Extended Selection Mechanisms in Genetic Algorithms

Thomas Bäck*

University of Dortmund
Department of Computer Science XI
P.O. Box 50 05 00 · D–4600 Dortmund 50

Frank Hoffmeister†

University of Dortmund
Department of Computer Science XI
P.O. Box 50 05 00 · D–4600 Dortmund 50

Abstract

Common selection mechanisms used in *Evolutionary Algorithms* are combined to form some generalized variants of selection. These are applied to a *Genetic Algorithm* and are subject to an experimental comparison. The feature of *extinctiveness* as introduced in *Evolution Strategies* is identified to be the main reason for a considerable speedup of the search in case of unimodal objective functions.

1 Introduction

Genetic Algorithms (GAs) [Hol75] and *Evolution Strategies* (ESs) [Rec73, Sch81] are two types of algorithms which try to imitate the mechanism of natural evolution. In this paper the generic term *Evolutionary Algorithms* is used to denote such algorithms with the common features of a population of individuals which undergo Darwinian selection of the fitter individuals and which are subject to mutation and sexual recombination processes [BH91, HB91]. The selection mechanism of such algorithms plays an important role for driving the search towards better individuals on the one hand and for maintaining a high genotypic diversity of the population on the other hand. This is directly related to the trade–off between high convergence velocity and high probability to find a global optimum in case of a multimodal problem, which is a well–known problem in current research concerning Evolutionary Algorithms [Bak85, Gol89, Sch81, Whi89].

Within this work we look at the selection techniques which are commonly used in Evolutionary Algorithms and describe a set of possible generalizations and recombinations of them in section 2. These new selection mechanisms are compared by experiments with respect to two simple but important topologies of objective functions in section 3.

2 Selection Schemes

Proportional selection [Hol75, Gol89] and ranking [Bak85, Whi89] are the main selection scheme used in GAs, while ESs are based on several variants of (μ,λ)–selection [Sch81].

To describe these techniques in a formal way the following notation is used: $P^t = (a_1^t, \ldots, a_\lambda^t) \in I^\lambda$ denotes the population at generation $t \in \mathbb{N}$, $\lambda > 1$ the population size, and I is the space of individuals a_i^t. The fitness function $f : I \to \mathbb{R}$ provides the environmental feedback for selection. Furthermore a mapping $rank : I \to \{1, \ldots, \lambda\}$ is given by the following definition:

$$\begin{array}{rl} & \forall i \in \{1, \ldots, \lambda\} : rank(a_i^t) = i \\ \iff & \forall j \in \{1, \ldots, \lambda - 1\} : f(a_j^t) \Box f(a_{j+1}^t) \end{array} \quad (1)$$

where $\Box$ denotes the $\leq$ relation in case of a minimization task and $\geq$ in case of a maximization problem. Consequently, we can use the index i to denote the rank of an individual. In the following we assume that individuals are always sorted according to their fitness, with a_1^t being the best individual of P^t.

Selection in Evolutionary Algorithms is defined by selection (reproduction) probabilities $p_s(a_i^t)$ for each individual within a population. At present the following selection schemes exist:

- *Proportional Selection* [Hol75]:

$$p_s(a_i^t) = f(a_i^t) / \sum_{j=1}^{\lambda} f(a_j^t)$$

- *Linear Ranking* [Bak85]:

$$p_s(a_i^t) = \frac{1}{\lambda}\left(\eta_{max} - (\eta_{max} - \eta_{min})\frac{i-1}{\lambda - 1}\right)$$

where $\eta_{min} = 2 - \eta_{max}$ and $1 \leq \eta_{max} \leq 2$.

*baeck@lumpi.informatik.uni-dortmund.de
†iwan@lumpi.informatik.uni-dortmund.de

- *(μ,λ)-Uniform Ranking* [Sch81]:

$$p_s(a_i^t) = \begin{cases} 1/\mu & , \ 1 \le i \le \mu \\ 0 & , \ \mu < i \le \lambda \end{cases}$$

While the latter two schemes are rank-based (i.e. instead of the actual fitness their rank-index i is used), proportional selection is directly based upon the fitness values of all individuals. When (μ,λ)–selection is also taken into account, a selection scheme can be classified with respect to the following criteria:

- *Dynamic* versus *static* selection:

  The selection probabilities can depend on the actual fitness-values (proportional selection) and hence they change between generations, or they can depend on the rank of the fitness-values only (linear ranking, (μ,λ)–selection) which results in fixed (static) values for all generations. This can easily be formalized as follows:

  DEFINITION 1 (Dynamic Selection)
  A selection scheme is called *dynamic*:

  $$\Longleftrightarrow \ \nexists\, i \in \{1,\ldots,\lambda\} \ \forall\, t \ge 0 \ : \ p_s(a_i^t) = c_i$$
  where the c_i are constants.

