Fuzzy recombination for the Breeder Genetic Algorithm*

H.–M. Voigt
T.U. Berlin
13355 Berlin
Germany

H. Mühlenbein
GMD
53754 St. Augustin
Germany

D. Cvetković
GMD
53754 St. Augustin
Germany

Abstract

A new recombination operator, called fuzzy recombination (FR) is introduced for continuous genes. The performance of the operator is analyzed by means of the equation describing the response to selection. The operator is evaluated according to a new design criterion: maximizing the product of heritability and standard deviation. The breeder genetic algorithm BGA with FR converges linearly for a test suite of benchmark functions. The computational complexity is also computed. We believe that linear convergence is the optimum to be achieved by random search methods. The question remains open: Can a random search method be found which gives the best linear convergence, i.e. the smallest constant for a well-defined class of functions?

1 Introduction

Let an optimization problem be given on a domain $G \subset R^n$

$$f^* = f(x^*) = \min_{x \in G} f(x), \quad G \subset R^n. \tag{1}$$

We make no assumptions concerning the convexity and differentiability of the function $f(x)$. For the minimization a number of algorithms have been proposed. In this paper we apply the Breeder Genetic Algorithm BGA [MSV93] to obtain approximations. The BGA uses a continuous representation. We will investigate several continuous recombination operators, both empirically and theoretically. Recombination operators can best be analyzed by the equation for the *response to selection*. The equation states that the progress of the average fitness of the population is proportional to the *selection intensity*, the *heritability* and the *standard deviation* of the fitness.

This equation leads to a design principle for operators. An efficient recombination operator should maximize the progress of the average fitness for a number of generations. This means that it should *maximize the product of the heritability and the standard deviation.* Obviously these two goals — maximizing the heritability and maximizing the standard deviation — are antagonistic. This makes the design and the analysis of recombination operators particularly difficult. Recent investigations of operators [MS94] [DW94] have only used one goal, maximizing the correlation between parents and offspring. But the correlation coefficient is directly related to heritability used in the equation for the response to selection [AM94]. Therefore maximizing the correlation alone is not a useful design criterion for a recombination operator. The reduction of the standard deviation has to be taken also into account.

In this paper and also in previous papers we distinguish between empirical laws and theorems. Empirical laws are derived from carefully performed computer experiments. Theorems are obtained by purely mathematical reasoning. Empirical laws are by no means less true than theorems. They are laws carefully deduced from the results of numerical experiments. This procedure was and is successfully used in physics. A historical example are the laws describing the movements of the planets. Kepler derived his famous laws empirically. They explained all the available data, in addition they could be used for prediction. Newton was able to derive the same laws by postulating a gravitational force between the sun and the planets. Thus in Newton's theory Kepler's laws can be proven mathematically. In my terminoloy Newton converted an empirical law to a theorem by a theory. Now back to our application. We also would like to derive some of our empirical laws from a mathematical theory, or at least part of the laws. Unfortunately many of the empirical laws seem mathematically almost intractable.

*This work was partly funded by the Real World Computing Partnership under the project SIFOGA;appeared in Proc. Sixth Inter. Conf. Genetic Algorithms pp.104-111, Morgan Kaufmann, 1995

But we succeeded to derive theoretically part of main empirical law of this paper.

The main source of confusion is that the word "empirical" has one sense in which it refers to something based purely on observation without theoretical depth. But we use the word in its classical sense. Our thesis is that the classical empirical approach is a viable alternative that should be pursued more consciously and more rigorously.

The outline of the paper is as follows. In section 2 we introduce the mathematical terms which are necessary to evaluate random search algorithms. Then the equation for the response to selection is used to investigate several recombination operators. For the evaluation unimodal test functions will be used. It will be shown that a new recombination operator called *fuzzy recombination* gives on the average the best results. Fuzzy recombination is not derived from the theory of fuzzy sets, it only resembles certain aspects of fuzzyfication. We empirically derive prediction formulas for gen^*, the number of generations needed to achieve a solution quality of $|f^* - f| \leq \epsilon$. We show that for a test suite of unimodal functions the convergence is linear. In section 4 this result is derived theoretically. We conclude the paper with some results for multimodal functions.

