A Mathematical Modelling Technique for the Analysis of the Dynamics of a Simple Continuous EDA

Bo Yuan, *Student Member*, *IEEE,* and Marcus Gallagher, *Member*, *IEEE*

Abstract — **This paper presents some initial attempts to mathematically model the dynamics of a continuous Estimation of Distribution Algorithm (EDA), based on a Gaussian distribution. Case studies are conducted on both unimodal and multimodal problems to highlight the effectiveness of the proposed technique and explore some important properties of the EDA. With some general assumptions, we show that, for one-dimensional unimodal problems and with the (μ, λ) scheme: (1). The behaviour of the EDA is dependent only on the general shape of the test function, rather than its specific form; (2). When initialized far from the global optimum, the EDA has a tendency to converge prematurely; (3). Given a certain selection pressure, there is a unique value for the proposed amplification parameter that could help the EDA achieve desirable performance; for one-dimensional multimodal problems: (1). The EDA could get stuck with the (μ, λ) scheme; (2). The EDA will never get stuck with the (μ+λ) scheme.**

I. INTRODUCTION

Estimation of Distribution Algorithms (EDAs)[8] refers to a relatively new paradigm of Evolutionary Algorithms (EAs) that is based on statistical Machine Learning techniques instead of conventional genetic operators such as crossover and mutation. The major advantage of EDAs is that they can explicitly build a probability density function based on the distribution of promising individuals and utilize the dependence information to conduct optimization efficiently, which is particularly important in the presence of strong epistasis among variables.

Usually, EAs are evaluated on an empirical basis while only limited progress has been made in theoretical analysis, especially compared to the large number of algorithms in the literature[1]. This is mainly due to their complex dynamics and massive parallel behaviour as well as a lack of appropriate modelling techniques.

In traditional EAs, new individuals are generated by directly manipulating current individuals through various stochastic genetic operators. Consequently, it is often difficult to precisely estimate the new population and thus predict the dynamics of the evolutionary process in each generation. Fortunately, this issue may be less challenging in EDAs because the search process is solely driven by a high-level statistical model and all individuals are generated by

Bo Yuan is with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (phone: +61-7-3365-1636; email: boyuan@itee.uq.edu.au)

Marcus Gallagher is with the School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD 4072, Australia (phone: +61-7-3365-6197; email: marcusg@itee.uq.edu.au)

sampling from this model. In order to describe the behaviour of an EDA, only the model needs to be estimated, which is often specified by a few parameters. This clarity in the mechanisms of EDAs indicates that it may be possible to conduct more detailed theoretical analysis[2, 4, 7, 9-11, 14].

In this paper, a general technique is proposed to model the behaviour of a continuous EDA based on a Gaussian distribution and truncation selection, which is aimed at precisely predicting the Gaussian model in each generation (specified by the mean and standard deviation parameters). The importance of this technique is that it can be used to formally investigate some fundamental issues of EDAs such as the influence of various algorithm factors and the problem structure on the observed behaviour of the algorithm.

In previous related work, González et. al and Grahl et. al [5, 6] present some theoretical results for the $UMDA_c$ algorithm, which is equal to the EDA used in this paper for one dimensional problems. This paper builds on and extends this work in several different respects. Specifically, we consider not only unimodal but also multimodal problems on which some interesting dynamics can be observed. Furthermore, in addition to the (μ, λ) scheme, the more complex (μ+λ) scheme is also taken into account (i.e., in both cases μ=λ).

II. ESTIMATION OF DISTRIBUTION ALGORITHMS

TABLE I
THE GENERAL FRAMEWORK OF THE GAUSSIAN EDA

Initialize and evaluate the starting population P
While stopping criteria not met
Select the top individuals P^{sel}
Fit a Gaussian model $G(\mu, \sum)$ to P^{sel}
Sample a set of individuals P' from $G(\mu, \sum)$
Evaluate new individuals in P'
Use P' as the new population
or
Select the top individuals from P U P'
End While

The general mechanism of EDAs is an iterative process of evolving statistical models such as Gaussian models, Gaussian Mixture models and Gaussian Networks that specify the distribution of the promising individuals in the current population. The major motivation is that, with the guidance of the statistical model, EDAs are expected to be

able to explicitly capture the dependence relationships among problem variables in selected individuals and may achieve faster convergence speed when complex dependence relationships exist in the problem of interest.