  DEFINITION 2 (Static Selection)
  A selection scheme is called *static*:

  $$\Longleftrightarrow \ \forall\, i \in \{1,\ldots,\lambda\} \ \forall\, t \ge 0 \ : \ p_s(a_i^t) = c_i$$
  where the c_i are constants.

- *Extinctive* versus *preservative* selection:

  The term preservative describes a selection scheme, which guarantees a non-zero selection probability for each individual, i.e. each individual has a chance to contribute offspring to the next generation. On the other hand, in an extinctive selection scheme some individuals are definitely not allowed to create any offspring, i.e. they have zero selection probabilities.

  DEFINITION 3 (Preservative Selection)
  A selection scheme is called *preservative*:

  $$\Longleftrightarrow \ \forall\, t \ge 0 \ \forall P^t = (a_1^t,\ldots,a_\lambda^t) \ \forall\, i \in \{1,\ldots,\lambda\} : p_s(a_i^t) > 0$$

DEFINITION 4 (Extinctive Selection)
A selection scheme is called *extinctive*:

$$\Longleftrightarrow \ \forall\, t \ge 0 \ \forall\, P^t = (a_1^t,\ldots,a_\lambda^t) \ \exists\, i \in \{1,\ldots,\lambda\} : p_s(a_i^t) = 0$$

- *Left* versus *right extinctive* selection:

In case of extinctive selection (def. 4) there is a major special case where the worst performing individuals have zero reproduction rates, i.e. do not reproduce. This situation is referred to as right extinctive selection. Although it might be of no practical relevance there may be also the opposite situation (left extinctive selection) where some of the best performing individuals are prevented from reproduction in order to avoid premature convergence due to super-individuals.

DEFINITION 5 (Left Extinctive Selection)
A selection scheme is called *left extinctive*:

$$\Longleftrightarrow \ \forall\, t \ge 0 \ \forall\, P^t = (a_1^t,\ldots,a_\lambda^t) \ \exists\, l \in \{1,\ldots,\lambda-1\} : \ i \le l \implies p_s(a_i^t) = 0$$

DEFINITION 6 (Right Extinctive Selection)
A selection scheme is called *right extinctive*:

$$\Longleftrightarrow \ \forall\, t \ge 0 \ \forall\, P^t = (a_1^t,\ldots,a_\lambda^t) \ \exists\, l \in \{2,\ldots,\lambda\} : \ i \ge l \implies p_s(a_i^t) = 0$$

Of course, in any case the condition $\sum_{i=1}^{\lambda} p_s(a_i^t) = 1$ must be satisfied. With regard to this classification proportional selection is a dynamic, preservative scheme, while linear ranking realizes a static, preservative scheme. (μ,λ)-uniform ranking is static and extinctive. Hence, the main difference is the preservativeness and extinctiveness of the selection schemes, respectively.

Apart from different assignments of reproduction rates there are other characteristics of selection:

- *Elitist* versus *pure* selection:

  Normally, parents are allowed to reproduce in one generation only. Then, they die out and are replaced by some offspring. A selection scheme which enforces a life time of just one generation for each individual regardless of its fitness is referred to as pure selection. In an elitist selection scheme some or all of the parents are allowed to undergo selection with their offspring [Jon75]. This might result in 'unlimited' life times of super-fit individuals.

  DEFINITION 7 (Elitist Selection)
  A selection scheme is called *elitist* or *k-elitist*:

  $$\Longleftrightarrow \ \exists\, k \in \{1,\ldots,\lambda\} \ \forall\, t > 0 \ \forall\, i \in \{1,\ldots,k\} : f(a_i^t) \ \square \ f(a_i^{t-1})$$

  DEFINITION 8 (Pure Selection)
  A selection scheme is called *pure* iff there is no $k \in \{1,\ldots,\lambda\}$ which satisfies the k-elitist property.

- *Generational* versus *steady-state* selection:

  With generational selection the set of parents is fixed until λ offspring, the members of the next generation, are completely produced. In case of selection on-the-fly or steady-state selection an offspring immediately replaces a parent if it performs better. Thus, the set of parents may change for every reproduction step [Whi89].

  It should be noted, that steady-state selection is a special variant of elitist selection (def. 7) where the set of parents incorporated into selection is larger than the set of offspring of size 1.

