Convergence Analysis of $UMDA_C$ with Finite Populations: A Case Study on Flat Landscapes

Bo Yuan
Division of Informatics
Graduate School at Shenzhen
Tsinghua University
Shenzhen 518055, P.R. China
+86-755-26036067
yuanb@sz.tsinghua.edu.cn

Marcus Gallagher
School of Information Technology and Electrical Engineering
The University of Queensland
QLD 4072, Australia
+61-7-33656197
marcusg@itee.uq.edu.au

ABSTRACT

This paper presents some new analytical results on the continuous Univariate Marginal Distribution Algorithm ($UMDA_C$), which is a well known Estimation of Distribution Algorithm based on Gaussian distributions. As the extension of the current theoretical work built on the assumption of infinite populations, the convergence behavior of $UMDA_C$ with finite populations is formally analyzed. We show both analytically and experimentally that, on flat landscapes, the Gaussian model in $UMDA_C$ tends to collapse with high probability, which is an important fact that is not well understood before.

Categories and Subject Descriptors

G.1.6 [**Numerical Analysis**]: Optimization – *global optimization.*

General Terms

Algorithms, Experimentation, Theory

Keywords

EDAs, $UMDA_C$, Theory, Finite Population

1. INTRODUCTION

As an alternative paradigm to traditional Evolutionary Algorithms (EAs), Estimation of Distribution Algorithms (EDAs) conduct searching by employing a statistical model instead of applying genetic operators at the individual level [11]. In each generation, a statistical model (e.g., Bayesian Networks or Gaussian Networks) is estimated using a set of selected individuals (typically the best ones according to the fitness function) from the current population. All new individuals in the next generation are generated by sampling from this model, which represents the probability distribution (structure) of the best current individuals.

Since new individuals in EDAs are generated by sampling from a well-defined statistical model, the process of evolution can be fully specified by the model parameters over all generations. This provides the opportunity to analyze and draw specific theoretical conclusions on the convergence behavior of EDAs. By contrast, due to the lack of such a centralized mechanism guiding the evolution process, it is usually very difficult to characterize the detailed behavior of traditional EAs unless significant assumptions are made to simplify the algorithm and/or the problem [8, 9].

Most of the existing theoretical results on the convergence and properties of EDAs have been developed in combinatorial and discrete domains [4, 5, 10, 12, 13, 15, 18]. For example, Shapiro [15] proves that, on a flat landscape, Population Based Incremental Learning (PBIL, with binary representation) [2] will converge to one of the corners of the solution space hypercube due to the phenomenon known as drift.

In recent years, theoretical research in the continuous domain has received growing attention from the EDA community. Zhang et. al [19] gives an important general proof on the global convergence of EDAs with three widely used selection operators: proportional, truncation and tournament selection. It is shown that with an infinite population, an EDA with a model given as an arbitrary probability density function will converge to the global optimum with probability 1 as long as the inferred model for the next generation has identical statistics as the selected set of parent solutions.

In addition to theoretical work focusing on the global asymptotic convergence, some progress has also been made towards a more detailed understanding of the dynamic behavior of EDAs. For example, González et. al [6] present some analysis of the continuous Univariate Marginal Distribution Algorithm ($UMDA_C$) [11] based on Gaussian distributions with tournament selection. Their analysis shows that the $UMDA_C$ model will converge prematurely (after a finite number of steps) on a linear fitness function (approximating far from optimum behavior) while it converges to the optimum on a quadratic function (approximating behavior close to an optimum) becomes slower when the dimensionality increases. Grahl et. al [7] give some theoretical results for $UMDA_C$ with truncation selection on monotonic functions. One of their major findings is that the distance that the Gaussian model can travel is bounded, with this distance dependent on the selection pressure instead of the structure of the landscape. This is in agreement with the previous result for a linear function[6]. Recently, Yuan et. al [16,17] present some analysis of $UMDA_C$ with truncation selection on unimodal and multimodal problems as well as the difference between two offspring selection strategies.

GECCO'09, July 8–12, 2009, Montréal Québec, Canada.