In the continuous EDA used in this paper (Table 1), a portion of top individuals in the current population P are selected using truncation selection at each generation and a multivariate Gaussian model specified by its mean and covariance matrix is then built using the maximum likelihood estimate. All new individuals P' are generated by sampling from the Gaussian model and the new population is created using either the (μ, λ) scheme (i.e., replace all old individuals) or the $(\mu+\lambda)$ scheme (i.e., choose the fittest individuals from the union of old and new individuals). Note that P and P' are of the same size in this framework $(\mu=\lambda)$.

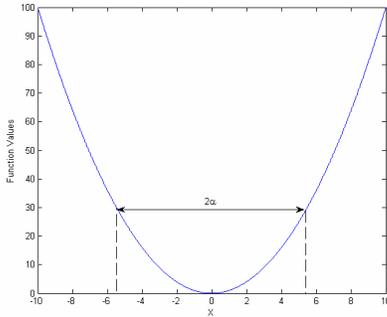## III. MODELLING UNIMODAL PROBLEMS

Fig. 1. A typical one-dimensional unimodal problem.

The 1D unimodal problems considered in this section have a general shape like the quadratic function $y=x^2$ shown in Figure 1. The only assumption made is that each problem is symmetric with regard to its global minimum, which is placed at the origin without loss of generality (i.e., there are no upper and lower boundaries on the search space). For the EDA, an infinitely large population size is assumed throughout the paper, meaning that new individuals have exactly the same statistics as the Gaussian model from which they are sampled. It is also assumed that the initial population is generated from a Gaussian with mean μ_0 and standard deviation δ_0. During evolution, the Gaussian model will be evolved towards the origin with changing mean and standard deviation values. An illustration of the dynamics of the EDA is shown in Figure 2.
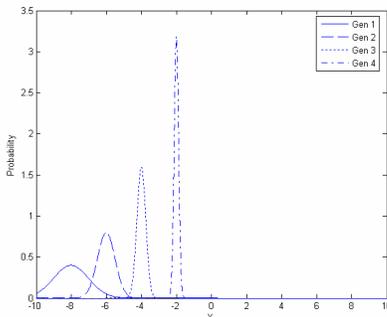

Fig. 2. A demo of the dynamics of the EDA on the unimodal problem.

With the (μ, λ) scheme, the Gaussian model constructed in the ith generation is completely determined by selected individuals in the ith population sampled from the Gaussian model built in the $(i-1)^{th}$ generation. As a result, there is a simple recursive relationship among the models in different generations. Another important feature is that the selected individuals in any generation are restricted within a continuous range $[-\alpha, \alpha]$ as shown in Figure 1 with the value of α depending on the model parameters as well as the selection pressure τ $(0<\tau<1)$. This is because truncation selection applies a deterministic fitness threshold and only individuals strictly better than it are chosen.

The specific value of α in the $(i+1)^{th}$ generation can be calculated by solving the following equation:

$$\int_{-a}^{a} \frac{1}{\sqrt{2\pi\delta_i^2}} e^{-\frac{(x-\mu_i)^2}{2\delta_i^2}} dx = \tau \qquad (1)$$

The meaning of Eq. 1 is that the cumulative probability of the Gaussian distribution within $[-\alpha, \alpha]$ should be equal to τ, which is the portion of individuals to be selected. Although it may be difficult to come up with an analytical solution to Eq.1, it is easy to see that the value of the left hand side increases from zero monotonically as $\alpha \rightarrow +\infty$, which makes it possible to use a simple line searching method to find the appropriate α value at a desired accuracy level.

Once the value of α is available, the mean parameter μ_{i+1} and standard deviation parameter δ_{i+1} are determined by the statistics of the individuals within $[-\alpha, \alpha]$:

$$\mu_{i+1} = \frac{1}{\tau} \cdot \int_{-a}^{a} x \cdot \frac{1}{\sqrt{2\pi\delta_i^2}} e^{-\frac{(x-\mu_i)^2}{2\delta_i^2}} dx \qquad (2)$$

$$\delta_{i+1}^2 = \frac{1}{\tau} \cdot \int_{-a}^{a} (x-\mu_{i+1})^2 \cdot \frac{1}{\sqrt{2\pi\delta_i^2}} e^{-\frac{(x-\mu_i)^2}{2\delta_i^2}} dx \qquad (3)$$

Now we have enough tools to describe the EDA's behaviour on unimodal problems. It should be pointed out that, in the above three equations, there is no utilization of any specific information of test functions. This shows that the convergence behaviour of the EDA in this situation is only dependent on the general structure of the problem (i.e., strictly unimodal, symmetric with regard to the origin).