2 Order of convergence and computational complexity

The definition of acceptable norms by which to evaluate and compare the efficiency of random search techniques remains a major research question. The best method seems to be the study of the distribution of the number of steps required to reach the essential infimum. The algorithms can then be evaluated by comparing the expected number of steps and/or higher moments of this distribution. To do this, we must rely on idealized benchmark situations. Clearly, not all possible functions can serve as test functions for such an investigation.

As a first step in this direction we proposed in [MSV94] to investigate the scaling of a given algorithm for a suite of test functions. Scaling defines the *computational complexity* of the algorithm. If a test function is defined for an arbitrary number of variables n, then the expected number of steps required to reach the essential infimum as a function of n has to be computed. Computational complexity has been a very useful concept in computer science.

In numerical analysis the *order of convergence* is used as a first criterion for evaluation. The order of convergence measures how fast the approximations converge to the infimum. The function is held fixed. There are two convergence measures, one defined for the function values $f(x)$, one defined for the x values. For simplicity we restrict our definition to the case that the infimum is unique.

Definition 1: *Let x^* denote the infimum and $f^* = f(x^*)$. Let $\|x\|$ be a norm in R^n. Then the order of convergence is* **linear** *in f (or in $\|x\|$), if there exists a constant $c_f < 1$ (or $c_x < 1$) such that*

$$|f(x_{t+1}) - f^*| \leq c_f \cdot |f(x_t) - f^*| \tag{2}$$

or

$$\|x_{t+1} - x^*\| \leq c_x \cdot \|x_t - x^*\| \tag{3}$$

The difficult relation between convergence in f and in $\|x\|$ will not be discussed here. Algorithms that converge as a higher power, i.e.

$$|f(x_{t+1}) - f^*| \leq c_f \cdot |f(x_t) - f^*|^m \quad m > 1$$

are said to converge superlinearly. A famous example is the Newton-Raphson algorithm which converges quadratically. But linear converge is not bad at all. Linear convergence means that succesive significant digits are won linearly with computational effort. In other contexts the above linear convergence would be termed "exponential" or "geometrical". We believe that linear convergence is the best one can achieve for random search methods which do not use the derivative of the given function.

Unfortunately linear convergence is differently defined if the function has to be approximated by some series of known functions. Here it means

$$\|f - s_t\| \leq \frac{c}{t}\|f - s_{t-1}\|$$

where s_t is the approximation obtained with e.g t data points.

The following lemma can be easily proven. It gives the average number of steps required to reduce the error by a factor of ϵ.

Lemma 1: *If the order of convergence is linear in f or in x, then the avarage number of steps s_f or s_x required to reduce $|f(x_t) - f^*|$ or $\|x_t - x^*\|$ by a factor of ϵ is bounded by*

$$s_f \leq \frac{\ln(\epsilon)}{\ln(c_f)}, \quad and\ s_x \leq \frac{\ln(\epsilon)}{\ln(c_x)}. \tag{4}$$

We will show in the next section that for specific recombination operators the BGA converges linearly in f for a class of unimodal functions. Furthermore we will estimate the computational complexity.

3 Analysis of recombination operators

A number of different recombination operators have been proposed for continuous genes. Some of the most popular are discrete recombination [Sch81], [MSV93],

intermediate recombination [Sch81], extended intermediate recombination [MSV93], extended line recombination [MSV93], fuzzy Min-Max recombination [Voi92], linear crossover [Wri91] and BLX-O.a crossover [ES92]. A thorough evaluation of these recombination operators has not yet been done. In this section we will analyze discrete, intermediate and extended intermediate recombination and a new soft recombination scheme gleaned from fuzzy set theory.

Let $(x_1, \ldots, x_n)$ and $(y_1, \ldots, y_n)$ be the selected parent chromosomes. With discrete recombination (DR) the offspring variable z_i is chosen randomly from x_i and y_i. With intermediate recombination (IR) the offspring variable is given by $(x_i \leq y_i)$

$$z_i = x_i + \alpha_i \cdot (y_i - x_i)$$

where α_i is either fixed to 0.5, chosen randomly in the interval $[0, 1]$ (IR in the narrow sense) or chosen randomly in the interval $[-d, 1+d]$ (extended intermediate recombination (EIR) [MSV93]). The rationale behind EIR is to introduce more variance. Fuzzy recombination (FR) is inspired from fuzzy set theory. The probability that the offspring has the value z_i is given by a bimodal distribution,