A key assumption of the abovementioned work is that of an infinite population. This assumption can significantly reduce the complexity of the analysis, as new individuals will have exactly the same statistics as the model from which they are sampled. In this sense, the behavior of $UMDA_C$ during evolution can actually be described by a deterministic process. In practice, however, population sizes are always finite, and may grow slowly compared to the dimensionality of the search space. Variations in the process of sampling from the probabilistic model in this situation are inevitable, and will subsequently have some impact on the new model to be built. As a result, with a finite population, the behavior of $UMDA_C$ can no longer be specified deterministically and its dynamics may not be well described by theory based on the infinite population assumption.

In this paper, an alternative approach towards more accurate understanding of the practical dynamics of $UMDA_C$ is developed, in terms of a probability distribution over each model parameter. Section 2 gives a brief description of the basic working mechanism $UMDA_C$. In Section 3, a formal analysis of the Gaussian model in $UMDA_C$ on flat landscapes is presented, which provides deep insights into the convergence behavior of $UMDA_C$ with finite populations. Experimental studies are conducted in Section 4 to provide empirical justification of our theoretical results. This paper is concluded in Section 5 with some discussions and future work.

2. THE FRAMEWORK OF $UMDA_C$

The framework of $UMDA_C$ is given in Table 1. Given that the $UMDA_C$ model is a product of univariate Gaussian distributions, it is sufficient here to consider a one-dimensional optimization problem and single univariate Gaussian model (i.e., the flat fitness functions are fully decomposable). The population P of size m is initialized by sampling from N (μ_0, σ_0^2), with the subscript on model parameters denoting time (generation number). In each generation, the n ($n<m$) best individuals are selected via truncation selection, from which a Gaussian model is built through maximum likelihood estimation of its parameters. The next population of m individuals is then generated by sampling from this newly created Gaussian model.

Table 1. The framework of $UMDA_C$ (1D)

Initialize the population of size m: $P \leftarrow N$ (μ_0, σ_0^2)
REPEAT for i=1,2,… until stopping criteria are met
Evaluate P
Select the best n individuals ($n<m$)
Estimate μ_i and σ_i^2 by maximum likelihood
$P \leftarrow N$ (μ_i, σ_i^2)
i=i+1
END REPEAT

Given a flat fitness function and an infinite population, it is easy to imagine that $UMDA_C$ will keep building new models (re-estimating model parameters) identical to the initial values, resulting in a stationary model. However, with a finite population, due to sampling errors, the process of model building is likely to be stochastic and it is interesting to give a formal investigation of the evolution of the Gaussian model.

3. $UMDA_C$ ON FLAT LANDSCAPES

3.1 Theoretical Analysis of σ^2

Firstly, consider the behavior of the model variance parameter σ_i^2 over generations. Since the landscape is assumed to be flat (identical fitness everywhere), using truncation selection, the n individuals are actually selected from m individuals at random (without replacement) and can be regarded as being directly sampled from the current Gaussian model. Also, the new model is to be built on these individuals using maximum likelihood estimation. As a result, its mean μ_i and variance σ_i^2 are identical to the sample mean and sample variance of the n individuals respectively. The selection pressure directly determines the size of the sample (n) used to estimate the model for the next generation. According to Cochran's Theorem[1], the relationship between the sample variance s^2 and the population (or model) variance σ^2 of Gaussian distributions is:

$$(n-1)\frac{s^2}{\sigma^2} \sim \chi^2_{n-1} \qquad (1)$$

According to Eq. 1, the ratio between the sample variance and the variance of the Gaussian model follows a scaled chi-square distribution with n-1 degrees of freedom. To describe this ratio, suppose that Z is a random variable following the chi-square distribution with n-1 degrees of freedom. It is known that the expected value and variance of Z are n-1 and 2(n-1) respectively. The sample variance s^2 can be rewritten as:

$$s^2 = \frac{Z \cdot \sigma^2}{n-1} \qquad (2)$$

Applying Eq. 2 iteratively (over generations), the variance of the Gaussian model to be built in the i^{th} generation is:

$$\sigma_i^2 = s_i^2 = \prod_{j=1}^{i} \frac{Z_j}{n-1} \cdot \sigma_0^2 = \prod_{j=1}^{i} Z'_j \cdot \sigma_0^2 \qquad \text{where} \quad Z'_j = \frac{Z_j}{n-1} \qquad (3)$$