In the following, some case studies are presented to formally investigate the dynamics of the EDA with regard to various algorithm factors. Figures 3&4 shows the standard deviations and the mean values of the EDA with $\mu_0 = -20$ and $\tau=0.3$, where the three lines in each plot represent results with $\delta_0=0.1$, 1.0 and 5.0 respectively.

There are a few interesting things that can be observed from this example:

- When the mean of the Gaussian model was far from the global optimum (i.e., several standard deviations away from the origin), the trajectories of the standard deviation values were approximately log-linear, which means that the standard deviations of the Gaussian model decreased exponentially.

- The gradient of the trajectory was independent of δ_0 (i.e., all standard deviation trajectories were parallel to each other).

- The EDA can get stuck (i.e., converge towards somewhere other than the global optimum) on unimodal problems even with an infinite population, especially with small initial standard deviations.
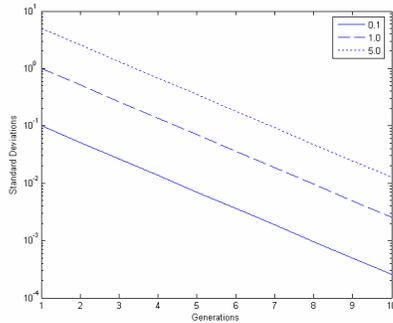

Fig. 3. The trajectories of the standard deviation parameter of the Gaussian model on unimodal problems with different initial δ_0 values.
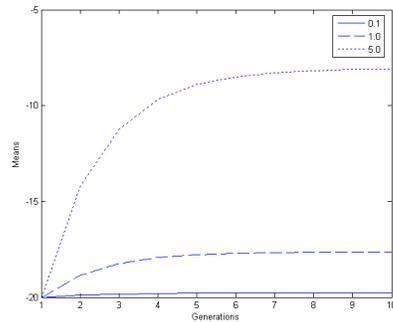

Fig. 4. The trajectories of the mean parameter of the Gaussian model on unimodal problems with different initial δ_0 values.

In order to demonstrate the influence of selection pressure on the dynamics of the EDA, the standard deviations and the mean values of the EDA with $\mu_0 = -20$ and $\delta_0=1.0$ are shown in Figures 5&6 where the three lines represent results with selection pressure $\tau =0.1$, 0.3 and 0.5 respectively. It is clear that as the selection pressure increased, the convergence speed of the EDA increased accordingly (i.e., the slope of the trajectory of the standard deviation values became steeper). Again, no satisfactory performance could be achieved as the EDA quickly became stuck somewhere far from the optimum.

Note that with weak selection pressure (e.g., τ=0.5), the EDA converged relatively slowly but still only made limited progress before getting stuck. This is because the larger the fraction of individuals to be selected, the smaller the distance between μ_i and μ_{i+1} becomes. In the extreme situation where τ=1.0, the EDA will simply keep building a new model based on the population generated from the previous model, which will result in a stationary statistical model and no progress could be expected. In practical situations where the population sizes are limited, the EDA is expected to present random walk behaviour.
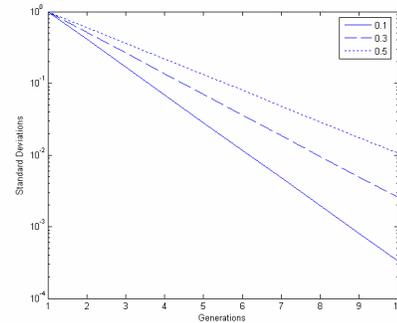

Fig. 5. The trajectories of the standard deviation parameter of the Gaussian model on unimodal problems with different selection pressures.
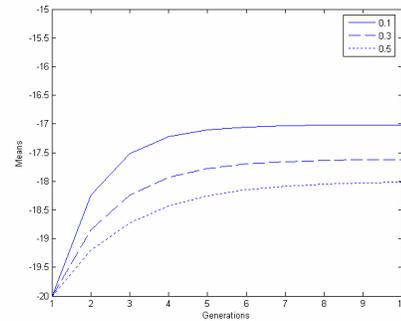

Fig. 6. The trajectories of the mean parameter of the Gaussian model on unimodal problems with different selection pressures.

