Counteracting Genetic Drift and Disruptive Recombination in $(\mu \overset{+}{,} \lambda)$-EA on Multimodal Fitness Landscapes

Mike Preuss
Systems Analysis Group
Comp. Science Department
University of Dortmund,
Germany
mike.preuss
@uni-dortmund.de

Lutz Schönemann
Systems Analysis Group
Comp. Science Department
University of Dortmund,
Germany
lutz.schoenemann
@cs.uni-dortmund.de

Michael Emmerich
Natural Computing Group
Leiden Institute for Advanced
Computer Science (LIACS),
Niels Bohrweg 1, 2333 CA
Leiden, NL
emmerich@liacs.nl

ABSTRACT

The impact of operator disruption and genetic drift on the extinction of EA subpopulations on multimodal landscapes is estimated by means of idealized two-peak landscape models. To establish upper and lower bounds for extinction times the behavior of an EA that employs $(\mu \overset{+}{,} \lambda)$ selection and recombination mechanisms is studied, assuming disruptive recombination. Markov chain and statistical simulation studies reveal that panmictic selection mechanisms as used in evolution strategies (ES) do not allow for maintaining several populations of similar fitness at the same time. Moreover, when using comma selection, good individuals might easily get lost if forming the minority of a population, an effect seemingly amplified by recombination. Niching techniques are suggested to facilitate coexistence of populations on distant attractors; conducted studies confirm their aptitude.

Categories and Subject Descriptors: G.3: Probabilistic algorithms

General Terms: Algorithms

Keywords: multimodal optimization, evolution strategies, Markov simulation, niching, clustering, genetic drift

1. INTRODUCTION

State-of-the-art evolutionary algorithms (EA) usually rely on recombination and mutation as variation operators. Current results show that on some problems, recombination provably speeds up global search essentially [7]. On the other hand, it is also known that it might disrupt good individuals [4]. In this study, we take a closer look at the effect of recombination on EA behavior for simple multimodal fitness landscapes.

Experimental experience often provides us with disappointing results when tackling multimodal problems. Es-

pecially (μ, λ)-EA, although potentially able to leave local optima, seem vulnerable to losing track of the global optimum, despite having entered its basin of attraction. We will provide evidence that a single individual entering a new attraction basin is lost under comma selection even if it is the best currently detected one. A first naive explanation is that it does not manage to pass on its search space position to the next generation due to lack of appropriate mating partners.

To enable judging the validity of this conjecture, we herein present a simple model for the behavior of a $(\mu \overset{+}{,} \lambda)$-EA on a two-attractor fitness landscape. The derived probabilistic formulas permit Markov chain simulation for a number of cases, including the genetic drift effect. On the border of practical infeasibility, these can be extended by monte carlo simulations. Using both techniques, we tackle questions for the likelihood of the undesired events described above. Note that our approach is related to, but still different from the investigation of takeover times [6] which focuses on selection schemes. Undoubtedly, selection favors good individuals if available. Nevertheless, these have to be generated first – this production aspect is the rationale behind our model.

One may argue that the two-attractor case is too specific and thus our model lacks explanatory power for the general case. However, as recombination in panmictic EA tends to concentrate the whole population to a certain search space area [2], it may be reasonable to expect that different attractors are often discovered by a single individual due to one large mutation.

The paper is structured as follows. First, Sect. 2 relates our approach to previous work on genetic drift. Section 3 describes our model together with the basic EA. We then derive the transition matrices used for Markov simulation in Sect. 4. Results of analyses and simulations are presented in Sect. 5, whereas the following Sect. 6 explains our new approach to hold different local optima. Finally, Sect. 7 summarizes the paper.

2. GENETIC DRIFT RELATED WORK

The term genetic drift is known from population genetics. It describes the effect of a loss of population diversity that occurs due to the stochastic nature of selection in a finite population even in case of neutral fitness [9].

Investigations of this effect on evolutionary algorithms

were carried out by several authors. Asoh and Mühlenbein examine convergence of a population to one genotype. They only take reproduction by uniform crossover into account [1], neglecting mutation and selection. This stands contrary to our investigations, as we also look at selection. For a representation with two alleles per gene they perform a Markov chain analysis to calculate the mean convergence time. For one gene, this time is linear in the number of individuals. For multiple genes it grows logarithmic in their number.

Rogers and Prügel-Bennett investigate the effect of selection schemes on the population fitness variance ([11]). The selection is split into two operators. First, it randomly chooses a parent to generate one offspring by producing an exact copy. In a second step, the new parents are selected out of the set of offspring and possibly the parents. At this stage, the selection method prefers individuals with better fitness values. The higher the selection pressure (the ratio of generated offspring and the number of parents), the less is the probability to select an offspring with worse fitness. But even if the same number of offspring as the number of parents is generated, the stochastic nature of the selection operator provides for a loss in population fitness variance. To analyze this effect, the authors use different selection schemes for solving the ONEMAX problem. They come to the conclusion that 'definitive statements about the performance of different selection schemes are difficult to make'. Further related investigations were conducted by Shapiro and Prügel-Bennett [13] and Prügel-Bennett [10].