In Eq. 3, Z_j is a random variable following the chi-square distribution and σ_0^2 is the initial model variance. The expected value and variance of Z_j' are 1 and 2/(n-1) respectively. The expected value of σ_i^2 is given by Eq. 4, which shows that it remains unchanged from the initial model variance value σ_0^2.

$$E(\sigma_i^2) = E(\prod_{j=1}^{i} \frac{Z_j}{n-1} \cdot \sigma_0^2) = \prod_{j=1}^{i} E(Z'_j) \cdot \sigma_0^2 = \sigma_0^2 \qquad (4)$$

The variance of σ_i^2 is given by:

$$Var(\sigma_i^2) = Var(\prod_{j=1}^{i} \frac{Z_j}{n-1} \cdot \sigma_0^2) = Var(\prod_{j=1}^{i} Z'_j) \cdot \sigma_0^4 \qquad (5)$$

Eq. 5 contains the variance of the product of i random variables. It is well known that the variance of the product of two independent variables u and v can be written as [3]:

[1] http://en.wikipedia.org/wiki/Cochran's_theorem

$$
\begin{aligned}
Var(u\cdot v) &= E(u^2\cdot v^2)-E(u)^2\cdot E(v)^2\\
&= E(u^2)\cdot E(v^2)-E(u)^2\cdot E(v)^2\\
&= (E(u)^2+\sigma_u^2)(E(v)^2+\sigma_v^2)-E(u)^2\cdot E(v)^2\\
&= E(u)^2\cdot\sigma_v^2+E(v)^2\cdot\sigma_u^2+\sigma_u^2\cdot\sigma_v^2
\end{aligned} \tag{6}
$$

For example, using this expansion the variance of σ_2^2 is given by:

$$
\begin{aligned}
Var(\sigma_2^2) &= Var(Z_1^{'}\cdot Z_2^{'})\cdot\sigma_0^4\\
&= (E(Z_1^{'})^2\cdot Var(Z_2^{'})+E(Z_2^{'})^2\cdot Var(Z_1^{'})\\
&\quad +Var(Z_1^{'})\cdot Var(Z_2^{'}))\cdot\sigma_0^4\\
&= \left(\frac{2}{n-1}+\frac{2}{n-1}+\frac{2}{n-1}\cdot\frac{2}{n-1}\right)\cdot\sigma_0^4=\frac{4n}{(n-1)^2}\cdot\sigma_0^4
\end{aligned} \tag{7}
$$

In the general case, it then follows that:

$$
\begin{aligned}
Var(\sigma_i^2) &= Var\left(\prod_{j=1}^{i-1}Z_j^{'}\cdot Z_i^{'}\right)\cdot\sigma_0^4\\
&= (E(\prod_{j=1}^{i-1}Z_j^{'})^2\cdot Var(Z_i^{'})+E(Z_i^{'})^2\cdot Var(\prod_{j=1}^{i-1}Z_j^{'})\\
&\quad +Var(\prod_{j=1}^{i-1}Z_j^{'})\cdot Var(Z_i^{'}))\cdot\sigma_0^4\\
&= \left(\frac{2}{n-1}+Var(\prod_{j=1}^{i-1}Z_j^{'})+\frac{2}{n-1}\cdot Var(\prod_{j=1}^{i-1}Z_j^{'})\right)\cdot\sigma_0^4\\
&> Var(\prod_{j=1}^{i-1}Z_j^{'})\cdot\sigma_0^4=Var(\sigma_{i-1}^2)
\end{aligned} \tag{8}
$$

According to the above analysis, it is now clear that: i). the expected value of σ_i^2 is a fixed value (equal to its initial value) during generations; ii).its variance increases monotonically.

