

ON THE EXTINCTION OF SUB-POPULATIONS ON MULTIMODAL LANDSCAPES

L. Schönemann, M. Emmerich, M. Preuss

Systems Analysis Group, Computer Science Department,

University of Dortmund, Germany

Keywords: Evolutionary algorithms, multimodal fitness landscape, niching techniques, random genetic drift, gambler's ruin

Abstract Population based evolutionary algorithms (EA) are frequently used to optimize on multimodal functions. A common assumption is that during search several sub-populations might coexist in different attraction regions of the search space. Practical experience and takeover-time considerations suggest that this is not true in general. We therefore analyze the stability of sub-populations within a simplified EA on a two-attractor model, focussing on two extreme cases: (1) Function values of both local minima are exactly the same and (2) function values on the first attractor are always better than on the second. Realistic scenarios for bimodal optimization are assumed to be located in between these two extremes, such that upper and lower bounds for extinction times can be estimated, e.g. by Markov chain analysis and empirical studies. The obtained results provide new insights into the effect of $(\mu \dagger \lambda)$ selection on the stability of sub-populations and the effect of genetic drift. Moreover, the effect of idealized niching on the same scenarios is investigated, leading to an immense increase of the EA's ability to perform concurrent search.

1. Introduction

EA are preferable tools for optimization on multimodal functions. It has often been assumed that the strength of EA stems from the ability to search concurrently in different high performance regions of the search space. Contrary to this, experimental results on multimodal function optimization suggest that EA using the panmictic $(\mu \dagger \lambda)$ selection tend to rapidly concentrate on a single attractor, even if all optima have the same size and function values. It seems impossible to maintain individuals in different regions at the same time, without employing niching techniques.

In this paper we trace back the effect of extinction on neutral landscapes to random genetic drift dynamics, which can be observed in a simplified scenario. Though, dealing with an theoretical issue, the paper provides valuable insights for the practitioner on how to design niching techniques, when searching on multimodal landscapes.

The paper is structured as follows: First, we demonstrate the effect on a simple test-case (2-sphere model) (Sect. 2). Second, we ascribe extinction to random genetic drift dynamics that can be reproduced and analyzed with a simple Markov model which is set up and analyzed in Sect. 3. Based on the theoretical observation we motivate a design principle for niching techniques and demonstrate its benefit on the test case (Sect. 4).

2. Population dynamics on a bimodal test-case

As an example, the extinction of sub-populations has been observed for the minimization of a simple two-sphere problem $f(\mathbf{x}) = \min((x_1 - 1)^2 + x_2^2, (x_1 + 1)^2 + x_2^2)$ (cf. Figure 2.1).

Algorithm 1 describes the $(\mu^\dagger\lambda)$ -EA [2] that will be studied in this paper. Let \mathbb{I} define the individual space. Each individual $a \in \mathbb{I}$ consists of information on its position in the search space and its objective function value. Furthermore, let $P_t \in \mathbb{I}^\mu$, $Q_t \in \mathbb{I}^\lambda$ and $M_t \in \mathbb{I}^\nu$ denote multisets of individuals (or *populations*) with $\nu = \mu + \lambda$ for the $(\mu + \lambda)$ selection and $\nu = \lambda$ for the (μ, λ) selection. P_t will be termed the *parent populations*, while Q_t will be termed the *offspring population* for $t = 0, \dots, t_{\max}$. The

Algorithm 1

```

 $t \leftarrow 0$ 
 $P_t \leftarrow \text{init}()$  /* Initialize population  $P_t \in \mathbb{I}^{\mu*}$  */
while  $t < t_{\max}$  do
   $Q_t \leftarrow \text{gen}(P_t)$  /* Generate  $Q_t \in \mathbb{I}^\lambda$  by variation operators */
   $M_t \leftarrow \begin{cases} Q_t & \text{for } (\mu, \lambda) \text{ selection} \\ Q_t \cup P_t & \text{for } (\mu + \lambda) \text{ selection} \end{cases}$ 
   $P_{t+1} \leftarrow \text{sel}(M_t)$  /* Select  $\mu$  best individuals from  $M_t$  for  $P_{t+1}$  */
   $t \leftarrow t + 1$ 
end while