3. A TWO-ATTRACTOR $(\mu \overset{+}{,} \lambda)$-EA MODEL

We deliberately restrict our model to a $(\mu \overset{+}{,} \lambda)$-type EA where mating selection is regarded as random choice and only environmental selection exerts pressure in direction of better fitness values. The basic algorithm is modeled after [3], later also taking the assignment of individuals to basins of attraction into account. Note that we do not apply clustering at this stage to allow for use of arbitrary test problems. This is deferred to Sect. 6. Instead, we first tackle the second problem: Given that attractor identification is possible, how do subpopulations evolve under different (reproduction and selection) operators?

At first we assign a color tag to each individual, either black if located at the first attractor, or white if located at the second, or gray for all others. Further on, we refer to subpopulations as multisets of individuals sharing the same color.

The color characteristic is described with the following notation: A population

$$P = (\underbrace{\bullet, \ldots, \bullet}_{n_1}, \underbrace{\bullet, \ldots, \bullet}_{n_2}, \underbrace{\circ, \ldots, \circ}_{n_3}) \tag{1}$$

is denoted with $P = (n_1 \bullet n_2 \circ n_3)$. Moreover, for a population P let $\bullet(P)$ denote the number of black individuals, $\bullet(P)$ the number of gray individuals and $\circ(P)$ the number of white individuals.

We now detail the model EA according to the pseudocode given in Alg. 1. Let $\mathbb{I}$ define the individual space. An individual $a \in \mathbb{I}$ consists of its search space position, the associated fitness and a color tag $c_a \in \{\bullet, \bullet, \circ\}$. 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. P_t is termed *parent population*, while Q_t denotes the *offspring population* for $t = 0, \ldots, t_{\max}$.

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

The EA starts with initializing generation counter t and parent population P_t, the latter is placed in the individual space $\mathbb{I}$. Then, as long as the generation counter does not exceed a defined maximum $t_{\max}$, the following loop is repeated: Generate a multiset of λ offspring by means of variation operators (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 $M_t = Q_t \cup P_t$ for $(\mu + \lambda)$ selection. Finally, increase the generation counter and jump to the beginning of the loop.

We use two versions of the generate procedure $gen(P_t)$ with different mating selection schemes. *Standard recombination* randomly chooses two individuals from the whole population, whereas *niching recombination* only mates individuals from the same attractor (black or white). Furthermore, the latter distributes the number of produced offspring evenly between subpopulations. Concerning color/attractor inheritance, our model utilizes the following rules:

- Mating of any two individuals on the same attractor will result in offspring preserving the parents color.

- Offspring generated by recombination between individuals on different attractors is not placed on any attractor and thus gets a gray color tag. Its fitness is worse than that on one of the attractors.

- Recombination with a gray individual involved produces a gray individual.

Note that for our model, the mutation operators influence is neglected insofar as we assume that mutation never leads from one to another attractor.

Concerning selection, we make the convention that in case of equal objective function values for M_t, $sel(M_t)$ randomly draws k out of the M_t individuals. Note that due to their low fitness gray individuals only get accepted for the next generation if $\bullet(Q) + \circ(Q) < \mu$ for (μ, λ)-selection. That is, the number of available black or white individuals is insufficient for filling up the next parent population. For $(\mu + \lambda)$-selection, this event is ruled out if $\bullet(P_0) + \circ(P_0) = \mu$.

4. MARKOV CHAIN MODEL

We are now interested in probabilistic transitions of populations caused by the selection mechanisms in order to perform Markov chain simulations [8]. In particular, we want to compute extinction probabilities and times for start conditions with a number k of black individuals representing a subpopulation on a newly detected attractor. All selection mechanisms investigated here have the Markov property, i.e.
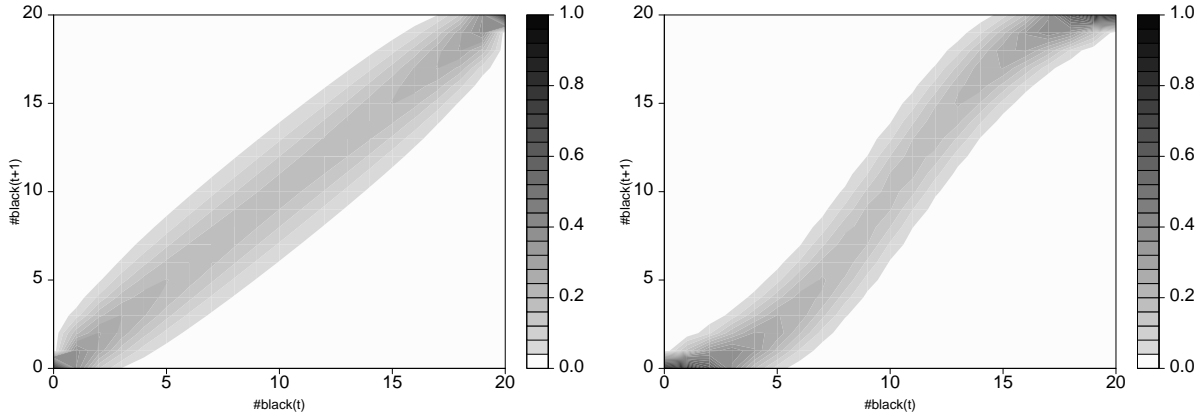
Figure 1: Visualization of transition matrices for an ES with $\mu = 20$, $\lambda = 140$ ($(\mu + \lambda)$ and (μ, λ) variants appear to be structurally identical) without (left) and with recombination (right) on neutral fitness. Recombination obviously reduces survival probabilities for the smaller subpopulation.

the probabilistic distribution of individuals in P_t can be derived from the distribution of individuals in P_{t-1}.