By "recombining" the major characteristics of the existing selection schemes, proportional selection and ranking can be generalized, which allows them to be also extinctive:

Definition 9 ((μ,λ)-*Proportional Selection*)

$$p_s(a_i^t) = \begin{cases} f(a_i^t)/\sum_{j=1}^{\mu} f(a_j^t) & , \ 1 \leq i \leq \mu \\ 0 & , \ \mu < i \leq \lambda \end{cases} \tag{2}$$

Definition 10 ((μ,λ)-*Linear Ranking*)

$$p_s(a_i^t) = \begin{cases} \frac{1}{\mu}\left(\eta_{max} - 2(\eta_{max}-1)\frac{i-1}{\mu-1}\right) & , \ 1 \leq i \leq \mu \\ 0 & , \ \mu < i \leq \lambda \end{cases} \tag{3}$$

Figure 1 tries to give an impression of the different selection schemes and their interdependencies. In each case selection probabilities versus the rank of the individual are sketched by step-functions. Remember the rank-ordering of individuals such that better individuals have lower ranks.

First, it should be noted, that each extinctive scheme turns into the corresponding preservative scheme for $\mu = \lambda$. For (μ,λ)-uniform ranking the case $\mu = \lambda$ should lead to random walk where the selective pressure towards better individuals is completely lost. The random walk variant is of no interest but mentioned here to complete the classification. (μ,λ)-uniform ranking in ESs is obviously a special case of the extinctive linear ranking selection ($\eta_{max} = 1$).

The selective pressure of the extinctive schemes can be guided by the exogeneous setting of μ. As μ approaches λ, the selective pressure towards the better individuals is decreasing continually and selection becomes "softer".

From theoretical investigations concerning (1,λ)–ES on a simple corridor and sphere model for the objective function, Schwefel derived values of $\lambda \approx 6.0$ for the corridor model and $\lambda \approx 4.7$ for the sphere model to achieve an optimum rate of convergence [Sch81]. The setting ($\mu/\lambda \approx 1/5$) emphasizes on the convergence speed for unimodal problems; for multimodal problems this ratio should be much higher in order to allow for the exploration of the search space to some extent.

With respect to super-individuals with a high fitness value or individuals with just a poor fitness proportional selection appears to be rather "hard", since the resulting rates of reproduction effectively discard the poor ones while high preference is given to the good ones, thus decreasing the genetic diversity quickly. With rank-based schemes the same situation is less drastic since the actual fitness does not influence the realized rate of reproduction, thus yielding a slower reduction of the genetic diversity. Hence, uniform and linear ranking appear to be "softer" than proportional selection.

3 Experimental Results

For the experimental comparison of the selection mechanisms it is concentrated on the two examples of objective functions given in table 1.

The functions f_1 and f_7 are representing the classes of unimodal as well as multimodal functions. For f_1 a high convergence velocity is expected to be sufficient to approach the optimum, while for f_7 a more explorative behaviour of the algorithm would give a chance to find the global optimum. To obtain the results, a modified version of Grefenstette's GENESIS–GA [Gre87] was used here. The GA is defined by the following parameter and configuration settings: Mutation rate $p_m = 0.001$; crossover rate $p_c = 0.6$; population size $\lambda = 50$; length of an individual $l = 32n$, where n denotes the dimension of the objective function[1]; two–point crossover; Gray code. For ranking the usual setting of $\eta_{max} = 1.1$ (maximum expected value) was chosen according to [Bak85].

Different values of $\mu \in \{5, 10, 15, 20, 30, 40, 50\}$ have been used for the test runs, and for a comparison the best values in each generation are plotted. The results are based on the averaged values of 10 runs in each case.

In the left parts of figures 2–4 the performances of (μ,λ)–proportional selection, (μ,λ)–linear ranking, and (μ,λ)–uniform ranking are shown for f_1. Obviously the performance is maximized for rather small values of $\mu \in \{5, 10, 15\}$. In each case performance decreases for growing values of μ, finally turning into the familiar normal ranking and proportional selection plots and a random walk wandering for (μ,λ)–uniform ranking, respectively.

A comparison of the different selection mechanisms for the same values of μ does not lead to a clear general statement, since no large differences exist. A tendency towards favouring (μ,λ)-linear ranking compared with proportional selection and the latter compared with (μ,λ)-uniform ranking can be deduced from a set of graphics not shown here. The major improvement is introduced by the idea of extinctive selection. This result can be interpreted as an indication of the validity of Schwefel's result for an optimum setting of the ratio μ/λ in case of f_1 [Sch81] not only for ESs, but also for GAs. Thus we can formulate the hypothesis, that the effect of an extinctive selection mechanism is

[1] A length of 32 bits per object variable is used for the representation of the real interval $[x_{\min}, x_{\max}]$ to which the bitstrings are mapped, in order to achieve a maximum resolution $\Delta x = (x_{\max} - x_{\min})/(2^{32} - 1)$.