In the above two case studies, the EDA was never able to converge to the global optimum and, as suggested by our previous research[12,13], it is necessary to explicitly maintain the population diversity to prevent the EDA from converging prematurely. A simple approach is to introduce an algorithm parameter γ, which is used to amplify the original standard deviation parameter (i.e., $\delta'_i= \gamma \cdot \delta_i$) in each generation of the algorithm. For γ=1.0, the EDA is unchanged. Figures 7 & 8 show the standard deviations and the mean values of the EDA with $\mu_0 = -20$, δ_0=1.0 and τ =0.3, where the three lines in each plot represent results with γ=1.0, 1.5 and 2.0 respectively. It is evident that this new parameter dramatically changed the dynamics of the EDA. With γ>1.0, the standard deviations were often orders of magnitude larger than in the original EDA. In the meantime, the mean of the Gaussian model also consistently moved towards the global optimum with γ=2.0, showing some significantly improved performance.

The above results do not, however, mean that such large γ values should be used in practice where the population size is limited. The issue is that the standard deviation may increase during evolution and become quite large, which makes the behaviour of the EDA similar to random search. Since the distance between the current model and the global optimum is typically not known, having the standard deviation increase or decrease may all produce undesirable results. As a result, a good strategy may be to choose a value of γ that keeps the current standard deviation value unchanged. Recall that, when the EDA is far from the global optimum, its convergence trajectory is only influenced by τ and is independent of δ_0. So, for different τ values, there are

different "optimum" γ values (i.e., $\gamma^* = \delta_i / \delta_{i+1}$). For example, the γ^* values for $\tau = 0.1$, 0.3 and 0.5 are approximately 2.4315, 1.9443 and 1.6589 respectively.
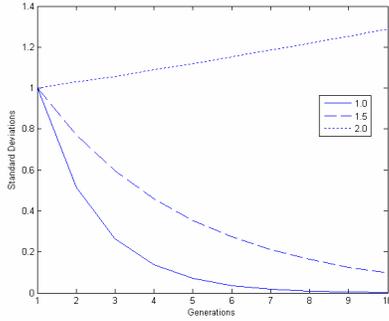

Fig. 7. The trajectories of the standard deviation parameter of the Gaussian model on unimodal problems with diversity maintenance.
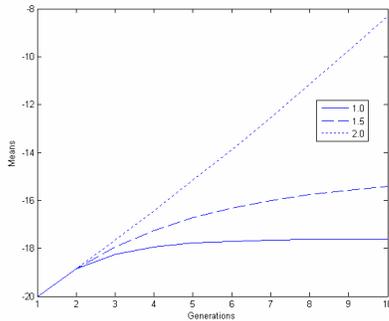

Fig. 8. The trajectories of the mean parameter of the Gaussian model on unimodal problems with diversity maintenance.

In order to demonstrate the effect of the optimum γ values, we investigated both empirically and theoretically the EDA with $\gamma^* = 1.9443$ in accordance with $\tau = 0.3$. The EDA started with $\mu_0 = -20$ and $\delta_0 = 1.0$ and, for empirical studies, the population size was 5000 and results were averaged over 25 independent trials.

Figure 9 shows that the standard deviation of the EDA model was almost unchanged within the first 16 generations due to the γ^* value. After that, when the EDA was quite close to the global optimum as shown in Figure 10, it started converging very quickly and changed its behaviour to local searching. This is because when the EDA model is far from the global optimum, the problem could be regarded as being monotonic and its behaviour is approximately consistent. However, the gradient of the convergence trajectory of the EDA gets significantly larger when it is close to the global optimum and the original γ^* value is no longer large enough to keep the standard deviation unchanged. This phenomenon could be roughly explained by the fact that when the Gaussian model is close to the optimum, selected individuals tend to be distributed within a smaller range than when the Gaussian model is far from the optimum.