Let k denote the number of black individuals in the initial population. Then we are interested in the probability $p_{j,n}(k,i)$ for j black individuals and n gray individuals given k black individuals and i gray individuals in the initial population. Let

$$P_t = (k \bullet i \circ \mu - k - i), Q_t = (l \bullet r \circ \lambda - l - r), P_{t+1} = (j \bullet n \circ \mu - j - n) \quad (2)$$

describe a possible generational transition. We calculate:

$$p_{j,n}(k,i) = \sum_{l=0}^{\lambda} \sum_{r=0}^{\lambda} p_{l,r}^{gen}(k,i) p_{j,n}^{sel}(k,i,l,r) \quad (3)$$

with

$$p_{l,r}^{gen}(k,i) := \Pr(Q_t = (l \bullet r \circ \lambda - r - l) | P_t = (k \bullet i \circ \mu - k - i)) \,,$$

and

$$p_{j,n}^{sel}(k,i,l,r) := \quad\quad\quad (4)$$

$$\Pr(P_{t+1} = (j \bullet n \circ \mu - j - n) |$$

$$Q_t = (l \bullet r \circ \lambda - r - l) \wedge P_t = (k \bullet i \circ \mu - k - i)) \,.$$

Generate Transition Matrix

Regardless of selection operator and fitness distribution the probability for $p_{l,r}^{gen}(k,i)$ is always the same and can be described by a multinomial distribution:

$$p_{l,r}^{gen}(k,i) = A^l \cdot B^r \cdot C^{\lambda-l-r} \cdot D \quad (5)$$

with

$$A := \left(\frac{k^2}{\mu^2} \right)$$

$$B := \left(\frac{\mu^2 - (\mu - k - i)^2 - k^2}{\mu^2} \right)$$

$$C := \left(\frac{(\mu - k - i)^2}{\mu^2} \right)$$

$$D := \frac{\lambda!}{l! r! (\lambda - l - r)!}.$$

Fractions A, B, and C in this term are the probabilities of generating a black, gray or white individual from P_t within a single recombination. Assuming, the individuals are sequentially generated, fraction D describes the number of possible sequences of length λ that contain l black and r gray individuals (multinomial distribution) and $A^l \cdot B^r \cdot C^{\lambda-l-r}$ denotes the probability for each of these sequences to be generated. It is assumed that the two parents are drawn randomly from the population with uniform probability and that selecting the same individual twice is allowed.

Selection Transition Matrix: Neutral Fitness

If we assume equal fitness on both attractors, we obtain the transition probabilities:

$$p_{j,n}^{sel}(k,i,l,r) = \frac{\binom{q}{j} \cdot \binom{m-q-s}{\mu-j-n}}{\binom{m}{\mu}} \cdot I(n = \max(0, \mu-m+s)) \quad (6)$$

Here, $m = \lambda + \mu$, $q := k + l$, $s := i + r$ in case of a $(\mu + \lambda)$ selection and $m := \lambda$, $q := l$, $s := r$ in case of a (μ, λ) selection. In other words, the merged population writes $M_t = (q \bullet s \circ m - q - s)$. I is the indicator function which equals 1, if the condition is true, and 0 otherwise. Moreover, we use the extended definition of binomial coefficients. The idea behind equation 6 is that — following the Laplacian approach — we relate the number of possibilities to select a population with n gray individuals out of M_t (the denominator of 6) to the number of possibilities of choosing j individuals out of the black partition, while choosing $\mu - j - n$ individuals out of the white partition (in the numerator of 6).

Selection Transition Matrix: Boolean Fitness

If we assume that individuals on the black attractor have better fitness than the ones on the white attractor, the selection step transition gets deterministic:

$$p_{j,n}^{sel}(k,i,l,r) = I(j = \min(\mu, q) \wedge n = \max(0, \mu-m+s)) \quad (7)$$

Here, the numbers m, q, s are defined as in 6, i.e. $M_t = (q \bullet s \circ m - q - s)$.

Mating Restrictions Transition Matrix

If we apply mating restrictions in a way that only individuals located on the same attractor (niche) can mate, the equations of the Markov model simplify, because gray individuals need no longer be considered.

Let $k = \bullet(P_t)$, $l = \bullet(Q_t)$, and $j = \bullet(P_{t+1})$. Then we get the general transition matrix:

$$p_j(k) = \sum_{l=0}^{\lambda} p_l^{gen}(k) p_j^{sel}(l, k). \tag{8}$$

For the generation of new individuals we get the equation:

$$p_l^{gen} = \left(\frac{k}{\mu}\right)^l \cdot \left(\frac{\mu - k}{\mu}\right)^{\lambda - l} \cdot \binom{\lambda}{l}. \tag{9}$$

In order to obtain the transition probability $p_j^{sel}(l, k)$, we have to distinguish two cases - equal fitness values and distinct fitness values on both attractors. In the first case we get:

$$p_j^{sel}(l, k) = \frac{\binom{q}{j} \cdot \binom{m-q}{\mu-j}}{\binom{m}{\mu}}, \tag{10}$$

with $m = \lambda, q = l$ in case of comma selection and $m = \lambda + \mu, q = k + l$ in case of plus selection. For boolean selection the equation simplifies to: $p_j^{sel}(l, k) = I(j = min(\mu, q))$. For a detailed derivation of these equations we refer to [12].