```

EA starts with initializing the population of parents P_t in the individual space \mathbb{I} . Then the following procedure is repeated while the generation counter does not exceed a user defined maximum t_{\max} : Generate a multiset of λ offspring by means of variation operators (usually recombination and mutation), then select the best μ individuals out of M_t . Here $M_t = Q_t$ in case of (μ, λ) selection and $M_t = Q_t \cup P_t$ in case of

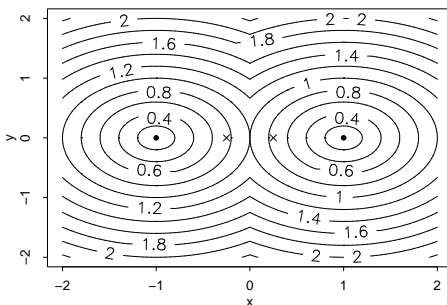


Figure 2.1. Two-sphere model: Crosses mark starting points for the EA.

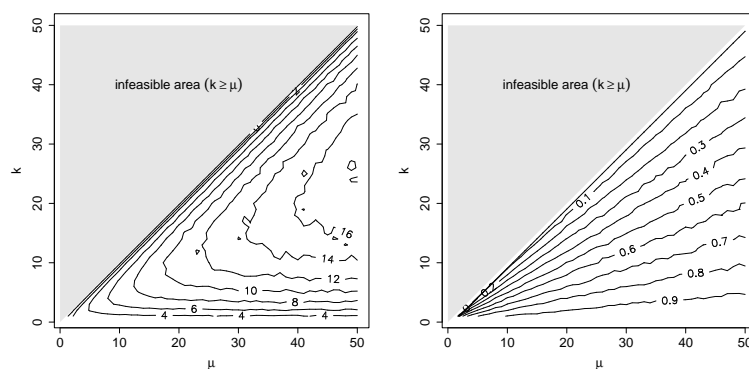


Figure 2.2. Extinction times (left) and probabilities (right) for a $(\mu + 7\mu)$ -ES with Gaussian mutation on the two sphere model (with 2 dimensions) averaged from 5000 runs.

$(\mu + \lambda)$ selection. Finally increase the generation counter and jump to the beginning of the loop.

We make the convention that in case of equal objective function values for M_t , $\text{sel}(M_t)$ draws randomly k individuals out of the M_t individuals, without choosing one of the individuals twice and without preferring offspring individuals in case of $(\mu + \lambda)$ selection.

Figure 2.2 shows the average extinction time and probabilities for the $(\mu + \lambda)$ -EA. It can be observed that the takeover probability of one sub-population grows proportionally with its ratio in the initial population.

3. Markov Model for the Extinction Dynamics

Imagine an objective function (for minimization) with two large local optimal regions with equal or slightly different optimal function values.

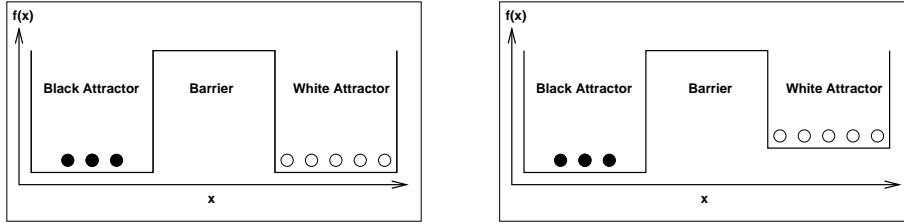


Figure 3.3. Schematic draw of the instantiations of the *two-attractor model*. The left figure describes the case with equal function values in both attractor basins and the right figure describes the case of better function values for the black individuals than that for the white individuals.

In between these plateaus there is a large barrier with extremely high function values (Fig. 3.3), such that it is very improbable that an individual from one area crosses the barrier by a single mutation. This is similar to the case that the optimization has reached the bottom of two equal or similar local optima of a bimodal function with flat bottoms.