In fact, this property is very useful as it can adapt the EDA's behaviour depending on its current situation. This is an important advantage over a hill-climbing algorithm with fixed step size. At last, the empirical results marked by '+' have a good match against the theoretical analysis, which verifies the soundness of our methods.
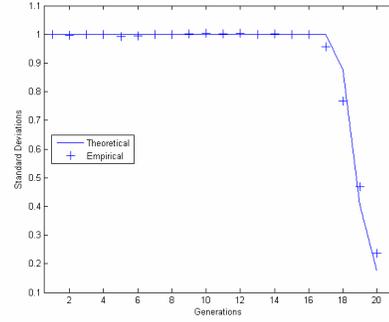

Fig. 9. The trajectory of the standard deviation parameter of the Gaussian model on unimodal problems with diversity maintenance ($\gamma^* = 1.9443, \tau = 0.3$).
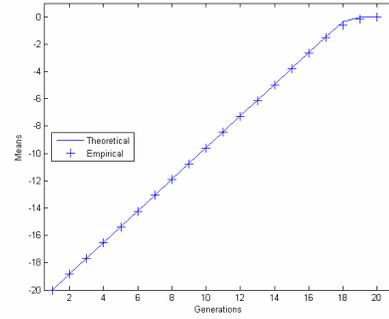

Fig. 10. The trajectory of the mean parameter of the Gaussian model on unimodal problems with diversity maintenance ($\gamma^* = 1.9443, \tau = 0.3$).

IV. MODELLING MULTIMODAL PROBLEMS

In this section, we will apply the techniques developed above in a more general way. Firstly, the problems to be solved have a global optimum and a local optimum. Secondly, in addition to the relatively simple (μ, λ) scheme, the $(\mu + \lambda)$ scheme is also considered, which selects the best individuals from the union of both old and new populations.

For the following analysis and experiments, test problems (i.e., to be maximized) are constructed by two Gaussian functions due to their smoothness and symmetric shape. Note that this has nothing to do with the Gaussian model of the EDA. The fitness value of an individual x is determined by the Gaussian function that gives it the larger value. Each Gaussian function is governed by its μ_j and δ_j while a parameter ω_j is used to adjust its height.

$$F(x) = \max\{G_1(x), G_2(x)\} \qquad (4)$$

with $\qquad G_j(x) = \omega_j \cdot \dfrac{1}{\sqrt{2\pi\delta_j^2}} e^{-\frac{(x-\mu_j)^2}{2\delta_j^2}} \qquad \omega_j > 0$

Two example functions are shown in Figures 11&12 with different parameter values. When there are two or more peaks in the problem, it is not always sufficient to describe the range of selected individuals by a single interval because they may come from different peaks. In Figure 11, when the worst fitness value (selection threshold) indicated by line A is greater than the value of the local optimum at x=0.2, all selected individuals will come from the peak corresponding

to the global optimum at x=-0.3 and bounded within [x₁, x₂], which is similar to the unimodal case. On the other hand, when the threshold is lower, as indicated by line B, the selected individuals will be from both peaks and two intervals [x₁, x₂] and [x₃, x₄] are required. Furthermore, when the two peaks are very close, it is possible that again only one interval is needed, as shown in Figure 12.
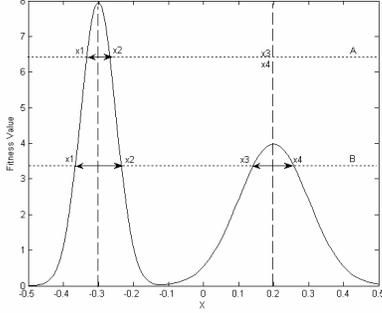

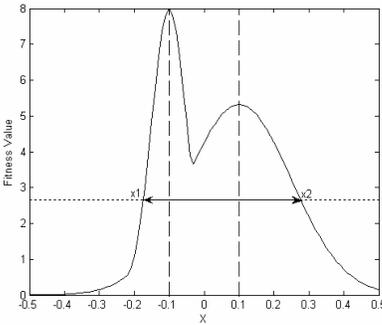Fig. 11. A multimodal problem with two separate optima.

Fig. 12. A multimodal problem with two close optima.