Visualization of Transition Matrices

Figure 1 shows a graphical representation of the transition matrices of $(\mu \overset{+}{,} \lambda)$-EA for the neutral fitness case. According to further generated graphs, selection mechanism and population size only change concrete probabilities but not the overall shape. The only important difference lies in application or non-application of recombination which drastically reduces the chances of small subpopulations to survive. The state with equal numbers of individuals in both subpopulations becomes even more unstable.

5. SIMULATION AND ANALYSIS

Our first tests have shown that without counteractions, two subpopulations never survive for long if global environmental $(\mu \overset{+}{,} \lambda)$-selection is applied. All further simulations and experiments are thus targetted to answering the following questions:

- What is the probability for a new subpopulation to become extinct?

- How long do both attractors remain in focus of the population?

With our model and the formulas derived in the last sections, we now have a number of tools available for generating numerical results needed to answer these two questions. Markov simulations were carried out for smaller populations and thus manageable transition matrix sizes. Additionally, we used simple Monte Carlo simulations to approximate cases with larger populations where Markov simulations get difficult to handle (see Fig. 2). Furthermore, model results were validated by comparison to the averaged output of a real EA, started under similar conditions.
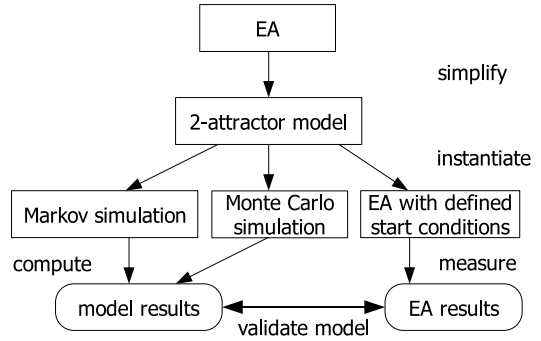

Figure 2: The two-attractor model is derived from a real EA by simplification: Only mechanisms regarded as dominant are included. Defining exact start conditions, it can be validated by comparing model output to EA results.

Another very elegant way to directly compute extinction times and probabilities for the black subpopulation is to analyze the transition matrix of the markov process. In case of mating restrictions no gray individuals need to be considered and the transition matrix is simply given by the $\mu + 1 \times \mu + 1$ matrix $p_j(k)$. The markov process is a process with two absorbing boundaries, either the extinction of black individuals $k = 0$ or the extinction of white individuals $k = \mu$. Thus, it suffices to analyse the properties of the fundamental matrix in order to directly compute the mean extinction time and extinction probability of the black subpopulation [12].

Under plus selection, gray individuals can not enter the population P_{t+1} provided that no gray individuals existed in P_t. Thus, the $\mu + 1 \times \mu + 1$ transition matrix of size $p_{l,0}(k, 0)$ can be applied to precisely compute mean extinction times and probabilities of the black subpopulation, in a similar manner as suggested for the aforementioned case with mating restrictions.

In case of comma selection, gray individuals appear in the parent population with positive probability, even if we start from black and white individuals only. Thus the number of states for the transition matrix increases from $\mu + 1$ to $(\mu + 1) \times (\mu + 2)/2$. As the transition matrix has size (number of states)^2, doubling μ leads to a $\approx 2^4 = 16$ times larger matrix, resulting in high computational cost.

Real EA Experimental Setup

As real optimization problem examples, we consider symmetric and asymmetric two spheres problems of type $f(\mathbf{x}) = \min(\sum(x_i - a_i)^2, \sum(x_i - b_i)^2 - c)$, with $\mathbf{a}$ and $\mathbf{b}$ giving sphere centers and $c = 0$ for the symmetric case. Here, we utilize discrete (dominant) recombination as common in evolution strategies, and mutation with small, constant mutation strength. Two subpopulations of size k and $\mu - k$ are initialized slightly inside their particular basin of attraction. Note that at this stage, we provide the EA with attractor information for each individual by initialization — this is equivalent to an ideal initial clustering.

Extinction Probabilities and Times

Concerning extinction probabilities, Fig. 3 documents the results of our Markov simulations. For the case of equal fitness values, these grasp the behavior of a real EA well if
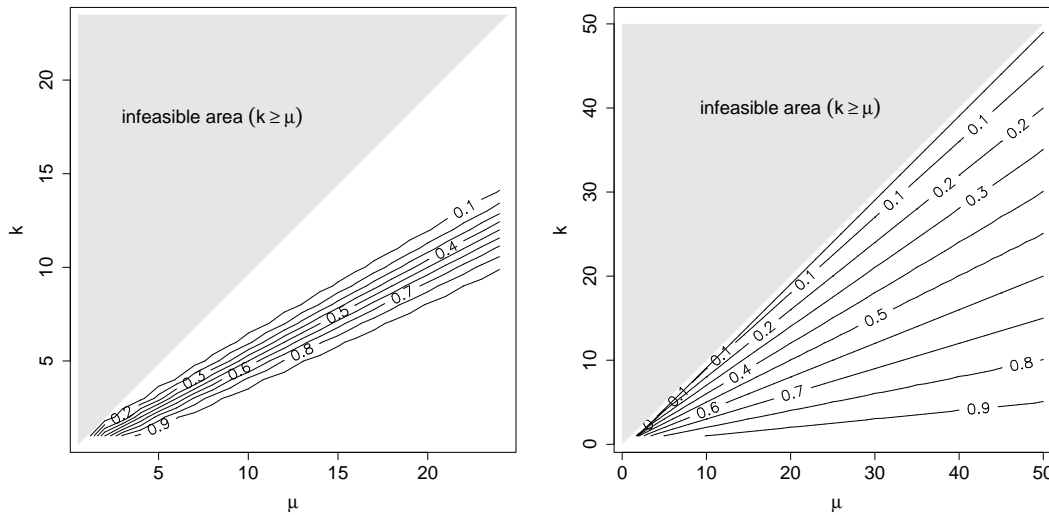
Figure 3: Extinction probabilities computed by means of Markov chain simulations, $(\mu + 7\mu)$-EA, with (left, scaling differs) and without (right) recombination on neutral fitness.