In order to simulate the dynamics of the $(\mu \dagger \lambda)$ -EA on such a system, let us define the following rules of the game:

For a population P_t at a time t let $\text{black}(P_t)$ define the number of individuals on the first attractor (we will call them *black individuals*). Accordingly, $\mu - \text{black}(P_t)$ individuals are located on the second attractor (we call them *white individuals*). Furthermore, let us assume that all individuals on an attractor have the same function value. Individuals cannot move from one attractor to another attractor or leave their attractor by means of mutation. Hence, the reduced mutation operator simply results in a copy of the individual.

Assuming an initial population with a specified number of black individuals, we are now interested in the dynamics of the simplified EA for the case that (1) the function value for both plateaus is equal and (2) the function value for the plateau that contains the black individuals is better than the function value on the plateau that contains the white individuals. Markov chain analysis can be a powerful tool for understanding simple models of evolution [3, 5]. Next, we provide the reader with the derived Markov chain model.

3.1 Deriving the Transition Probabilities

Let k denote the number of black individuals in the initial population. Then we are interested in the probability $p_j(k)$ for j black individuals in the subsequent population. This can be obtained by dividing the EA into two steps - the generation of individuals and the selection of P_{t+1} .

A possible generational transition could be described as

$$P_t = \underbrace{\{\bullet, \dots, \bullet\}}_k, \underbrace{\{\circ, \dots, \circ\}}_{\mu-k} \xrightarrow{\text{generate}} Q_t = \underbrace{\{\bullet, \dots, \bullet\}}_l, \underbrace{\{\circ, \dots, \circ\}}_{\lambda-l} \quad (3.1)$$

$$\xrightarrow{\text{replace}} P_{t+1} = \underbrace{\{\bullet, \dots, \bullet\}}_j, \underbrace{\{\circ, \dots, \circ\}}_{\mu-j} \quad (3.2)$$

Then the transition matrix of the whole evolution step reads:

$$\mathbf{P} := (p_{k,j})_{k \in \{0, \dots, \mu\}, j \in \{0, \dots, \mu\}}, \quad (3.3)$$

with

$$p_{k,j} = \sum_{l=0}^{\lambda} p_l^{\text{gen}}(k) \cdot p_j^{\text{sel}}(l, k) . \quad (3.4)$$

Here $p_l^{\text{gen}}(k)$ describes the transition probabilities of the procedure $\text{gen}(P_t)$

$$p_l^{\text{gen}}(k) = \Pr(\text{black}(Q_t) = l | \text{black}(P_t) = k) \quad (3.5)$$

and $p_j^{\text{sel}}(l, k)$ describes the transition probabilities for the procedure $\text{sel}(M_t)$

$$p_j^{\text{sel}}(l, k) = \Pr(\text{black}(P_{t+1}) = j | \text{black}(P_t) = k \wedge \text{black}(Q_t) = l) . \quad (3.6)$$

The transition probabilities $p_l^{\text{gen}}(k)$ for the generate function are the same for all selection schemes studied here:

$$p_l^{\text{gen}}(k) = \left(\frac{k}{\mu}\right)^l \cdot \left(\frac{\mu-k}{\mu}\right)^{\lambda-l} \cdot \binom{\lambda}{l} . \quad (3.7)$$

Table 3.1 shows the transition probabilities that are instantiated for different selection methods and assumptions about the function values on the two attractors.

3.2 Markov Chain Analysis

Now, we can apply Markov chain analysis [7] in order to analyze the dynamics of the system. Recall from probability theory, for $t > 0$ and any given state vector \mathbf{p}_t we can calculate the probability distribution for the resulting subsequent state j by means of $\mathbf{p}_{t+1} = \mathbf{P} \cdot \mathbf{p}_t$. The limit value for \mathbf{p}_t as $t \rightarrow \infty$ can be obtained by means of the fundamental matrix:

The Markov process of the two-attractor model has absorbing boundaries $k = 0$ and $k = \mu$. If one of the absorbing states has been reached the

