \text{where } c(\alpha) &= \left\{ 1 + \frac{\Phi^{-1}(1-\alpha)\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} - \right. \\ &\qquad \left. \left[\frac{\phi\left(\Phi^{-1}(\alpha)\right)}{\alpha} \right]^2 \right\} \end{aligned} \qquad (12)$$

Now, we can compute the population variance in $t+1$, given the population variance in t. The factor $c(\alpha)$ is plotted in figure 2. It can be seen, that if $\alpha \to 1$, the factor $c(\alpha)$ converges to 1, leaving the variance in generation $t+1$ unchanged.

4.5 Population statistics in generation t

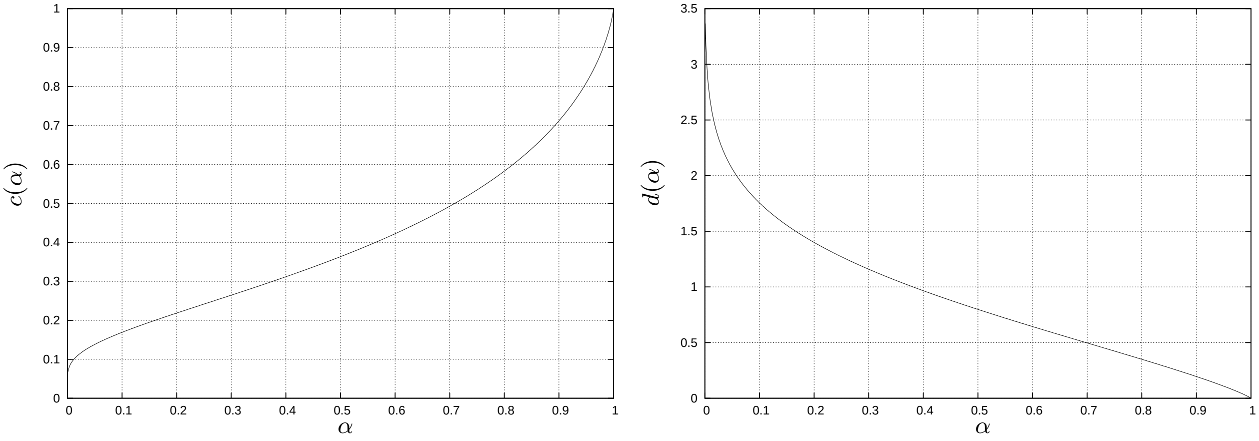

Figure 2: Illustration of $c(\alpha)$ and $d(\alpha)$

The last two subsections examined the change of the population statistics from one generation to the next generation.

Now, we calculate how the population mean and variance depend on t. To obtain the corresponding population statistics, we sum up the iterative formulae that have been developed in 4.3 and 4.4.
Doing this we get the following result for the mean after some calculations. In generation $t > 0$, the mean μ_i^t can be computed as:

$$\mu_i^t = \mu_i^0 + \sigma_i^0 \cdot c(\alpha) \cdot \sum_{i=1}^{t} \sqrt{d(\alpha)^{i-1}} \tag{13}$$

Similarly, in generation t ,the variance σ_t^2 can be computed as:

$$(\sigma_i^t)^2 = (\sigma_i^0)^2 \cdot c(\alpha)^t \tag{14}$$

4.6 Convergence of population statistics for $t \to \infty$

In this section, we analyze convergence of UMDA$_c$. That means that we analyze how the population statistics develop over time, assuming that $t \to \infty$.

First, we consider the mean. Therefore, we make use of (13). Note that the sum is the only part of the expression that depends on t. This leads us to:

$$\lim_{t\to\infty} \mu_i^t = \mu_i^0 + \sigma_i^0 \cdot c(\alpha) \cdot \underbrace{\lim_{t\to\infty} \sum_{k=1}^{t} \left[\sqrt{d(\alpha)^{(k-1)}}\right]}_{\text{infinite geometric series}} \tag{15}$$

The last part is an infinite geometric series that can be simplified. After doing this, we get the following result:

$$\lim_{t\to\infty} \mu_i^t = \mu_i^0 + \sigma_i^0 \cdot c(\alpha) \cdot \frac{1+\sqrt{d(\alpha)}}{1-d(\alpha)} \tag{16}$$

We can decompose the last factor, leading to:

$$\lim_{t\to\infty} \mu_i^t = \mu_i^0 + \sigma_i^0 \cdot \frac{1+\sqrt{1+e(\alpha)\cdot z(\alpha)}}{-z(\alpha)}, \tag{17}$$

with:

$$\begin{aligned} e(\alpha) &= \Phi^{-1}(1-\alpha) \\ h(\alpha) &= \frac{\phi(e(\alpha))}{\alpha} \\ z(\alpha) &= e(\alpha) - h(\alpha) \end{aligned} \tag{18}$$

This expression allows us to compute the maximum distance, that UMDA$_c$'s mean will move across the search space for a given selection intensity of $\alpha \cdot 100\%$ and monotonous fitness functions.