$$p(z_i) \in \{\phi(x_i), \phi(y_i)\}, \tag{5}$$

with triangular probability distributions $\phi(r)$ having the modal values x_i and y_i with

$$\begin{array}{l} x_i - d \cdot |y_i - x_i| \leq r \leq x_i + d \cdot |y_i - x_i| \\ y_i - d \cdot |y_i - x_i| \leq r \leq y_i + d \cdot |y_i - x_i| \end{array} \tag{6}$$

for $x_i \leq y_i$ and $d \geq 0.5$. We mainly used $d = 0.5$ for the simulations. All recombination operators are volume oriented. They create offspring randomly within a hyper-rectangle defined by the parent points.

We now derive a design criterion for recombination operators. The analysis is based on the equation for the *response to selection* [MSV93], [MSV94]. The response R is defined as the difference between the population mean fitness $\bar{f}$ of generation $t+1$ and the population mean of generation t, $R(t) = -\bar{f}(t+1) + \bar{f}(t)$. (In population genetics a trait is normally maximized, so $\bar{f}(t+1) - \bar{f}(t)$ is used.) Breeders measure selection with the *selection differential*, which is symbolized by S. It is defined as the difference between the mean fitness of the selected parents $\bar{f}_s(t)$ and the mean fitness of the population, $S(t) = -\bar{f}_s(t) + \bar{f}(t)$.

The prediction of the response to selection starts with

$$R(t) = b_t \cdot S(t). \tag{7}$$

b_t is called *realized heritability* in quantitative genetics. The breeder either measures b_t in previous generations or estimates b_t using different methods [MSV94]. It is normally assumed that b_t is constant for a certain number of generations. There is no genetics involved in this equation. It is simply an extrapolation from direct observation. The prediction of just one generation is only half the story. The breeder (and the GA user) would like to predict the cumulative response R_s for s generations of his breeding scheme.

$$R_s = \sum_{t=1}^{s} R(t) = b \sum_{t=1}^{s} S(t) \tag{8}$$

The response to selection is the product of the heritability and the sum of the selection differentials. For predicting the response to selection b and the selection differentials have to be estimated. Breeders often use *truncation selection* or *mass selection*. In truncation selection with threshold T, the T% best individuals will be selected as parents. T is normally chosen in the range 10% to 50%. The problem of estimating the selection differential is a problem of order statistics. If the fitness values form a normal distribution, the selection differential $S(t)$ can be computed from

$$S(t) = I \cdot \sigma(t) \tag{9}$$

where σ is the phenotypical standard deviation. I is called the *selection intensity*. It depends nonlinearly on T ([Bul80]). For arbitrary distributions the following estimate can be shown [Nag81]

$$S(t) \leq \sqrt{\frac{100 - T}{T}} \sigma(t) \tag{10}$$

The equation for the response to selection leads to a design criterion for genetic operators. In order to maximize the cumulative response, the genetic operator should *maximize the product of the realized heritability* and *the standard deviation* of the offspring generation. This design criterion we will subsequently use to analyze the recombination operators defined above.

For the evaluation the following unimodal test functions will be used. They consist of a standard function (*sphere*), a function where the variables have different importance for the fitness function (*ellipse*), a function where the minimum is at the boundary (*sum*) and a function which is not differentiable at the infimum (*pyramid*).

$$\begin{array}{ll} F_{sphere}(x) = \sum_{i=1}^{n} x_i^2 & |x_i| \leq 1 \\ F_{ellipse}(x) = \sum_{i=1}^{n-1} x_i^2 + 10^4\, x_n^2 & |x_i| \leq 1 \\ F_{sum}(x) = \sum_{i=1}^{n} x_i & 0 \leq x_i \leq 1 \\ F_{pyramid}(x) = (\sum_{i=1}^{n} (1 - |x_i|))/n & |x_i| \leq 1 \end{array}$$

The performance of the recombination operators is shown in Figure 1. If a good approximation is required, then FR or EIR with $d = 0.5$ should be used. They linearly converge to the solution up to a precision of 10^{-12}. EIR with $d = 1$ does not converge at all. Selection reduces the search space, but EIR places the offspring in the whole area. So selection is counterbalanced by this recombination operator. In contrast, DR converges very early. DR would need a huge population size to achieve a good approximation. For IR

the mean fitness decreases the fastest, but it also converges prematurely. IR reduces the variance too fast. Before we analyze the results in more detail, we formulate the most important result of our simulations as a law.