Table 3.1. Selection probabilities $p_j^{sel}(l, k)$ (I is the indicator function)

selection method	equal function values	black better than white
$(\mu + \lambda)$	$\frac{\binom{k+l}{j} \cdot \binom{\mu+\lambda-k-l}{\mu-j}}{\binom{\mu+\lambda}{\mu}}$	$I(j = \min(\mu, k + l))$
(μ, λ)	$\frac{\binom{l}{j} \cdot \binom{\lambda-l}{\mu-j}}{\binom{\mu+\lambda}{\mu}}$	$I(j = \min(\mu, l))$

system remains stable. The absorption probabilities and mean absorption times correspond to the extinction probabilities and mean extinction times. Both can be derived from the transition matrix \mathbf{P} and the initial state k_0 . First, let us partition the transition matrix as follows:

$$\mathbf{P} = \begin{pmatrix} 1 & 0 & \dots & 0 & \dots & 0 & 0 \\ a_{1,1} & c_{1,1} & \dots & c_{1,j} & \dots & c_{1,\mu-1} & a_{1,2} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{k,1} & c_{k,1} & \dots & c_{k,j} & \dots & c_{k,\mu-1} & a_{k,2} \\ \vdots & \vdots & & \vdots & & \vdots & \vdots \\ a_{\mu-1,1} & c_{\mu-1,1} & \dots & c_{\mu-1,j} & \dots & c_{\mu-1,\mu-1} & a_{\mu-1,2} \\ 0 & 0 & \dots & 0 & \dots & 0 & 1 \end{pmatrix}. \quad (3.8)$$

Now, the fundamental matrix \mathbf{T} of the transition matrix \mathbf{P} reads:

$$\mathbf{T} := (\mathbf{I} - \mathbf{C})^{-1}, \quad (3.9)$$

and from Markov chain theory ([7], Chap. 3) an expression for the extinction of black individuals, i.e. for reaching the absorbing state $k = 0$, can be derived as

$$p_E(k_0) = \sum_{i=1}^{\mu-1} a_{i,1} t_{k_0,i}, \quad k_0 = 1, \dots, \mu - 1 \quad (3.10)$$

It is also known that $t_{i,j}$ of \mathbf{T} equals the mean number of iterations that the system is in state i when starting in state j before absorption takes place. Thus

$$E(k_0, \mu) = \sum_{i=1}^{\mu-1} t_{k_0,i}, \quad k_0 = 1, \dots, \mu - 1 \quad (3.11)$$

is the mean absorption time, or - translated to our model - the average time that two species in the evolutionary system can coexist when working with the generational transition described by \mathbf{P} and starting with k individuals.

4. Analysis of Selection Mechanisms

Now, we can use the Markov chain techniques proposed in the previous section to determine some characteristics of selection mechanisms on the two-attractor model. We start with the case of equal fitness.