As before, the initial population is sampled from a Gaussian (μ_0, δ_0). In general, with the (μ, λ) scheme, the two intervals at the (i+1)th generation represented by [x₁, x₂] and [x₃, x₄] can be calculated by solving the following equation (i.e., solved by line searching in practice):

$$\int_{x_1}^{x_2} P(x,i)dx + \int_{x_3}^{x_4} P(x,i)dx = \tau \quad (5)$$

with

$$P(x,i) = \frac{1}{\sqrt{2\pi\delta_i^2}} e^{-\frac{(x-\mu_i)^2}{2\delta_i^2}}$$

subject to

$$G_1(x_1)=G_1(x_2)=G_2(x_3)=G_2(x_4)$$

In situations like the threshold A in Figure 11, the additional condition will reduce to $G_1(x_1)=G_1(x_2)$, assuming $G_1(x)$ corresponds to the global optimum. In situations like Figure 12, the two intervals [x₁, x₂] and [x₃, x₄] satisfying the additional condition may overlap with each other. Consequently, the two intervals must be merged together. In both cases x₃ and x₄ are set to some identical values, representing an empty interval.

Once the intervals are determined, the new model parameters are calculated in a similar manner as in the unimodal case (i.e., $\Omega = [x_1, x_2] \cup [x_3, x_4]$):

$$\mu_{i+1} = \frac{1}{\tau} \cdot \int_\Omega x \cdot P(x,i)dx \quad (6)$$

$$\delta_{i+1}^2 = \frac{1}{\tau} \cdot \int_\Omega (x-\mu_{i+1})^2 \cdot P(x,i)dx \quad (7)$$

The major difference between the (μ, λ) scheme and the ($\mu+\lambda$) scheme is that the latter may transfer old individuals of high quality into the next generation. That is to say, the ith population is not solely sampled from the statistical model built in the (i-1)th generation. Instead, it can be decomposed into a subset of individuals sampled from the initial Gaussian model and i-1 subsets of individuals sampled from the previous i-1 models respectively. Following this rule, it is possible to write down the new equations with the similar meaning as Eqs. 5-7:

$$\sum_{k=0}^{i}\left[\int_{x_1}^{x_2} P(x,k)dx + \int_{x_3}^{x_4} P(x,k)dx\right] = \tau \quad (8)$$

$$\mu_{i+1} = \frac{1}{\tau} \cdot \left[\sum_{k=0}^{i}\int_\Omega x \cdot P(x,k)dx\right] \quad (9)$$

$$\delta_{i+1}^2 = \frac{1}{\tau} \cdot \left[\sum_{k=0}^{i}\int_\Omega (x-\mu_{i+1})^2 \cdot P(x,k)dx\right] \quad (10)$$

A test problem was constructed according to Eq. 4 with the global optimum created by a Gaussian (-0.5, 0.05) as well as a local optimum created by a Gaussian (0.5, 0.2), which was scaled up by a factor of 3.5 to make it a competitive optimum (i.e., a large basin as well as 87.5% of the fitness value of the global optimum).

The initial Gaussian model was set as μ_0=0 and δ_0=0.5. Since it was centred in the middle of the space between the two optima, it is likely that, at least at the beginning of evolution, quite a large portion of selected individuals might come from the peak corresponding to the local optimum, due to its significantly larger basin size. From this point of view, this problem is deceptive and it is interesting to see how the EDA can handle this difficulty with the two different selection schemes.

The dynamics of the EDA with the ($\mu+\lambda$) scheme is shown in Figures 13&14 (with τ=0.3). Compared to the monotonic behaviour observed in the unimodal case, the behaviour of the EDA is much more complicated on multimodal problems. For example, in the first five generations, the standard deviation values were shrinking and, in the meantime, the mean of the Gaussian was moving towards the local optimum.