Alternatively, the behaviour of σ_i^2 can be characterized as follows. Let Z_i^* represent the ratio between σ_i^2 and σ_0^2:

$$
Z_i^* = \prod_{j=1}^{i} Z_j^{'} \tag{9}
$$

From Eq. 9, it follows that:

$$
\log(Z_i^*)=\log\left(\prod_{j=1}^{i}Z_j^{'}\right)=\sum_{j=1}^{i}\log(Z_j^{'}) \tag{10}
$$

Since *log* (Z_i^*) is the sum of *i* independent random variables following the same distribution, it approximately follows a normal distribution as *i* gets large (Central Limit Theorem). As a result, Z_i^* is log-normally distributed [1]:

$$
f(Z_i^*;u,v)=\frac{e^{-(\log Z_i^*-u)^2/2v^2}}{Z_i^* v\sqrt{2\pi}} \tag{11}
$$

$$
\begin{aligned}
u &= \log(E(Z_i^*))-\frac{1}{2}\log\left(1+\frac{Var(Z_i^*)}{(E(Z_i^*))^2}\right)\\
&= -\frac{1}{2}\log(1+Var(Z_i^*))
\end{aligned} \tag{12}
$$

$$
v^2=\log\left(1+\frac{Var(Z_i^*)}{(E(Z_i^*))^2}\right)=\log(1+Var(Z_i^*)) \tag{13}
$$

Since Var (Z_i^*) increases monotonically (Eq. 8), the values of *u* in Eq. 12 and v^2 in Eq. 13 will consistently move towards negative and positive infinity respectively.

The cumulative distribution function of Z_i^* is known as:

$$
F(Z_i^*;u,v)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{\log(Z_i^*)-u}{v\sqrt{2}}\right] \tag{14}
$$

Consequently, the probability that Z_i^* is less than $0<\varepsilon\le 1$ is given by ($u=-0.5\,v^2$):

$$
F(\varepsilon;u,v)=\frac{1}{2}+\frac{1}{2}erf\left[\frac{2\log(\varepsilon)+v^2}{v2\sqrt{2}}\right] \tag{15}
$$

According to Eq. 15, for any ε between 0 and 1, the probability that Z_i^* is less than ε consistently increases during generations (the value of *v* increases monotonically and *log* (ε) is always negative).

In the meantime, the median (*exp* (*u*)) and mode (*exp* (u-v^2)) of Z_i^* both converge towards zero. Since Z_i^* is the ratio between σ_i^2 and σ_0^2, the values of σ_i^2 are expected to be mostly distributed close to zero as the evolution progresses. Note that its expected value is a constant, which means that σ_i^2 values significantly larger than σ_0^2 can be found in some cases.

Hence, the expected value of σ_i^2 is found to be always identical to the initial variance σ_0^2 but the variance of σ_i^2 monotonously increases during generations. However, the distribution of σ_i^2 is not symmetric with regard to its mean. Instead, it is proven that the values of σ_i^2 are mostly close to 0 when the number of generations gets large. Consequently, in practice, the Gaussian model is expected to collapse with high probability.

The above phenomenon is intuitively interesting and will happen in any experimental results with $UMDA_C$. Note that, with a small probability, the values of σ_i^2 could also get very large.

3.2 Theoretical Analysis of μ

The other parameter of the $UMDA_C$ model is the Gaussian mean μ_i. Given the initial Gaussian model *N* (μ_0, σ_0^2), it is well known that the sample mean of *n* individuals (equal to μ_1) follows a normal distribution:

$$
E(\mu_1)=\mu_0 \tag{16}
$$

$$
Var(\mu_1)=\frac{\sigma_0^2}{n} \tag{17}
$$

Applying Eq. 16 and Eq. 17 iteratively, μ_i can be represented by the sum of i random variables:

$$\mu_i = \mu_0 + \sum_{j=0}^{i-1} a_j \cdot \frac{\sigma_j}{\sqrt{n}} \quad (18)$$

In Eq. 18, a_j is a random variable following the standard normal distribution. The expected value and variance of μ_i are:

$$E(\mu_i) = E(\mu_0) + \sum_{j=0}^{i-1} E\left(a_j \cdot \frac{\sigma_j}{\sqrt{n}}\right) = \mu_0 \quad (19)$$

$$Var(\mu_i) = Var(\mu_0) + \sum_{j=0}^{i-1} Var\left(a_j \cdot \frac{\sigma_j}{\sqrt{n}}\right) \quad (20)$$