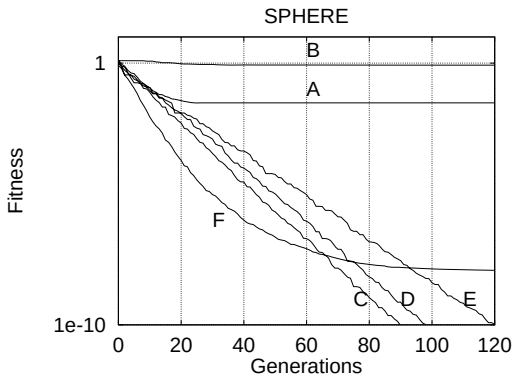


Figure 1: *Sphere*, n=32, popsize N=512, I=1.4; A: DR, B: EIR with $d = 1.0$, C: FR with $d = 0.5$, D: EIR with $d = 0.5$, E: FR with $d = 1.0$, F: IR : mean fitness

Empirical law 1: *The early generations of the BGA simulation always converge linearly in f, for all recombination operators, excluding EIR with $d = 1$. The slope of this linear region is independent of the size of the population N. A larger N increases the length of the linear region (but not the slope) and locates the optimum with higher accuracy.*

Our subsequent analysis will be restricted to the region giving linear convergence. The challenge for the GA designer is to find a recombination operator which gives a large linear region and a steep slope. This means that this operator has to introduce the right portion of variance into the offspring population. If the variance is too large then the algorithm does not converge at all, if it is too small then it converges prematurely.

This problem is investigated for EIR in more detail in figure 2. Here we vary d from 0 to 1. As already known, IR (EIR with $d = 0$) converges the fastest, but the convergence stops early. EIR with $d = 0.25$ gives the best results, if a solution accuracy of at least 10^{-6} is required. EIR with $d = 1$ does not converge at all. The value of the best d gets larger if the infinum is not contained in the interior. Here it would be obviously better to have a larger d. Simulations confirm this statement. We just report the results for the function *sum*. For $d = 0.25$ we have premature convergence, $d = 0.5$ gives the best results. Therefore it seems that $d = 0.5$ is a good choice for a large class of functions. The same is true for fuzzy recombination FR.

Next we make a more detailed analysis of the recombination operators by using the design criterion for recombination operators defined earlier. First we compute the realized heritability $b_t = R(t)/S(t)$ for DR,

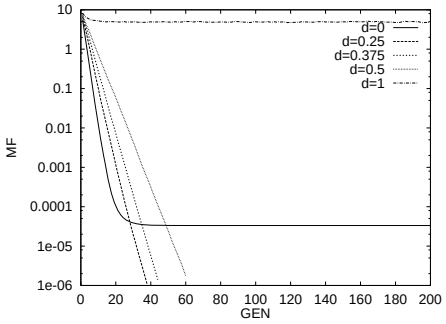


Figure 2: *Sphere*, $n = 32$, popsize $N = 1024$, $I = 1.2$: Mean fitness MF

IR, EIR and FR with $d = 0.5$. In figure 3 the results for the linear function *sum* are displayed. The heritability of DR and IR is about 1.0. FR has a heritability of about 0.83 and EIR of about 0.73. Heritability favours DR and IR, but we have already seen that DR and IR converge prematurely. This confirms our statement made in the introduction that a large heritability is not enough. The operator has also to create enough variance. A closer analysis of the reduction of variance will be made for fuzzy recombination in the next section. Figure 1 indicates that FR performs about 10% better than EIR. The reason for this fact shows figure 2(b). The heritability of FR is more than 10% larger than for EIR.