This indicates that the EDA was being misled by the large portion of selected individuals from the local optimum. However, after that point, the EDA moved back and consistently converged towards the global optimum, correcting its previous trajectory.
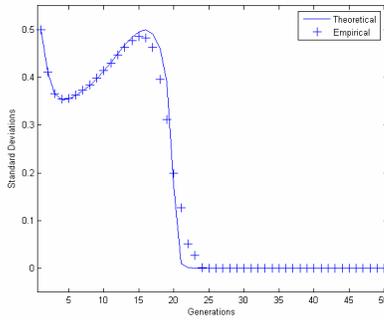
Fig. 13. The trajectory of the standard deviation parameter of the Gaussian model on the multimodal test problem with the (μ+λ) scheme.
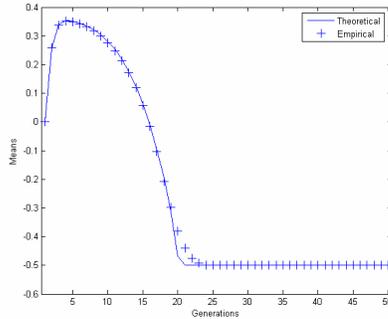

Fig. 14. The trajectory of the mean parameter of the Gaussian model on the multimodal test problem with the (μ+λ) scheme.
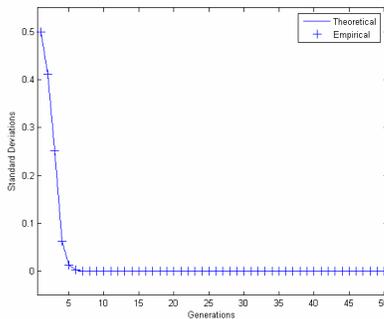

Fig. 15. The trajectory of the standard deviation parameter of the Gaussian model on the multimodal problem with the (μ, λ) scheme.
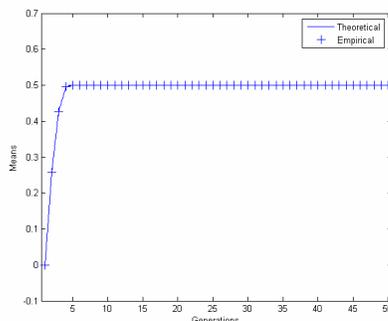

Fig. 16. The trajectory of the mean parameter of the Gaussian model on the multimodal problem with the (μ, λ) scheme.

It should be pointed out that the EDA with the (μ+λ) scheme will never get stuck at the local optimum with an infinitely large population, provided that every point in the search space has the chance to be sampled based on the statistical model in use, even with small probabilities. This is because the (μ+λ) scheme will always memorize good individuals that have been found. Individuals from the basin corresponding to the global optimum (i.e., it is guaranteed that some of them would exist in the initial population) will not only prevent the EDA from converging to the local optimum but also gradually update the model towards the global optimum.

By contrast, there could be a quite different story with the (μ, λ) scheme in which all old individuals are discarded. As long as individuals from the local optimum start dominating the population, the model will be updated towards the local optimum and the chance of getting new individuals from the global optimum will decrease, further reducing its influence in the population. In Figure 15, the standard deviation quickly dropped to nearly zero after five generations while, as shown in Figure 16, the mean got stuck at 0.5 where the local optimum was located. Again, in both cases, empirical results (i.e., population size = 10,000 and 25 independent trials) provide a good match against the theoretical analysis.

V. CONCLUSIONS

In this paper, a mathematical modelling technique has been proposed under the assumption of infinite population size, to investigate the dynamics of a continuous EDA with truncation selection and Gaussian models. The importance of this technique is that it provides a principled method to analyse the behaviour of the EDA under the influence of different algorithm/ problem factors.

In the case studies, this technique is applied to 1D unimodal and multimodal problems and some interesting properties of the EDA have been found some of which are unknown in previous research.

For 1D unimodal problems with the (μ, λ) scheme:

- The dynamics of the EDA is dependent only on the general shape of the test function, which is also true for the (μ+λ) scheme.

- When initialized far from the global optimum, the EDA has a tendency to converge prematurely.

- Given a certain selection pressure, there is a unique value for the amplification parameter γ that could help the EDA achieve desirable performance.

For 1D multimodal problems:

- The EDA could get stuck with the (μ, λ) scheme.

- The EDA will never get stuck with the (μ+λ) scheme, which is also true for unimodal problems for obvious reasons.

- The above conclusions also hold for high dimensional problems.

Furthermore, we expect that there are useful connections between the theoretical analysis of EDAs and the existing work on Evolution Strategies [3], which may help these two areas benefit from each other.