Since E (a_j) =0 and Var (a_j) =1, following the rule in Eq. 6:

$$\begin{aligned} Var\left(a_j \cdot \frac{\sigma_j}{\sqrt{n}}\right) &= \frac{1}{n} \cdot (E(a_j)^2 \cdot Var(\sigma_j) + Var(a_j) \cdot E(\sigma_j)^2 \\ &\quad + Var(a_j) \cdot Var(\sigma_j)) \\ &= \frac{1}{n} \cdot \left(E(\sigma_j)^2 + Var(\sigma_j)\right) = \frac{E(\sigma_j^2)}{n} = \frac{\sigma_0^2}{n} \end{aligned} \quad (21)$$

Substitute Eq. 21 into Eq. 20:

$$Var(\mu_i) = 0 + \sum_{j=0}^{i-1} \frac{\sigma_0^2}{n} = i \cdot \frac{\sigma_0^2}{n} \quad (22)$$

This shows that the expected value of the model parameter μ_i is always identical to the initial mean μ_0 while its variance increases linearly during generations. Consider the dynamics of the $UMDA_C$ model to be that of a certain type of random walk. The walk has a tendency to be symmetric about the initial mean parameter value, but its location becomes increasingly variable over time. After a large number of generations on a flat landscape, the model variance will often converge towards zero. However, it is also possible for the parameter to have an arbitrary large value after some number of generations. On other fitness functions, this behavior will still contribute to the model dynamics and could have an effect on the performance of the algorithm.

To illustrate the flat landscape dynamics of $UMDA_C$, Figures 1 and 2 show the mean and variance parameters respectively of the model for 5 trials, over 500 generations. The initial model parameters were determined from an initial population generated uniformly in [-5, 5]. The population size (m) was 100, and n=50. This shows the random walk behavior discussed above and the effect that the finite sample size has on the evolution of the model parameters.

4. EXPERIMENTS

To empirically verify the soundness of the above conclusions, an experiment was conducted using $UMDA_C$ on a flat landscape with the following parameters: m=50, n=20, μ_0 =0 and σ_0^2 =1. Note that the results are independent of the value of m. 100,000 independent trial runs of the algorithm were conducted.

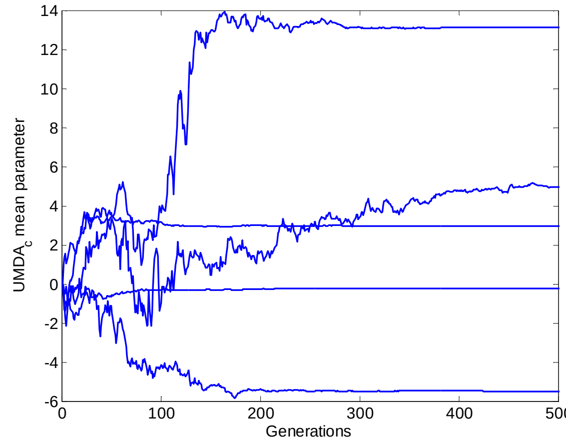

Figure 1. The evolution of the mean parameter over 5 individual trials.

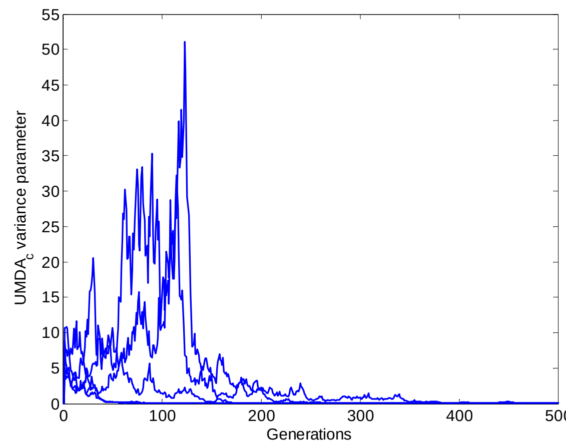

Figure 2. The evolution of the variance parameter over 5 individual trials.

The distributions of μ_i and σ_i^2 are plotted in Figures 3-6 respectively. The values predicted by the theoretical results above are shown as solid lines while the empirical simulation values are shown as circles. In all plots, both theoretical and empirical results match closely to each other. Figure 7 shows the histogram of the values of σ^2 after 5 generations. It is clear that the distribution is biased towards 0, although its mean value is around 1. However, there are trials, where the values of σ^2 did reach values significantly larger than σ_0^2.