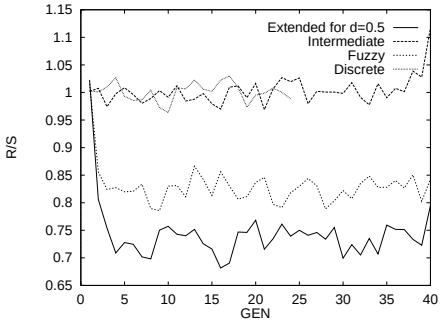
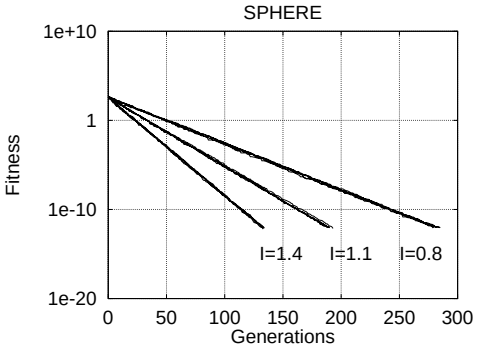


Figure 3: *Sum*, $n = 32$, $N = 1024$, $I = 1.6$: realized heritability R/S

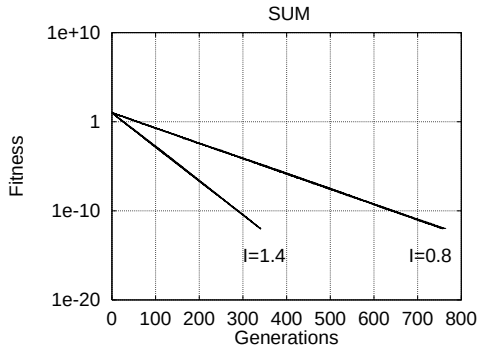
Our simulations have shown that the behavior of the fuzzy recombination algorithm is almost deterministic. In particular the average function value decreases similarly for runs with the same set of parameters but with different initial population. As stated in empirical law 1, the linear convergence part is independent of the population size N, if N is greater than a critical population size N^*. N^* is the size needed to approximate the infinum with the required precision. It depends om I, n and ϵ. The determination of N^* is extremely difficult and cannot be discussed in this paper. Instead we will empirically determine the computational complexity of the algorithm.

Definition 2: *Let N be the size of the population used by the breeder genetic algorithm. Then $gen^*(N, I, n, \epsilon, f)$ is defined as the number of generations needed to obtain an approximation $\tilde{f}$ such that $|f^* - \tilde{f}| \leq \epsilon$.*

We will first investigate the dependence of $gen^*(N, I, n, \epsilon, f)$ on I, afterwards on n and on f. We assume that we have a sufficiently large population, so that the required approximation accuracy can be obtained. Then gen^* is independent from N by empirical law 1. Therefore we will subsequently write $gen^*(I, n, \epsilon, f)$.



(a)



(b)

Figure 4: Average fitness of (a) *Sphere* (b) *Sum*, $n = 32$, $N = 512$, 5 runs overlaid.

Some of our simulation results are displayed in Figure 4. Note that the behavior of the algorithm is almost deterministic. There is no visible difference between the five runs. The figure suggests an inverse proportionate dependence of gen^* on I. Using additional simulation runs, we applied $Mathematica^{TM}$ to fit the data and got the relation:

$$gen^*(I, n, \epsilon, f) \approx \frac{c_1(n, f, \epsilon)}{I^{1.4}}. \tag{11}$$

In the same way, we also determined that

$$gen^*(I, n, \epsilon, f) = c_2(f, I, \epsilon) \cdot n^{0.7}. \tag{12}$$

These two empirical results can be combined, giving an estimate for the computational complexity

$$gen^*(I, n, \epsilon, f) = c(f, \epsilon)\frac{n^{0.7}}{I^{1.4}} \tag{13}$$

This estimate is surprisingly similar to the one obtained for the discrete ONEMAX function [MSV94]

$$gen^* = c\frac{\sqrt{n}}{I}.$$

The difference of the two formulas can be partially explained. A more precise look at the heritability of the recombination operators shows, that the heritability decreases with the number of dimensions. This decrease is very small ($O(n^{-0.2})$). Therefore we first did not notice this second order phenomenon.

The same analysis can be made also for convergence in $\|x\|$. We have used the usual Euclidian norm in R^n. This measure is independent from the function values. Therefore it is more general. Figure 5 shows that the BGA with FR also converges linearly in $\|\bar{x}\|$, where $\bar{x}$ is the average of the x values.