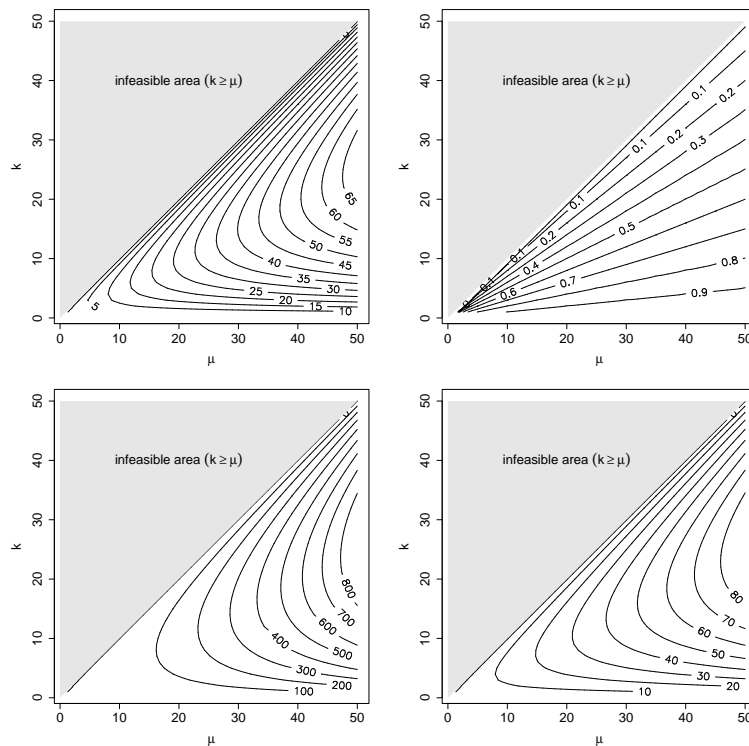


Figure 4.4. Expected extinction times (EET) and probabilities (PE) obtained by Markov theory. Upper left: EET $(\mu + 7\mu)$ -EA, Upper right: PE $(\mu + 7\mu)$ -EA, Lower left: EET $(\mu + 1)$ -EA, Lower right: EET $(\mu + \mu)$ -EA.

Figure 4.4 shows the mean EET and PE for some frequently used EA strategies. The figure reveals that the extinction times increase linearly with a growing μ if λ and k are set as a constant proportion of μ . Note, that the extinction times are measured in generations. In the case of $(\mu + 7\mu)$ selection and $\mu = 40, k = 20$ this means that one population dies

out on average after 53 generations or about 15,000 offspring. Contrary to this, for the $(\mu + 1)$ selection and the same settings for μ and k only 550 offspring are generated until one population dies out.

In addition to the mean extinction time we are interested in the probability of extinction for one population. As a closer look at the underlying data of the upper right diagram in Fig. 4.4 reveals, the extinction probability could be quantified by $\frac{k}{\mu}$. This is an astonishing simple formula and especially independent from λ .

Investigating the model of equal function values gives fundamental insights into the behavior of EA on a bimodal fitness landscape. But it is assumed that a real EA will produce different offspring on a two-attractor landscape. Therefore, we study the two sphere model depicted in Fig. 2.1 as a more realistic case.

For this function, a Markov chain analysis can not be applied any more. Hence, we obtain the results presented in Fig. 2.2 by a monte carlo simulation with a real EA. In order to prevent an acceleration of the extinction process cause by recombination, the EA shall only apply a variation operator that works with the mutation operator only. The mutation operator adds a normal distributed random variate to the object variables. The small value of the mutation step size assures that no individual is produced that jumps from one attractor to the other.

The results show that due to the stochastic mutation the extinction times are smaller than in the former model. This is an expected result because this scenario is lying in between the two models of equal and different function values. In contrast to this, the probabilities of extinction are the same.

Our investigations show that a species can die out quickly even if it has equal function values as the other species. To guarantee the survival of a fitter species and to prevent the extinction of sub-populations located on equally shaped attractors one could use several techniques. One of it is niching [4, 6, 8]. Some of the former experiments were repeated with a simple niching technique: Attraction areas to which individuals belong are identified by some clustering approach (cf. [9]) and we generate the same number of offspring individuals. In contrast to many other niching techniques, the suggested selection process is panmictic (alternative selection schemes are presented in [1, Sect. C2]). It was observed that using this kind of simple niching, the sub-populations were able to co-exist for a very much longer time (Table 4.2), even for the more realistic example of the 2-sphere function. Hence, the results in this paper affirm that for optimization on multimodal landscapes niching is a preferable technique to avoid the loss of information gained by sub-populations [9], even in the presence of equality.

Table 4.2. Illustrative cases for extinction times and probabilities.

Strategy	Model	Niching	k	Simulation Method	E(t)	p_E
(16,112)	boolean	-	1	Markov	2.0	$7 \cdot 10^{-4}$
(16,112)	neutral	-	1	Markov	6.7	0.94
(16,112)	neutral	x	1	Markov	$1.1 \cdot 10^5$	0.5
(16,112)	boolean	x	1	Markov	1.0	0
(4,28)	neutral	x	1	Markov	10.2	0.5
(50+1)	boolean	-	1	Markov	224.0	0
(16,112)	two-sphere	x	1	Experiment	$9.8 \cdot 10^4$	0.48
(16+112)	two-sphere	x	1	Experiment	$> 10^6$	0.52