strategy	k	neutral	reco	p_{ext}	niching	p_{ext}
(16,112)	1	x	1.1	1.0	$>10^5$	0.5
(16,112)	1		1.7	<u>0.96</u>	1.0	0
(16+112)	1	x	1.2	1.0	$>10^5$	0.5
(16+112)	1		5.7	0	1.0	0
(16,112)	8	x	4.8	0.5	$>10^5$	0.5
(16,112)	8		1.0	0	1.0	0
(16+112)	8	x	5.5	0.5	$>10^5$	0.5
(16+112)	8		1.0	0	1.0	0

Table 1: Extinction times/probabilities p_{ext} computed by Markov simulation. Reco: default recombination, niching: niching with mutation only or with recombination (equivalent in the model). For boolean fitness p_{ext} of the better subpopulation is recorded.

fitness values in the latter are very similar. Under our assumptions the simulations revealed that selection pressure or method (plus or comma) have little to no influence on the extinction probability characteristics. Recombination drives low and high probability lines towards the center, denoting that a small subpopulation is displaced with increased probability by its larger counterpart. In particular, this applies to groups of individuals resulting from detection of a new attractor. In a real EA with comma selection, it holds even for boolean fitness (one subpopulation is better).

Figure 4 illustrates the effects of recombination on the extinction times. Compared to the case without recombination, extinction times shrink to less than half the amount of the former. We infer that recombination provides for a clear distinction of subpopulations. The smaller subpopulation is afflicted with higher extinction probabilities as well as lower extinction times. Table 1 gives some numerical examples for Markov simulation results of borderline cases.

Both Subpopulations Become Extinct

It can be expected that under certain conditions, both original (black and white) subpopulations may vanish and be replaced with gray individuals. However, Markov chain analysis for small μ revealed that for $\lambda \geq 4\mu$ this probability is almost negligible. For neutral function values, μ even and $\mu/2$ black and $\mu/2$ white start individuals, the recombination operator would produce on average $\lambda \cdot (1/2)^2$ black and white offspring each, but $2 \cdot \lambda \cdot (1/2)^2$ gray offspring. Therefore, to keep both subpopulations the selection pressure must hold $\lambda/\mu > 4$. Otherwise, at least one subpopulation would rapidly die out.

Observations Concerning Gray Individuals

Early EA experiments revealed that our models assumptions about producing gray individuals and thus the expected disruptiveness of recombination may be too strong, especially in case of discrete (dominant) object parameter recombination. As this reproduction operator works like uniform crossover, the parent contributing an object parameter to the new individual is randomly chosen for each search space dimension. Producing an individual that inherits genetic material only from one parent thus has probability $1/2^{n-1}$ if n is the number of object parameters differing in both parents. For high dimensional search spaces with uneven local optima distribution, the gray individual fitness assumption seems realistic, whereas for very low dimensional problems, many gray individuals are in truth black or white.

A more general reason against the low fitness assumption of gray individuals is that recombination between subpopulations may lead to new local optima. However, the probability for these events entirely depends on the nature of the fitness function. We expect that local optima detection by recombination is likely for regularly structured optimization problems like e.g. the Rastrigin function. If local optima are distributed in an asymmetric fashion, this will happen far less often and thus our model is much more appropriate.

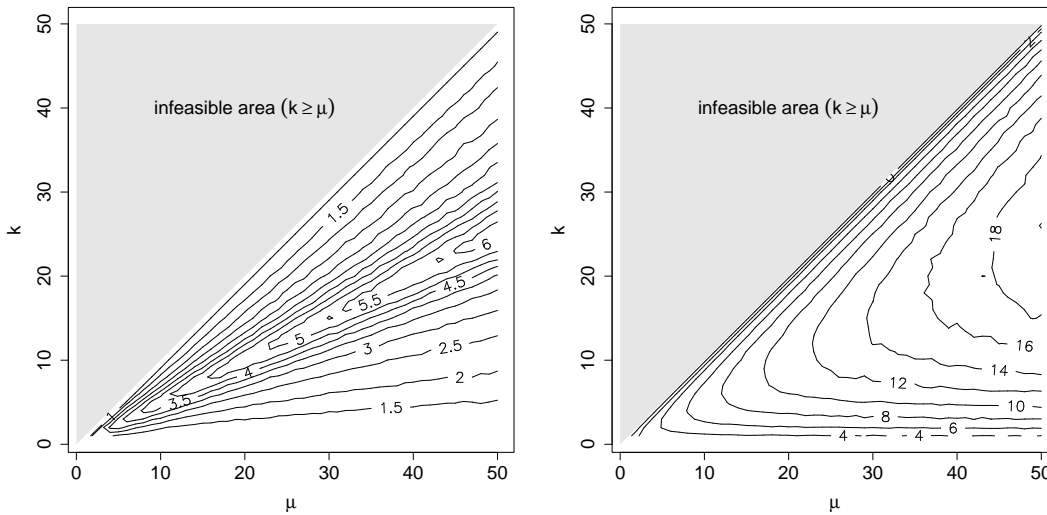Another unregarded possibility to produce gray individuals in a real EA is by mutation. In fact, it may happen