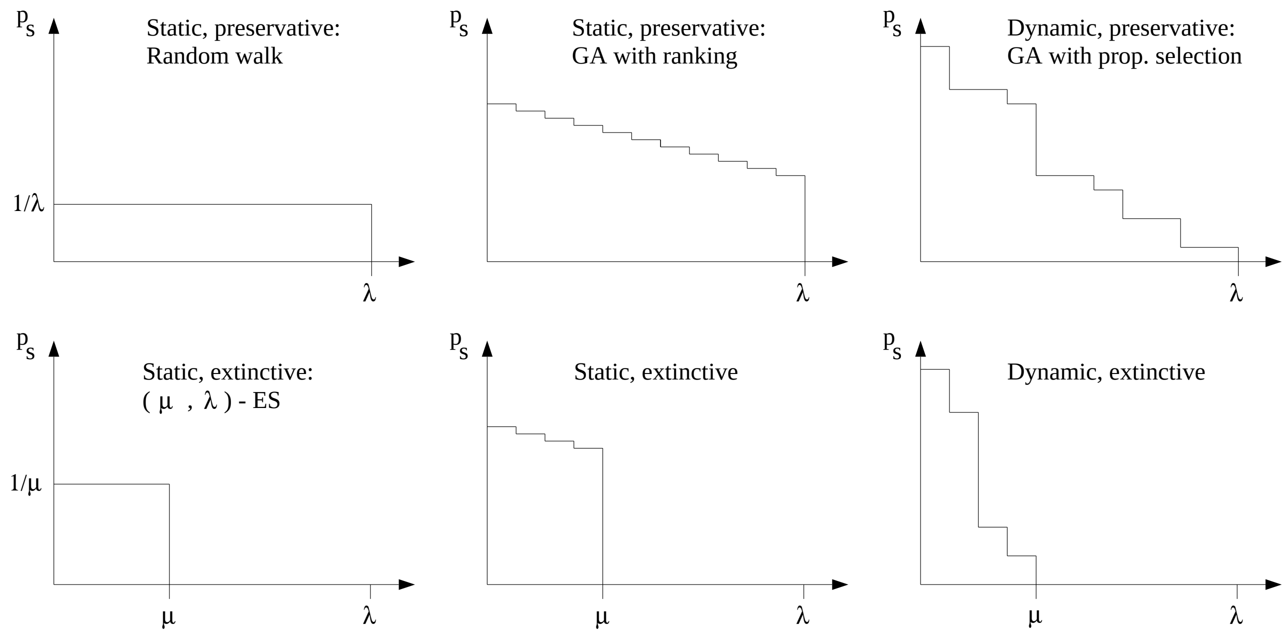

Figure 1: Sketch of selection schemes

Name	Description	Dim.	Characteristics	Ref.
f_1	sphere model $f_1(\vec{x}) = \sum_{i=1}^{n} x_i^2$ $-5.12 \leq x_i \leq 5.12$	$n = 30$	unimodal, high–dimensional	[HB91] [Jon75] [Sch81]
f_7	generalized Rastrigin's function $f_7(\vec{x}) = nA + \sum_{i=1}^{n} x_i^2 - A\cos(\omega x_i)$ $A = 10\,; \omega = 2\pi\,; -5.12 \leq x_i \leq 5.12$	$n = 20$	multimodal, high–dimensional, f_1 with sine wave superposition	[HB91] [TZ89]

Table 1: The set of test functions

very similar for both algorithms, independently of the representation of the individuals and the special kind of genetic operators.

For f_7 the results are completely different to the general results for f_1. The performance plots are given in the right parts of figures 2–4.

At first glance the similarity of all plots characterizing true extinctive selection mechanisms ($\mu < \lambda$) becomes apparent. Furthermore, all extinctive mechanisms as well as preservative proportional selection get stuck in local optima. Only preservative ranking, while progressing very slow in the early phase of a run, seems to promise better results in the long run. Besides random walk this is the only mechanism which behaves basically different compared to the rest. From these plots a cautious hypothesis about an optimum value of μ somewhere between 40 and 50 can be drawn up.

This completely inverse behaviour between f_1 and f_7 is expected to be caused by their topological differences solely.

The remarkably different shape of the preservative ranking mechanism for f_7 can be understood by looking at the genotypic diversity of the populations. This can be measured by the *bias* b ($0.5 \leq b \leq 1$) according to equation (4).

$$b = \frac{1}{l\lambda} \sum_{j=1}^{l} \max \left(\sum_{\substack{i=1 \\ \alpha_{i,j}^t = 0}}^{\lambda} \left(1 - \alpha_{i,j}^t\right) \; , \; \sum_{\substack{i=1 \\ \alpha_{i,j}^t = 1}}^{\lambda} \alpha_{i,j}^t \right) \tag{4}$$

$$\forall P^t = (a_1^t, \ldots, a_\lambda^t) \quad \forall a_k^t = (\alpha_{k,1}^t, \ldots, \alpha_{k,l}^t) \quad \forall t \geq 0$$