Regarding to higher dimensional problems, some preliminary work has been successfully conducted on 2D unimodal problems with dependences among parameters. However, a major challenge may come from specifying the boundaries of selected individuals and calculating multiple integrals over several variables. As a result, it seems that some further assumptions may be required to keep the complexity of such theoretical analysis at a reasonable level.

Note that modelling the behaviour of EDAs in high dimensional spaces is generally difficult. Grahl et. al [6] claim that an n-dimensional monotonic problem can be decomposed into n 1D problems and the mathematical modelling process only needs to be conducted on each of them independently. However, as will be shown in the next, this is not true in general.

Suppose that the fitness function is f(x,y)=x+y, which is to be maximized. In this case, the selection threshold is represented by a line y=-x+b with the value of b depending on the selection pressure τ.
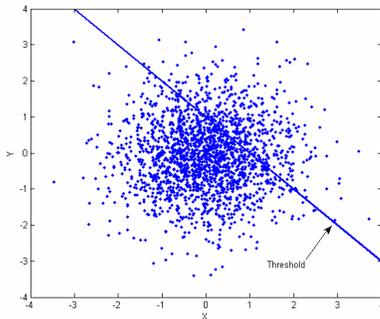

Fig. 17. A demonstration of the distribution of individuals and the selection threshold on f(x,y)=x+y with truncation selection.

A demo is given in Fig. 17, which shows the distribution of individuals sampled from a bivariate Gaussian distribution as well as the selection threshold. Note that only individuals above the threshold (i.e., x+y>b) will be selected to build the new Gaussian model. Through intuitive observation, it is easy to see that those selected individuals tend to indicate some correlation between x and y.

An experiment was conducted with 10,000 individuals sampled from a bivariate Gaussian distribution with μ=0 and δ=1.0 in each dimension. For τ=0.3, the covariance matrix estimated from those 3000 selected individuals was:

$$\begin{bmatrix} 0.6331 & -0.3692 \\ -0.3692 & 0.6350 \end{bmatrix}$$

It is clear that there was some strong empirical correlation between the two variables of a monotonic problem. Note that the variances of the two variables would be around 0.2773 if the original function is treated as two 1D functions and the modelling technique is applied on each of them.

Fig. 18 gives another example where the Gaussian model is sitting right on top of the minimum of a unimodal function f(x,y)=x^2+y^2. In this case, although there is no dependence structure in the selected individuals, the selection threshold

represented by a circle x^2+y^2=b still needs to be specified by both variables. As a result, it is again inappropriate to treat the original problem as two independent 1D problems.
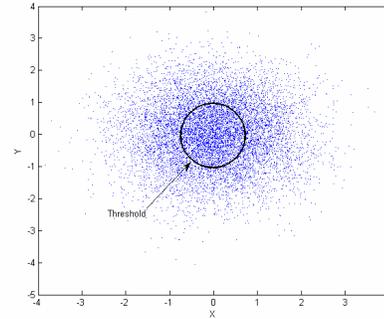

Fig. 18. A demonstration of the distribution of individuals and the selection threshold on f(x,y)=x^2+y^2 with truncation selection.

ACKNOWLEDGMENT

REFERENCES

[1] T. Bäck, D. B. Fogel, and Z. Michalewicz (eds.), *Handbook of Evolutionary Computation*. IOP Publishing Ltd and Oxford University Press, 1997.

[2] A. Berny, "Selection and Reinforcement Learning for Combinatorial Optimization". In Parallel Problem Solving from Nature VI, 2000, pp. 601-610.

[3] H.-G. Beyer, *The theory of evolution strategies*. Springer, 2001.

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

[5] 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*, vol. 31, pp. 313-340, 2000.

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

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

[8] P. Larrañaga and J. A. Lozano (eds.), *Estimation of Distribution Algorithms: A New Tool for Evolutionary Computation*. Kluwer Academic Publishers, 2001.

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

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

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

[12] B. Yuan and M. Gallagher, "Experimental Results for the Special Session on Real-Parameter Optimization at CEC 2005: A Simple, Continuous EDA". In Congress on Evolutionary Computation 2005, 2005, pp. 1792-1799.

[13] B. Yuan and M. Gallagher, "On the Importance of Diversity Maintenance in Estimation of Distribution Algorithms". In Genetic and Evolutionary Computation Conference 2005, 2005, pp. 719-726.

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