Figure 4: Extinction times as derived from results of a $(\mu, 7\mu)$-EA on a problem with two symmetric spheres in 10 dimensions, with (left) and without (right) recombination.

that by a huge lucky step in search space another attractor is found. If by then the remaining population has converged to one attractor, our model can be applied to this situation by marking this new individual as black and setting $k = 1$. Nevertheless, we expect that except for the first generations, such an event happens very rarely if we additionally take into account that the fitness value on the newly found attractor already has to be good enough to be selected for the next generation to reproduce.

6. REPRODUCTIVE NICHING

In contrast to other approaches that concentrate on selection mechanisms [5], niching and thus preservation of subpopulations may also be carried out by modifying reproduction. Considered we assume ideal knowledge about individual–attractor assignment, we could separately generate the same number of offspring from white and black subpopulations by combining only individuals located in the same attractor. We thereby avoid producing gray individuals. This is related to the tagging-based mating restriction method suggested in [14], but differs insofar as we explicitly require the same amount of offspring for each subpopulation. Assuming an even population size λ, for the transition probabilities we get the equations:

$$p_{l,r}^{gen}(k,i) = \begin{cases} I(l = \lambda/2) & \text{if } (k \neq 0) \wedge r = 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

For the case of equal fitness of individuals on both attractors we get

$$p_{j,n}^{sel}(k,i,l,r) = \frac{\binom{q}{j} \cdot \binom{m-q}{\mu-j}}{\binom{m}{\mu}} \quad , \quad (12)$$

setting $q = k + l, m = \mu + \lambda$ in case of plus selection and $q = l, m = \lambda$ in case of comma selection.

Accordingly, for the case that individuals on the black attractor have better fitness than individuals on the white attractor, we get

$$p_{j,n}^{sel}(k,i,l,r) = I(j = \min(\mu,q) \wedge n = \max(0, \mu-m)) \quad , \quad (13)$$

using the same settings for m and q as in equation 12. Note, that in case of comma selection no species can become extinct if $\lambda < 2\mu$. Hence, no absorbing state can be reached from any state with $0 < k < \mu$.

Figure 5 depicts how reproductive niching works for low population sizes μ under comma selection, obtained by running real EAs with correct initial clustering up to reproduction cycle 1000. Beyond a certain minimal population size that depends on selection pressure λ/μ, extinction times rapidly grow, standing for long periods of coexistence.
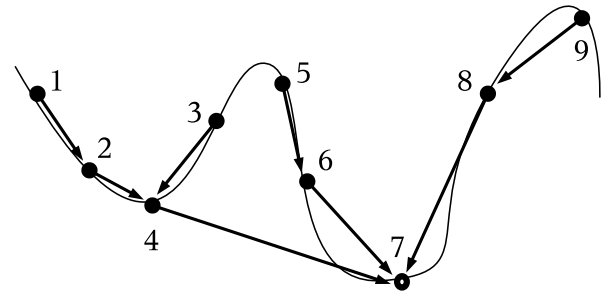
A New Clustering Algorithm

Figure 6: Schema drawing of our clustering algorithm, every individual connects to the nearest better neighbor and connections beyond a threshold length (here: 4 to 7) are interpreted as linking different clusters.

By now we have collected strong evidence in favor of a niching enhanced $(\mu \dotplus \lambda)$-EA when compared with a traditional variant. But to enable niching on unknown multimodal problems, a clustering algorithm is needed that identifies niches during the optimization run. This idea is not entirely new, as the work of Ursem [16] and Streichert [15] documents. Having in mind the difficulties arising from the niche identification methods of these two approaches — increased use of evaluations in the former and difficult para-
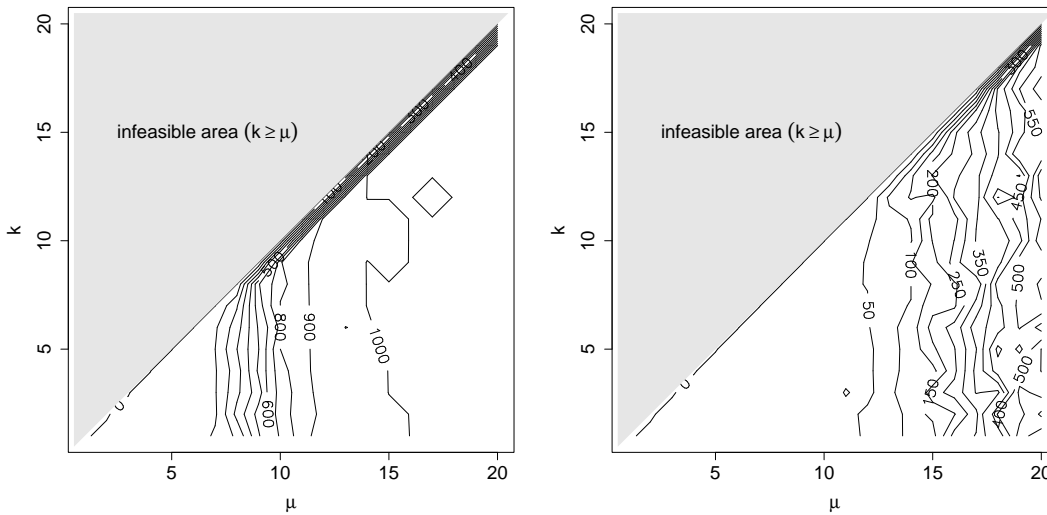
Figure 5: Extinction times with niching, real $(\mu, 2\mu)$ (left) and $(\mu, 7\mu)$-EA (right), on 10 dimensional two symmetric spheres problem. 1000 **is optimal.**

metrization in the latter one — we decided to design a new clustering method that makes use of search space and fitness information and does not necessitate much parameter experimentation.