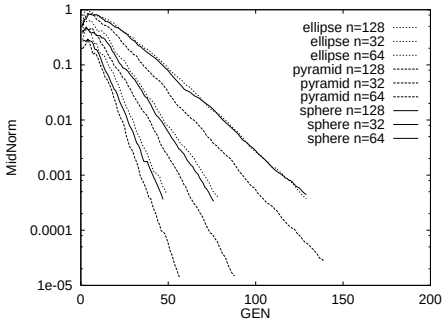


Figure 5: $N = 1024$, $I = 1.2$: Average Euclidian distance $Midnorm$

All curves obey the same scaling law formulated in equation (12) for f. The convergence behavior is only different at the very first generations. Here $\|x\|$ increases slightly. But the scaling laws are not effected by this behavior. Note that the results of *sphere* and *ellipse* are almost identical, despite that these functions are very different. Only the function *pyramid* converges faster. This has to be expected, because its level set is a hyper-rectangle which fits better to our recombination operator. Our recombination operator creates offspring within a hyper-rectangle defined by the selected parents. If the level set is spherical, selection will select parent points lying on spherical discs. The recombination operator will create offspring in a hyper-rectangle which contains the spherical disc.

In the next section we will derive the scaling law from a more theoretical view.

4 Approximate theoretical solution

It has been observed empirically for quite a time that the equation for the response to selection is valid for a large class of problems in quantitative genetics ([Fal81, VCF91, TB94]). Therefore our analysis starts with this equation

$$R(t) = b_l I \sigma(t) \tag{14}$$

b_t was already numerically computed in the previous section. In order to solve the above equation, $\sigma(t)$ has to be estimated. For simple binary functions, we successfully approximated $\sigma(t)$ by a binomial distribution ([MSV94]). In general, an estimate of $\sigma(t)$ for a genetic population under selection is very difficult.

In population genetics the following approach has been tried. The variance $V(t)$ and not the standard deviation σ is investigated. The variance is decomposed into a number of terms, usually into two terms. These are the variance of the selected parents and the additional variance introduced by random mating. Selection reduces the variance, and mating with recombination increases the variance. These two forces have to be balanced.

We just outline the behavior for a population with normal distributed fitness values. The detailed theoretical analysis will be described elsewhere. For normal distributed fitness values the variance V_s of the selected parents can be computed analytically ([Fal81]). It is given by

$$V_s(t) = (1 - I(I - X))V(t). \tag{15}$$

Here, X denotes the abscissa of the truncation point. The variance of the offspring is given by

$$V(t) = 0.5V_s(t-1) + V_r(t-1),$$

where V_r denotes the variance introduced by mating with recombination. The factor 0.5 has to be introduced because two parents give just one offspring, the fitness value of which depends on the midparent value and the some "noise" introduced by mating. If we do not select, then we have $V_s(t-1) = V(t-1)$ and $V(t) = V(t-1)$ because the population is in equilibrium. This gives $V_r(t-1) = 0.5V(t-1)$. We assume that this equation is valid also if selection is done. Combining the equations we obtain

$$V(t) = (1 - 0.5I(I - X))V(t-1). \tag{16}$$

Numerical simulations have confirmed this equation. A BGA with fuzzy recombination decreases the variance by a constant. This constant depends on I, but also on n and the function to be optimized. For the following theorem we will assume

$$\sigma(t) = c\sigma(t-1),$$

where $\sigma = V^{1/2}$. This equation has the solution

$$\sigma(t) = \sigma(0)c^t. \tag{17}$$

It is difficult to estimate c, especially its dependence on the selection intensity I and on the size of the problem n. We have made intensive simulations for $n = 32, 128, 512$ and $I = 2.6, 2.0, 1.6, 1.2$. Due to space limitations we are not able to show the numerical data. The results are summarized by the following law.

Empirical law 2: *For the fitness function sum the constant c_{sum} is given by*

$$c_{sum} \approx \sqrt{1 - \frac{I^{1.4}}{n^{0.7}}} \tag{18}$$

We are now able to prove the following theorem.