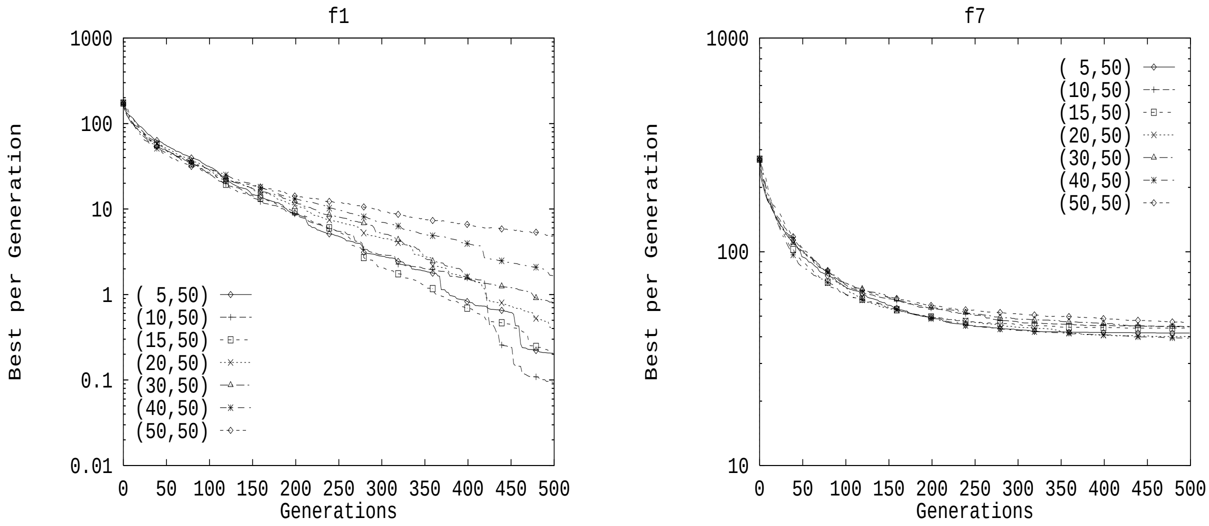

Figure 2: (μ,λ)–proportional selection schemes on f_1 and f_7

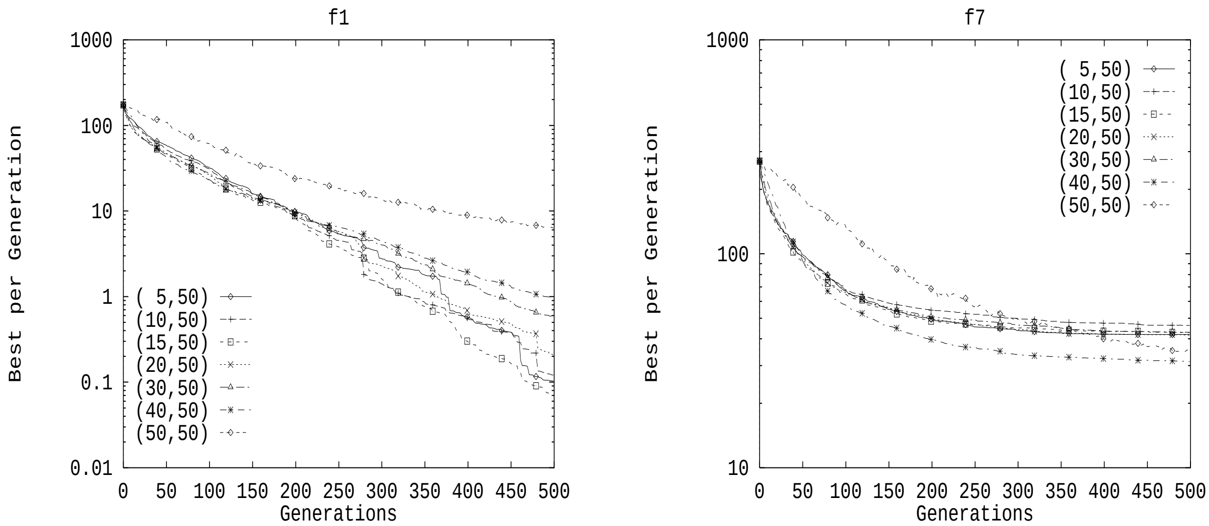

Figure 3: (μ,λ)–linear ranking schemes on f_1 and f_7

b indicates the average percentage of the most prominent value in each position of the individuals [Gre87]. Smaller (larger) values of b indicate higher (lower) genotypic diversity.

The preservative linear ranking mechanism shows a fundamentally larger genotypic diversity than the extinctive ones (left part of figure 5). This is the property which verifies the very slow convergence behaviour of such a selection mechanism. However, it is not experimentally checked whether such behaviour can lead to better solutions in the long run even for difficult surfaces like that of f_7.

A comparison of (50,50)-linear ranking and (40,50)-linear ranking for a longer run on f_7 seems to confirm this assumption, but the difference of the runs is rather small (right part of figure 5).