Our clustering method is layed out in high-level pseudocode in Alg. 2. It rests upon the assumption that the best yet found search points on different attractors are much further away from each other than all individuals are from their nearest better neighbor on average. Thus, it first creates a minimum spanning tree out of the current population (see Fig. 6). The longest edges are then cut, each time splitting a connected subgraph into two. We identify these by means of a heuristic rule that uses the new parameter ϕ. Throughout the following experiments, ϕ is always 2, which seems to be a feasible default value. Note that our algorithm neither uses minimal linkage distances, nor a preset number of required clusters.

Algorithm 2 Nearest-better Clustering

compute all individuals mutual distances in search space;
create an empty graph with num(individuals) nodes;
for all individuals **do**
    find nearest individual that is better; create edge to it;
end for/* result is minimum spanning tree */
delete edges of length $> \phi \cdot mean$(lengths of all edges);
find connected subgraphs;

2-Sphere Experiments

As a first test, we performed experiments with the clustering-based EA on the aforementioned symmetrical and asymmetrical 2-sphere problem under the same initial conditions as before (Table 2). Here, the clustering problem is rather trivial and the obtained results are thus comparable to the ones generated by Markov simulations. Interestingly, for the asymmetrical case and comma-selection, single-individual subpopulations on the better attractor always die out with high empirical probability ($p_{ext} > 0.9$) if default recombination is applied. For the symmetrical case, niching evenly

strategy	k	mut	reco	r-mat	niching	nich-reco
(16,112)	1	3.8	1.1	2.8	$1.4 \cdot 10^4$	$0.6 \cdot 10^4$
(16+112)	1	4.1	1.1	2.6	$>5 \cdot 10^4$	$2.2 \cdot 10^4$
(16,112)	8	9.7	4.5	7.7	$1.4 \cdot 10^4$	$0.7 \cdot 10^4$
(16+112)	8	9.6	4.6	7.5	$>5 \cdot 10^4$	$1.9 \cdot 10^4$
(20,80)	1	4.0	1.1	3.0	$>8 \cdot 10^4$	$>8 \cdot 10^4$
(20+80)	1	3.8	1.1	3.0	$>8 \cdot 10^4$	$>8 \cdot 10^4$
(20,80)	10	12.5	4.9	10.0	$>8 \cdot 10^4$	$>9 \cdot 10^4$
(20+80)	10	11.0	5.2	9.1	$>8 \cdot 10^4$	$>8 \cdot 10^4$
(50+50)	1	4.8	1.6	3.7	$>10^5$	10^5
(50+50)	25	25.0	11.0	21.8	$>10^5$	10^5

Table 2: Extinction times on a symmetric 2-sphere problem, mean values of 500 replicates, with a maximum of 10^5 generations. Mut: mutation only, reco: default recombination, r-mat: restricted mating, niching: mutation only with niching, nich-reco: recombination with niching.

divides the chances for either subpopulation to go extinct, regardless of its initial size. As reproductive niching enforces creation of the same number of offspring individuals for every attractor, small subpopulations are granted very high reproduction rates, thereby intensifying exploitation of their basins of attraction. During one optimization run, these proportions can shift several times (*swing effect*). It is our hope that this mechanism also enables optimizing multimodal problems with similarly as opposed to equally valued attractors.

N-Peaks Experiments

In order to verify reproductive niching as well as our clustering algorithm on a more demanding task, we ran the same EA on randomly generated 5-peaks problems of quadratic shape in 5 dimensions. All peaks are of equal height so that this setting is related to the neutral fitness case. Note that we are not interested in the performance of the EA in terms of speed, but in its ability to keep track of as many optima

as possible for a long time. Fig. 7 shows that with a larger population, reproductive niching with recombination is particularly successful by keeping 4 peaks covered on average. Around $\frac{1}{3}$ of its runs ended with 3,4, and 5 peaks, respectively. It seems that increasing the population size to at least ≈ 10 individuals per peak is necessary to support the clustering. At the same time, chances raise to put at least one indivual onto every peak during randomized initialization.
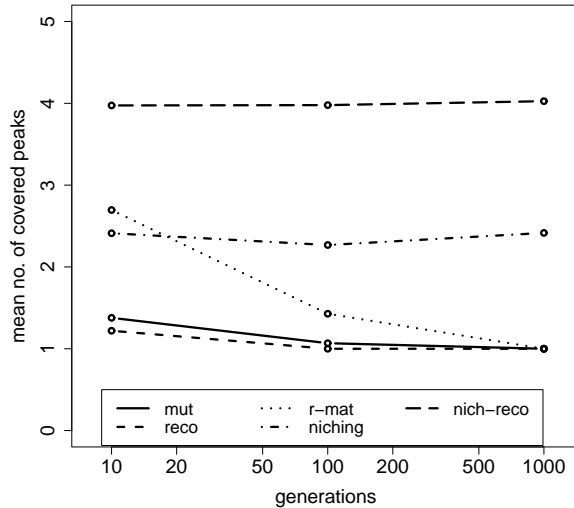

Figure 7: Peak coverage of (100,500)-ES with clustering on 5 dimensional random 5-peaks problem, under different reproduction schemes (see Table 2). Note: lines only added to enhance visibility, values sampled by 500 independent runs of 10, 100, and 1000 generations length.