Theorem 1: *Under the assumption of empirical law 2 the BGA converges linearly in f. The constants are given by*

$$c_{sum} = \sqrt{1 - \frac{I^{1.4}}{n^{0.7}}} \quad s_{sum} \approx -2\ln(\epsilon) \cdot \frac{n^{0.7}}{I^{1.4}}. \tag{19}$$

Proof: From $R(t) = -\bar{f}(t+1) + \bar{f}(t)$ we obtain from the equation for the response

$$|\bar{f}(t+1)| \leq (1 - bI\frac{\sigma(t)}{\bar{f}(t)})|\bar{f}(t)|. \tag{20}$$

For notational convenience we assume $f^* = 0$. The average fitness is obtained from the sum

$$-\bar{f}(t) + \bar{f}(0) = \sum_{s=1}^{t-1} R(s).$$

This equation can be solved easily

$$\bar{f}(t) = \bar{f}(0) - bI\sigma(0)\frac{1 - c^t}{1 - c}.$$

We assume that the population converges to the infimum, i.e $\bar{f}(t) \to 0$. Then

$$\bar{f}(0) = bI\sigma(0)\frac{1}{1 - c}.$$

Inserting this equation, we obtain

$$\bar{f}(t) = bI\sigma(0)\frac{c^t}{1 - c} = bI\frac{\sigma(t)}{1 - c}$$

Therefore we get the equation

$$\frac{\bar{f}(t)}{\sigma(t)} = \frac{bI}{1 - c} \tag{21}$$

Inserting this expression into the first equation we obtain $\bar{f}(t+1) \leq c \cdot \bar{f}(t)$. This proves the first conjecture.

The constant s_{sum} is obtained from lemma 1 by replacing $\ln\sqrt{1-x}$ in 4 with $-0.5x$. □.

The estimate given in equation (19) is quite accurate. If we set $\epsilon = 10^{-10}$, $I = 1.4$, $n = 32$, we obtain $s_f \approx 325$. From figure 3 we get $s_f \approx 290$. For different functions the constant c_f is different, but the constant s_f shows always the same asymptotic behavior concerning I and n.

Equation 21 is an interesting result by itself. It shows how the variance has to be balanced. In order to achieve linear convergence the mean distance to the optimum divided by the variance has to be a constant. Unfortunately the mean distance to the optimum is not known for a real BGA run, therefore this quotient cannot be directly used.

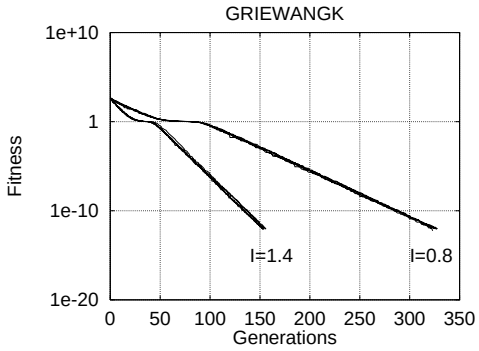
The BGA with fuzzy recombination has been applied to a number of real-life functions. In the next section we show just one interesting example in the domain of multimodal functions.

5 Multimodal functions

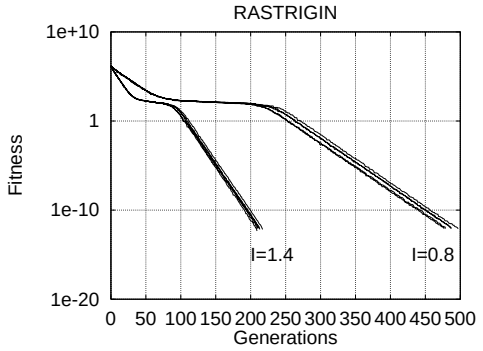
The BGA with FR is a volume-oriented search which is also able to locate the global minimum of multimodal functions. The analysis presented in this paper can be extended to multimodal functions. The results for two popular multimodal functions, Rastrigin's and Griewank's function [MSV93] are shown in Figure 6.

Figure 6 shows that the convergence speed at the beginning and at the end of the search is exactly like the convergence speed for the sphere. In both curves there is a plateau where the speed of convergence is reduced. The reason for this behavior lies in the structure of these functions. On a broad scale both look like a sphere. If the attractor region of the infimum is found, the functions again look like a sphere. In-between these two regions they oscillate like a sine function, and the convergence slows down because the heritability gets very small. The plateau region is very small for strong selection ($I = 1.4$). Therefore the number of generations needed to approximate the global minimum with an accuracy of $\epsilon = 10^{-10}$ is almost the same as for the unimodal sphere.

We postpone a more detailed investigation of multimodal functions to a larger paper. But the following observations can be made. Multimodal function where the smallest minima are clustered in an area are easily optimized.