Since the sample variance is likely to be less than the variance of the Gaussian model, in practice, $UMDA_C$ may actually converge faster than expected. To compensate for this effect, a simple modification to the algorithm is to scale the newly generated individuals so that their variance is identical to the variance of the Gaussian model.

In order to demonstrate the impact of this variance adjustment, an experiment was conducted with the following parameters: m=20, n=10, μ_0= -10 and σ_0^2 = 1. The test problem in this case was the quadratic (sphere) function f (x) $=\sum x_i^2$ (i=1, …, 5) and all results were averaged over 10,000 trials.

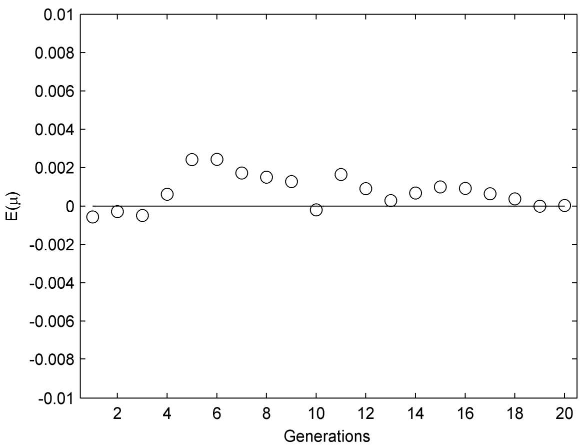

Figure 3. Theoretical and empirical mean values of *μ*.

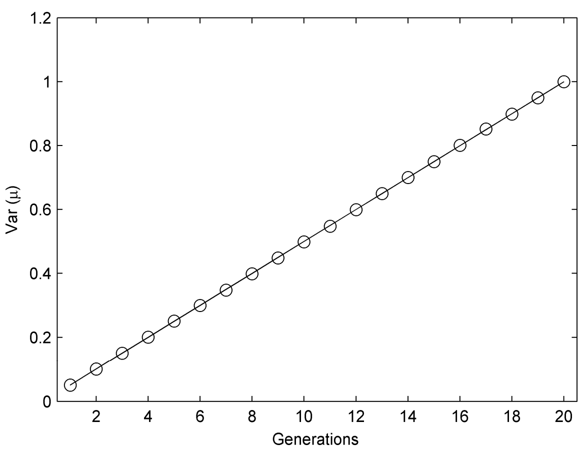

Figure 4. Theoretical and empirical variances of *μ*.

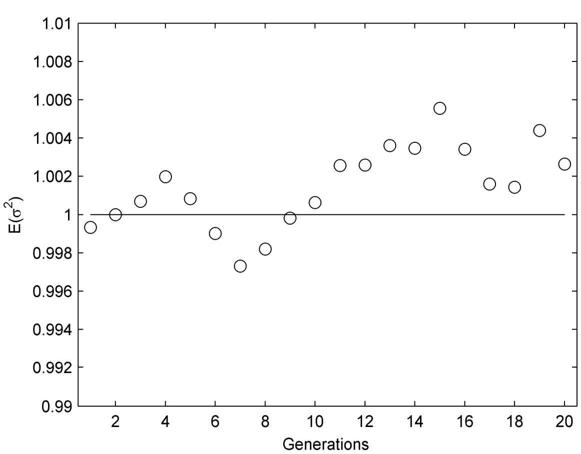

Figure 5. Theoretical and empirical mean values of σ^2.

The values of μ and σ^2 over generations are plotted in Figure 8 and Figure 9 respectively where the solid lines represent the original results and the solid lines with circles represent the results with variance adjustment. Note that there are five μ values and five σ^2 values plotted in each generation (one for each dimension), which

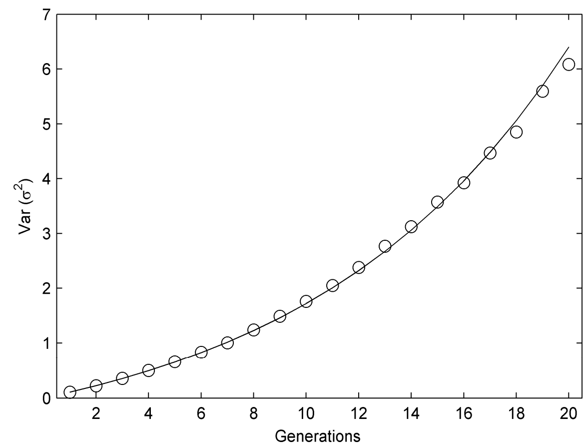

Figure 6. Theoretical and empirical variances of σ^2.