The results for some selected test cases are shown in Table 4.2. It contains the mean extinction times as well as the extinction probabilities. The cases were chosen because they reflect some frequently observed situations and provide the reader with some borderline cases. The mean extinction time shows the number of reproduction cycles both sub-populations survive. In contrast to this p_E measures the probability that the species with the black individuals die out. Some remarks need to be spend on the results of the (16+112)-ES on the two-sphere model. Due to time limitations the runs were limited to 10^7 generations. Until then, in only 15% of all runs one population died out. On average, a single run lasts longer than 10^6 generations. In the case that one population died out, the black sub-population was eliminated in 52%.

5. Conclusions

By means of this paper a better understanding of the extinction process on multimodal landscapes has been achieved. In detail our investigations give evidence for the following conjectures:

- For $(\mu \dagger 7\mu)$ -EA and the equal fitness model the extinction time grows linearly with μ . Thus, even in the best-case scenario the co-existence of populations will not occur for a long time. This result can be interpreted also in a way that mating restrictions alone will not suffice to prevent sub-populations from extinction. This is because in the studies on the simplified model, the recombination has been ommitted and thus it cannot accelerate the extinction of species.
- For different function values and comma selection it was observed that better individuals survive with probability near 1. Hence, the effect of random genetic drift unlikely biases the direction of evolution, if fitness values are clearly different.

- The extinction time for a $(\mu + 1)$ -EA are long even in the case of equal function values. However, if we regard the number of function evaluations as a criterion for the extinction time, proportions change and the mean extinction time for the $(\mu + 1)$ -EA is significant smaller than that for, e.g., the $(\mu + 7\mu)$ -EA.
- The extinction time is increased substantially by using the suggested niching technique.

In this paper only the bimodal case has been considered. However, we conjecture that similar results can be obtained for landscapes with more than two attractors. Future research will have to clarify this point. Furthermore, the effect of the recombination operator deserves further attention.

Acknowledgments This work was supported by the Deutsche Forschungsgemeinschaft (DFG) as part of the *Collaborative Research Center* “Computational Intelligence” (SFB 531).

References

- [1] Bäck, T., Fogel, D.B., Michalewicz, Z., eds.: Handbook of Evolutionary Computation. Oxford University Press, New York, and Institute of Physics Publishing, Bristol (1997) Release 97/1.
- [2] Beyer, H.-G., Schwefel, H.-P.: Evolution strategies: A comprehensive introduction. *Natural Computing* **1** (2002) 3–52
- [3] Chakraborty, U., Deb, K., Chakraborty, M.: Analysis of selection algorithms: A markov chain approach. *Evolutionary Computation* **2** (1996) 133–167
- [4] Deb, K., Goldberg, D.E.: An investigation of niche and species formation in genetic function optimization. In Schaffer, J.D., ed.: Proc. 3rd Int’l Conf. on Genetic Algorithms ICGA 89, San Mateo, CA, Morgan Kaufmann (1989) 42–50
- [5] Goldberg, D.E., Segrest, P.: Finite markov chain analysis of genetic algorithms. In Grefenstette, J.J., ed.: Proc. 2nd Int’l Conf. on Genetic Algorithms ICGA 87, L. Erlbaum Associates (1987) 1–8
- [6] Horn, J.: Finite markov chain analysis of genetic algorithms with niching. In Forrest, S., ed.: Proc. Fifth Int’l Conf. on Genetic Algorithms ICGA 93, San Mateo, CA, Morgan Kaufmann (1993) 110–117
- [7] Kemeny, J.G., Snell, J.L.: Finite Markov chains. D. Van Nostrand Company, Ltd., London (1969)
- [8] Mahfoud, S.W.: Niching Methods for Genetic Algorithms. PhD thesis, University of Illinois at Urbana Champaign (1995)
- [9] Streichert, F., Stein, G., Ulmer, H., and Zell, A.: A Clustering Based Niching EA for Multimodal Search Spaces. Proceedings of the 6th International Conference on Artificial Evolution, Marseille, France, October 27-30, 2003, p. 169-180