(a)



(b)

Figure 6: $N = 512$, $n = 32$, FR, 5 runs overlaid.

6 Conclusion

We have shown that a BGA with fuzzy recombination converges linearly for a set of benchmark functions. An exact mathematical definition of the class of functions where the BGA will converge linearly is difficult. But some comments can be given describing the class of functions where the BGA with FR alone will not converge. Fuzzy recombination is obviously a volume oriented search. The volume to be searched is a hyperrectangle defined by the parents. The hyperrectangle is parallel to the axes. If the minimum of the function is located at the end of a very steep and curved valley, this recombination method is obviously not able to locate the minimum. The steep valley is a very small part of the hyperrectangle. This problem can be solved by introducing additional operators, like line recombination [MSV93]. Another solution is to transform the axes, so that the direction of the valley is one of the axis.

In our opinion this paper makes two major contributions. First, it shows that the classical science of breeding can also be used as the theory for genetic algorithms. Second, we prove that the BGA converges linearly for a large class of unimodal functions. We suspect that several of the popular GA implementations do not converge linearly. We cannot be sure because most researchers still compare GA's on a very small and fixed set of benchmark functions. This is not a mature method of empirical science mentioned in the introduction. Such experiments do not give insight, but just numbers. Therefore we iterate our hope [MSV93] that researchers proposing new GA implementations investigate the order of convergence and the computational complexity of their algorithms. This is the only way to change the research from mystic and individual belief to a science with a solid foundation.

References

[AM94] H. Asoh and H. Mühlenbein. Estimating the heritability by decomposing the genetic variance. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature*, Lecture Notes in Computer Science 866, pages 98–107. Springer-Verlag, 1994.

[Bul80] M. G. Bulmer. *The Mathematical Theory of Quantitative Genetics.* Clarendon Press, Oxford, 1980.

[DW94] J. Dzubera and D. Whitley. Advanced correlation analysis of operators for the traveling salesman problem. In Y. Davidor, H.-P. Schwefel, and R. Männer, editors, *Parallel Problem Solving from Nature*, Lecture Notes in Computer Science 866, pages 68–77. Springer-Verlag, 1994.

[ES92] L.J. Eshelman and J.D. Schaffer. Real-coded Genetic Algorithms and Interval-schemata. In L.D. Whitley, editor, *Foundations of Genetic Algorithms*, pages 187–202, San Mateo, 1992. Morgan-Kaufman.

[Fal81] D. S. Falconer. *Introduction to Quantitative Genetics.* Longman, London, 1981.

[MS94] B. Manderick and P. Spiessens. How to select genetic operators for combinatorial optimization problems by analyzing their fitness landscape. In J.M. Zurada, R.J. Marks II, and Ch.J. Robinson, editors, *Computational Intelligence: Imitating Life*, pages 170–181, New York, 1994. IEEE Press.

[MSV93] H. Mühlenbein and D. Schlierkamp-Voosen. Predictive Models for the Breeder Genetic Algorithm I. Continuous Parameter Optimization. *Evolutionary Computation*, 1:25–49, 1993.

[MSV94] H. Mühlenbein and D. Schlierkamp-Voosen. The science of breeding and its application to the breeder genetic algorithm. *Evolutionary Computation*, 1:335–360, 1994.

[Nag81] H.N. Nagaraja. Some finite sample results for the selection differential. *Ann. Inst. Statist. Math.*, 33:437–448, 1981.

[Sch81] H.-P. Schwefel. *Numerical Optimization of Computer Models.* Wiley, Chichester, 1981.

[TB94] M. Turelli and N.H. Barton. Genetic and statistical analyses of strong selection on polygenic traits: What, me normal. *Genetics*, 138:913–941, 1994.

[VCF91] E. Verrier, J.J. Colleau, and J.L. Foulley. Methods for predicting response to selection in small populations under additive genetic models: a review. *Livestock Production Science*, 29:93–114, 1991.

[Voi92] H.-M Voigt. Fuzzy evolutionary algorithms. Technical report, International Computer Science Institute TR-92-038, 1992.

[Wri91] A.H. Wright. Genetic algorithms for real parameter optimization. In G. Rawlins, editor, *Foundations of Genetic Algorithms*, pages 205–220, San Mateo, 1991. Morgan-Kaufman.