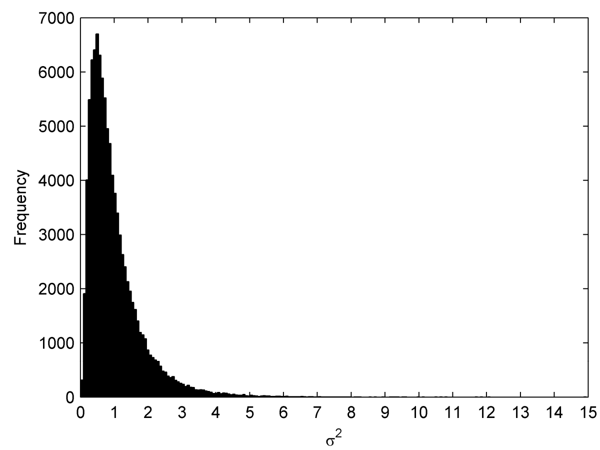

Figure 7. The distribution of σ^2 after 5 generations.

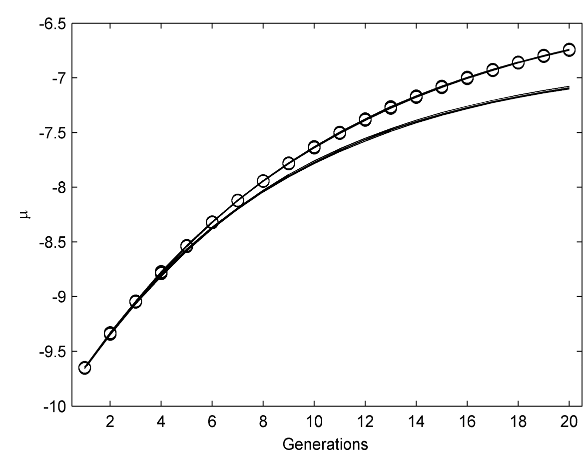

Figure 8. Comparison of *μ* with/without variance adjustment.

almost completely overlap each other. It is clear that the convergence speed (i.e., collapse of the model variance) was reduced when the variance adjustment was applied (Figure 9). Also, the mean of the Gaussian model made more progress towards the global optimum at the origin (Figure 8).

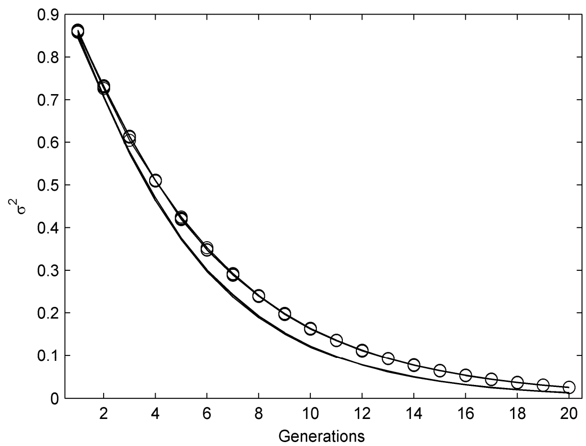

Figure 9. Comparison of σ^2 with/without variance adjustment.

5. CONCLUSIONS

The analysis presented in this paper provides a description of the dynamics of the $UMDA_C$ algorithm on a flat landscape when a finite population is used. As might be expected, the Gaussian model in $UMDA_C$ with finite populations tends to collapse (variance shrinks to zero, thereby causing the model to "converge") on flat landscapes. We show that this phenomenon is more likely to happen as the number of generations gets large. However, the analysis also reveals a more subtle phenomenon: it is also possible for the variances of the Gaussian model to go another way by becoming significantly larger than the initial variances, as the evolution of model parameters is governed by probability distributions. The implication is that, regardless of the test fitness function, the effect of finite sampling in model estimation should never be ignored, especially when the population size is small.