There are two major effects to observe for different degrees of extinctiveness ($0 < \mu < 50$) which depend on the number of optima. In general for a unimodal function like f_1 the best performance increases with "harder" selection, i.e. decreasing μ, while it stays on a similar level for most degrees of extinctiveness for multimodal functions like f_7. This is *not* a general fact. For different adaptation schemes the impact of selection varies as can be seen from figure 6 which summarizes the first results of a complete set of runs on f_1 and f_7 for all types of selection and various degrees

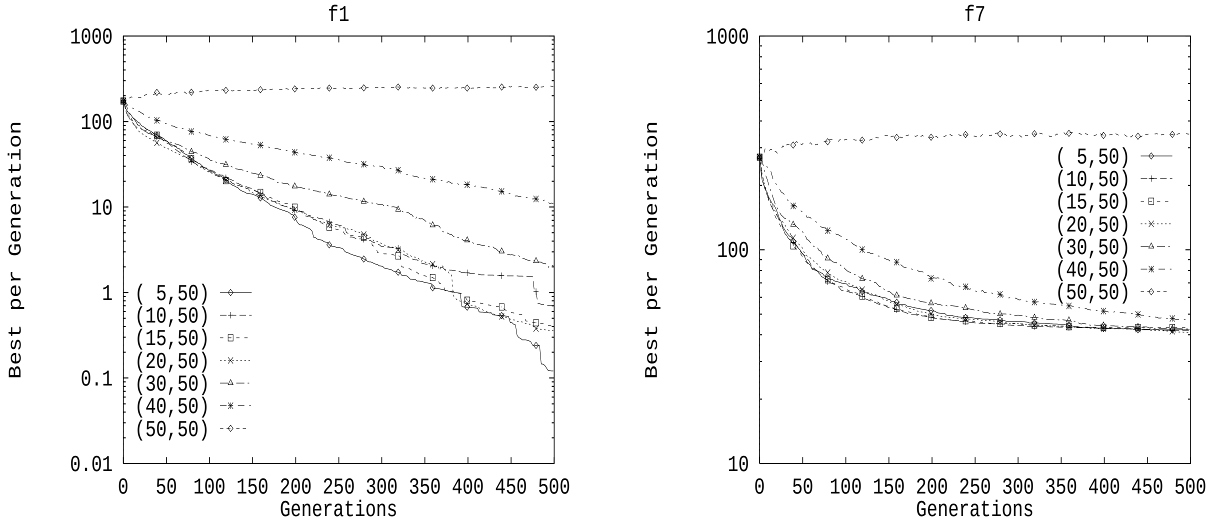

Figure 4: (μ,λ)–uniform ranking schemes on f_1 and f_7

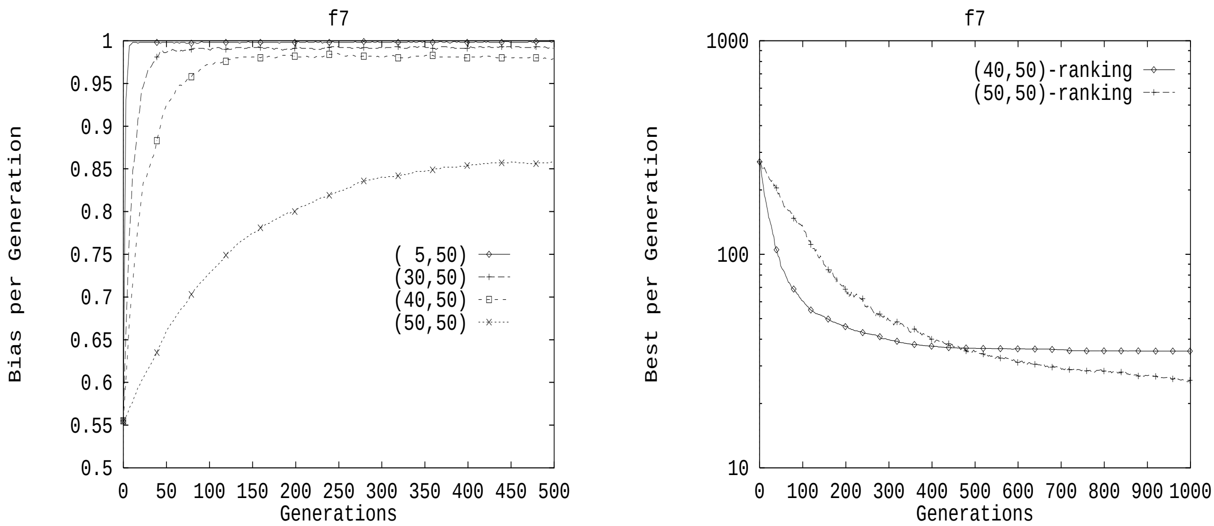

Figure 5: Bias and long runs for (μ,λ)–linear ranking selection schemes on f_7

of extinctiveness with an Evolution Strategy according to Schwefel [Sch81].