7. CONCLUSION

A two attractor model has been developed and validated that allows for the analysis of subpopulation coexistence in a $(\mu \overset{+}{,} \lambda)$-EA on bimodal fitness functions with and without recombination. By means of this simple model we can explain and quantify the frequently observed effect that the use of recombination in $(\mu\overset{+}{,}\lambda)$-EA accelerates the extinction of subpopulations, and the related observation that new individuals that are placed on a distant attractor might easily get lost. The obtained results show, that this also happens when fitness values are equal on both local optima, which is some kind of extreme assumption with regard to the difficulty of maintaining coexisting subpopulations.

Omitting the recombination operator or applying mating restrictions (which is similar in the model) increases the stability of subpopulations. Furthermore, it gets far less likely that individuals with better fitness but located on a distant attractor, die out in comma selection. But even for this scenario coexistence of populations is unlikely for a longer time even under neutral fitness. Further niching techniques are needed in order to allow populations with similar fitness to coexist. Here, the proposed reproductive niching proved to be very effective.

Future work will be needed, to extend the model for more complex landscapes with multiple peaks. On such land-scapes diversity raising effects of recombination, e.g. the detection of new suboptima by combining building blocks, may compensate for diversity loss due to the extinction of subpopulations.

8. REFERENCES

[1] H. Asoh and H. Mühlenbein. On the mean convergence time of evolutionary algorithms without selection and mutation. In *Proc. Parallel Problem Solving from Nature III*, pages 88–97, 1994.

[2] H.-G. Beyer. On the Dynamics of EAs without Selection. In W. Banzhaf and C. Reeves, editors, *Foundations of Genetic Algorithms, 5*, pages 5–26, San Mateo, CA, 1999. Morgan Kaufmann.

[3] H.-G. Beyer and H.-P. Schwefel. Evolution strategies: A comprehensive introduction. *Natural Computing*, 1(1):3–52, 2002.

[4] C. L. Bridges and D. E. Goldberg. An analysis of reproduction and crossover in a binary-coded genetic algorithm. In J. J. Grefenstette, editor, *Proc. 2nd Int'l Conf. on Genetic Algorithms ICGA 87*, pages 9–13. Lawrence Erlbaum Associates, 1987.

[5] K. Deb and D. E. Goldberg. An investigation of niche and species formation in genetic function optimization. In J. D. Schaffer, editor, *Proc. 3rd Int'l Conf. on Genetic Algorithms ICGA 89*, pages 42–50, San Mateo, CA, 1989. Morgan Kaufmann.

[6] D. E. Goldberg and K. Deb. A comparative analysis of selection schemes used in genetic algorithms. In G. J. E. Rawlins, editor, *Foundations of Genetic Algorithms FOGA 91*, pages 69–93, San Mateo, CA, 1991. Morgan Kaufmann.

[7] T. Jansen and I. Wegener. Real royal road functions - where crossover provably is essential. In L. Spector, editor, *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO-2001)*, pages 375–382, San Francisco, California, USA, 2001. Morgan Kaufmann.

[8] J. G. Kemeny and J. L. Snell. *Finite Markov chains*. D. Van Nostrand Company, Ltd., London, 1969.

[9] M. Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.

[10] A. Prügel-Bennett. Symmetry breaking in population-based optimization. *IEEE Transactions on Evolutionary Computation*, 1(8):63–79, 2004.

[11] A. Rogers and A. Prügel-Bennett. Genetic Drift in Genetic Algorithm Selection Schemes. *IEEE Transactions on Evolutionary Computation*, 3(4):298–303, 1999.

[12] L. Schönemann, M. Emmerich, and M. Preuss. On the extinction of evolutionary algorithms sub-populations on multimodal landscapes. *Informatica (Sl), Special Issue on Bioinspired Optimization*, 2004.

[13] J. L. Shapiro and A. Prügel-Bennett. Genetic algorithm dynamics in two-well potentials with basins and barrier. In *Proceedings of Foundations of Genetic Algorithms – 4*, pages 101–116. Morgan Kaufmann, 1997.

[14] W. Spears. Simple subpopulation schemes. In A. V. Sebald and L. J. Fogel, editors, *Proc. Third Annual Conf. Evolutionary Programming (EP'94)*, pages 296–307, Singapore, 1994. World Scientific.

[15] F. Streichert, G. Stein, H. Ulmer, and A. Zell. A clustering based niching method for evolutionary algorithms. In E. Cantú-Paz, editor, *Genetic and Evolutionary Computation – GECCO-2003*, pages 644–645, Berlin, 2003. Springer-Verlag.

[16] R. K. Ursem. Multinational evolutionary algorithms. In P. J. Angeline, editor, *Proceedings of the Congress of Evolutionary Computation (CEC-99)*, volume 3, pages 1633–1640, Piscataway, NJ, 1999. IEEE Press.