One of the major directions for future work is to perform a full analysis of $UMDA_C$ on non-flat landscapes. Also, it is important to extend the current results to higher dimensional spaces. It is expected that, for the general situations, the set of equations required could be very complicated, although the underlying principles and techniques used in this paper are still likely to be applicable to some extent.

6. REFERENCES

[1] Aitchison, J. and Brown, J. *The Lognormal Distribution with Special Reference to its Uses in Economics*. Cambridge: University Press, 1957.

[2] Baluja, S. *Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning*. Carnegie Mellon University Research Report CMU-CS-94-163, 1994.

[3] Barnett, H. The Variance of the Product of Two Independent Variables and Its Application to an Investigation Based on Sample Data. *Journal of the Institute of Actuaries*, 81, 190, 1955.

[4] Berny, A. Selection and Reinforcement Learning for Combinatorial Optimization. In *Proceedings of Parallel Problem Solving from Nature VI*, 2000, 601-610.

[5] Gonzalez, C., Lozano, J. and Larrañaga, P. Analyzing the PBIL algorithm by means of discrete dynamical systems. *Complex Systems*, 12 (4), 465-479, 2000.

[6] Gonzalez, C., Lozano, J. and Larrañaga, P. Mathematical Modelling of UMDAc algorithm with tournament selection. Behaviour on linear and quadratic functions. *International Journal of Approximate Reasoning*, 31, 313-340, 2000.

[7] Grahl, J., Minner, S. and Rothlauf, F. Behaviour of UMDAc with Truncation Selection on Monotonous Functions. In *Proceedings of 2005 Congress on Evolutionary Computation*, 2005, 2553-2559.

[8] He, J. and Yao, X. From an Individual to a Population: An Analysis of the First Hitting Time of Population-based Evolutionary Algorithms. *IEEE Transactions on Evolutionary Computation*, 6 (5), 495-511, 2002.

[9] He, J. and Yao, X. Towards an analytic framework for analysing the computation time of evolutionary algorithms. *Artificial Intelligence*, 145, 59-97, 2003.

[10] Hohfeld, M. and Rudolph, G. Towards a Theory of Population-Based Incremental Learning. In *Proceedings of IEEE International Conference on Evolutionary Computation (ICEC 97)*, pp. 1-6, 1997.

[11] Larrañaga, P. and Lozano, J. Eds., *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, 2002.

[12] Mühlenbein, H. The Equation for Response to Selection and Its Use for Prediction. *Evolutionary Computation*, 5 (3), 303 - 346, 1997.

[13] Mühlenbein, H. and Mahnig, T. Convergence Theory and Applications of the Factorized Distribution Algorithm. *Journal of Computing and Information Technology*, 7 (1), 19-32, 1999.

[14] Sebag, M. and Ducoulombier, A. Extending Population-Based Incremental Learning to Continuous Search Spaces. In *Proceedings of Parallel Problem Solving from Nature-V*, 1998, 418-427.

[15] Shapiro, J. Scaling of Probability-based Optimization Algorithms. In *Advances in Neural Information Processing Systems 15 (NIPS2002)*, 2002, 383-390.

[16] Yuan, B. *Towards Improved Experimental Evaluation and Comparison of Evolutionary Algorithms*. Ph.D. Thesis, The University of Queensland, 2006.

[17] Yuan, B. and Gallagher, M. A Mathematical Modelling Technique for the Analysis of the Dynamics of a Simple Continuous EDA. In *Proceedings of 2006 Congress on Evolutionary Computation*, 2006, 1585 - 1591.

[18] Zhang, Q. On stability of fixed points of limit models of univariate marginal distribution algorithm and factorized distribution algorithm. *IEEE Transactions on Evolutionary Computation*, 8 (1), 80-93, 2004.

[19] Zhang, Q. and Mühlenbein, H. On the convergence of a class of estimation of distribution algorithms. *IEEE Transactions on Evolutionary Computation*, 8(2), 127- 136, 2004.