Evolution Strategies (ESs) work on a phenotypic level, i.e. they operate directly on the set of real-valued object variables x_i. Mutation is realized by adding to each x_i a normally distributed random number with expected value 0 and standard deviation σ_i. Recombination may be discrete or intermediate. Theoretical considerations for a maximum rate of convergence suggest that the optimal settings of the σ_i may depend on the distance from the optimum, i.e. they are a local feature of the response surface. Therefore, the genetic information of each individual not only consists of the x_i, but also of the strategy parameters σ_i which also undergo mutation and recombination *before* they are used to mutate the x_i. Better adapted settings of the σ_i are expected to result in a better performance of the x_i with respect to the objective function. Hence, selection automatically favours advantageous settings of the strategy parameters σ_i. A detailed description of ESs may be found in [BHS91].

Each curve in figure 6 shows the best solution obtained after 250 generations for a particular selection scheme with respect to various degrees of extinctiveness. Like for the other experiments a population size of $\lambda = 50$ was chosen; all values are averaged over 10 runs.

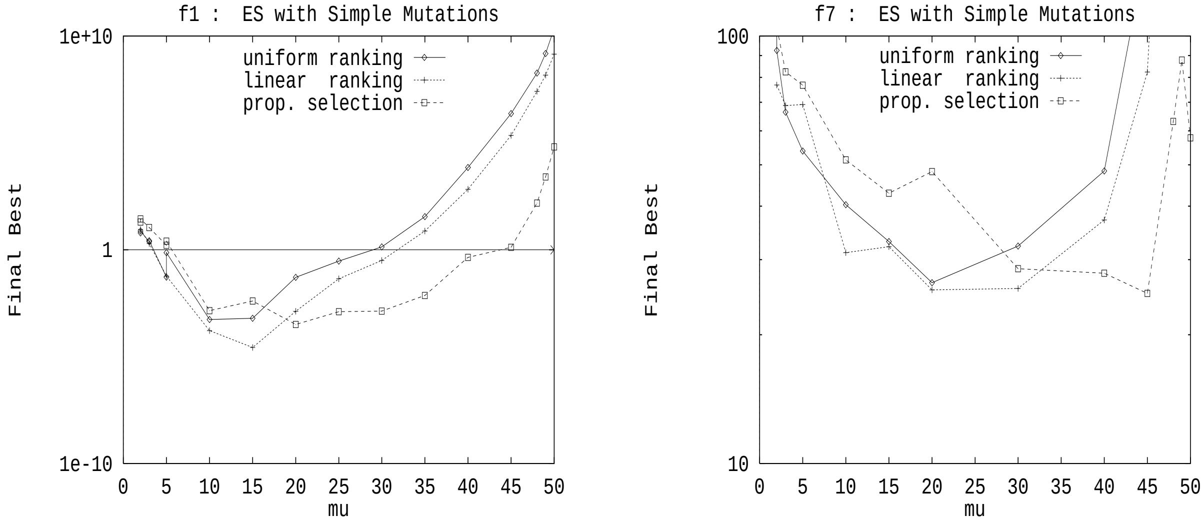

Figure 6: Best performance of ESs on f_1 and f_7 with respect to the degree of extinctiveness

In case of the unimodal function f_1 a high rate of convergence is required for optimal performance. For ranking this is achieved by a setting of $\mu/\lambda \approx 1/5$, which is pretty close to the theoretical results [Sch81]. The curves for f_1 also partially meet the expectation that a "harder" selection scheme (linear ranking) performs better than a "softer" one (uniform ranking). Proportional selection even results in a much "harder" selection, which for a high degree of extinctiveness, is not able to maintain a sufficient genetic diversity to allow for a rapid adaptation of the strategy parameters by means of recombination. This preference for the best is advantageous if extinctiveness is lowered. In that case the ranking schemes fail to maintain a proper setting of strategy parameters finally leading to a divergence from the optimum.

For a multimodal function like f_7 a high genetic diversity is required to explore the search space sufficiently. Hence, the "softer" selection schemes (uniform and linear ranking) perform best with respect to the quality of the final optimum. But if selection becomes too "soft" the population is unable to maintain partial solutions which may be used as a starting point for further improvements. This is why the right illustration in figure 6 shows an optimum at $\mu = 20$ for the ranking schemes while proportional selection performs better with growing μ. But even for proportional selection some degree of extinctiveness is required for an optimal performance.

4 Summary

Undoubtedly an extinctive selection mechanism produces a remarkable speedup for a unimodal function like f_1. This should be a sufficient reason to use such selection mechanisms as a further way of guiding the search of genetic algorithms.