Now, we consider the variance. We make use of (14) and let t tend to infinity. Note that $0 < c(\alpha) < 1$. This leads to

$$\begin{aligned} \lim_{t\to\infty} (\sigma_i^t)^2 &= \lim_{t\to\infty} \left[(\sigma_i^0)^2 \cdot c(\alpha)^t\right] \\ &= 0 \end{aligned} \tag{19}$$

Thus, the variance converges towards 0.

4.7 Interpretation of results

In the previous paragraphs, we derived expressions to describe the behaviour of the UMDA$_c$ algorithm with truncation selection on monotonous functions.

We have seen that the algorithm converges since the population variance converges towards 0. The maximal distance that the mean of the population can move across the search space is bounded. This distance solely depends on

- the mean of the first population,
- the variance of the first population,
- and the selection intensity α.

This has some important effects on the behaviour of UMDA$_c$. First, if the optimum of the search space lies outside this maximal distance, the algorithm can not find it and one will experience premature convergence. Furthermore, the first population is usually sampled uniformly in the space of feasible solutions. The exploration of the search space relies on density estimation and sampling. However, by choosing the amount of individuals that are selected, one can adjust the maximal distance that the mean of the population will move. One needs to be careful when choosing the selection intensity α. By wrongly setting α, the algorithm might not even be able to sample all feasible points.

5. CONCLUSION

In this article, we have analyzed the effect of iterative density estimation and sampling used in the UMDA$_c$ algorithm, when monotonous fitness functions and truncation selection is assumed. Monotonous functions include, but are not limited to, linear functions. We have developed mathematical expressions for the mean and the variance of the population.

Our findings are as follows. First, we have shown that using truncation selection of the best $\alpha \cdot 100\%$ of the individuals has an interesting effect. In this case, a linear fitness with any positive (negative) gradient can be used as a valid replacement for any increasing (decreasing) function. This replacement has no effect on the population statistics but makes the analysis easier. Furthermore, results obtained for linear fitness functions with positive (negative) gradient are valid for all increasing (decreasing) fitness functions.

Then we have analyzed the population statistics (mean and variance) for a given generation t. For doing this the fitness landscape does not need to be modeled explicitly.

We showed how the convergence behaviour of UMDA$_c$ depends on the selection pressure α. Furthermore, we obtained the maximal distance that UMDA$_c$'s mean will move across the search space for a monotonous function and found that UMDA$_c$ will behave identically on all of these fitness functions.

This can be problematic if the optimum lies outside this maximal distance. In this case, the algorithm will not find it but converge before. Furthermore, by wrongly setting α, the algorithm might even not be able to explore the entire search space at all. These are important limitations of the UMDA$_c$ algorithm with truncation selection.

Work is underway to apply the findings described above to predict success probability of UMDA$_c$. Also, we want to model UMDA$_c$ behaviour on peaks and analyze the case that covariances between the x_i are different from zero.

6. REFERENCES

[1] C. W. Ahn, R. S. Ramakrishna, and D. E. Goldberg. The real-coded bayesian optimization algorithm: Bringing the strength of boa into the continuous world. In *Proceedings of the GECCO 2004*, 2004.

[2] S. Baluja. Population-based incremental learning: A method for integrating genetic search based function optimization and competitive learning. Technical Report CMU-CS-94-163, Carnegie Mellon University, 1994.

[3] P. A. N. Bosman. *Design and Application of Iterated Density-Estimation Evolutionary Algorithms.* PhD thesis, University of Utrecht, Institute of Information and Computer Science, 2003.

[4] C. González, J. A. Lozano, and P. Larrañaga. Mathematical modelling of UMDAc algorithm with tournament selection. Behaviour on linear and quadratic functions. *International Journal of Approximate Reasoning*, 31(3):313–340, 2002.

[5] W. H. Greene. *Econometric analysis*, volume 5. Prentice Hall, Upper Saddle River, 2003.

[6] G. Harik, F. G. Lobo, and D. E. Goldberg. The compact genetic algorithm. In *Proceedings of the IEEE Conference on Evolutionary Computation*, pages 523–528, 1998.

[7] P. Larrañaga, R. Etxeberria, J. A. Lozano, and J. M. Peña. Optimization in continuous domains by learning and simulation of Gaussian networks. In A. S. Wu, editor, *Proceedings of the 2000 Genetic and Evolutionary Computation Conference Workshop Program*, pages 201–204, 2000.

[8] J. Ocenasek and J. Schwarz. Estimation of distribution algorithm for mixed continuous-discrete optimization problems. In *2nd Euro-International Symposium on Computational Intelligence*, pages 227–232, 2002.

[9] Martin Pelikan. *Bayesian optimization algorithm: From single level to hierarchy.* PhD thesis, University of Illinois at Urbana-Champaign, Dept. of Computer Science, Urbana, IL, 2002. Also IlliGAL Report No. 2002023.