In contrast by extinctiveness there is no improvement of the results for a multimodal surface. Due to this reason a superiority of selection mechanisms which maintain a high genotypic diversity can be concluded from the experimental runs. The question remains how to solve this contradiction concerning the algorithmic requirements caused by different topological surfaces of the actual optimization problem, which is noted as a list of characteristic properties in table 2.

The term convergence confidence is used to describe the probability to converge towards the global optimum. For a multimodal objective function a high convergence confidence is aspired, which requires an explorative character of the search. To achieve this behaviour a "soft" selection scheme can be used in order to maintain a large genotypic diversity of the population during the search. The resulting search process can be designated as volume oriented.

The corresponding appropriate properties for a unimodal problem aim at increasing the convergence velocity. A rather "hard" selection mechanism forces the search process into the gradient direction, resulting in a path oriented, exploiting search. Consequently, the genotypic diversity remains small.

Unfortunately, for a real–world application the user does not know anything about the objective function's properties. Besides the usual parameterization problem (which settings are appropriate for λ, l, p_m, p_c, η_{max}?) a additional parameter is introduced by extinctive selection.

To solve this problem at least two approaches can be

unimodal objective function	multimodal objective function
convergence velocity	convergence confidence
"hard" selection scheme	"soft" selection scheme
small genotypic diversity	large genotypic diversity
path oriented	volume oriented
exploitative character	explorative character

Table 2: Unimodal and multimodal search properties

thought of. As shown by Schwefel in the framework of Evolution Strategies [Sch81] the *self–learning* of strategy parameters provides a powerful mechanism of internal adaptation of the algorithm with respect to the objective function topology. This is often referred to as *second–level learning* and provides an alternative to the other approach of using a meta–level control algorithm as desribed in [Gre86, GBGK89].

References

[Bak85] James Edward Baker. Adaptive selection methods for genetic algorithms. In J. J. Grefenstette, editor, *Proceedings of the first international conference on genetic algorithms and their applications*, pages 101–111, Hillsdale, New Jersey, 1985. Lawrence Erlbaum Associates.

[BH91] Thomas Bäck and Frank Hoffmeister. Global optimization by means of evolutionary algorithms. In Alexander N. Antamoshkin, editor, *Random Search as a Method for Adaptation and Optimization of Complex Systems*, pages 17–21, Divnogorsk, UdSSR, March 1991. Krasnojarsk Space Technology University.

[BHS91] Thomas Bäck, Frank Hoffmeister, and Hans-Paul Schwefel. A survey of evolution strategies. In Richard K. Belew, editor, *Proceedings of the Fourth International Conference on Genetic Algorithms and their Applications*, San Diego, California, USA, 1991. Morgan Kaufmann Publishers.

[GBGK89] Yu. V. Guliaev, I. L. Bukatova, L. N. Golubeva, and V. F. Krapivin. Evolutionary informatics and "intelligent" special processors. Academy of Sciences of the USSR, Institute of Radio Engineering and Electronics, 1989. Preprint.

[Gol89] David E. Goldberg. *Genetic algorithms in search, optimization and machine learning.* Addison Wesley, 1989.

[Gre86] John J. Grefenstette. Optimization of control parameters for genetic algorithms. *IEEE Transactions on Systems, Man and Cybernetics*, SMC–16(1):122–128, 1986.

[Gre87] John J. Grefenstette. *A User's Guide to GENESIS.* Navy Center for Applied Research in Artificial Intelligence, Washington, D. C., 1987.

[HB91] Frank Hoffmeister and Thomas Bäck. Genetic algorithms and evolution strategies: Similarities and differences. In Hans-Paul Schwefel and Reinhard Männer, editors, *Parallel Problem Solving from Nature*, volume 496 of *Lecture Notes in Computer Science*, pages 455–470, Berlin, 1991. Springer.

[Hol75] John H. Holland. *Adaptation in natural and artificial systems.* The University of Michigan Press, Ann Arbor, 1975.

[Jon75] Kenneth De Jong. *An analysis of the behaviour of a class of genetic adaptive systems.* PhD thesis, University of Michigan, 1975. Diss. Abstr. Int. 36(10), 5140B, University Microfilms No. 76–9381.

[Rec73] Ingo Rechenberg. *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution.* Frommann–Holzboog Verlag, Stuttgart, 1973.

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

[Sch89] J. David Schaffer, editor. *Proceedings of the Third International Conference on Genetic Algorithms and Their Applications*, San Mateo, California, June 1989. Morgan Kaufmann Publishers.

[TZ89] A. Törn and A. Zilinskas. *Global Optimization*, volume 350 of *Lecture Notes in Computer Science.* Springer, Berlin, FRG, 1989.

[Whi89] Darrell Whitley. The GENITOR algorithm and selection pressure: Why rank–based allocation of reproductive trials is best. In Schaffer [Sch89], pages 116–121.