A Partial Order Approach to Noisy Fitness Functions

Günter Rudolph
Department of Computer Science
University of Dortmund
44221 Dortmund / Germany
rudolph@LS11.cs.uni-dortmund.de

Abstract-
If the fitness values are perturbed by noise then they do not have a definitive total order. As a consequence, traditional selection procedures in evolutionary algorithms may lead to obscure solutions. A potential remedy is as follows: Construct a partial order on the set of noisy fitness values and apply those evolutionary algorithms that have been designed for finding the minimal elements of partially ordered sets. These minimal elements are the only reasonable candidates for the unperturbed true solution. A method for reducing the number of candidate solutions is suggested. From a theoretical point of view it is worth mentioning that all convergence results for evolutionary algorithms with partially ordered fitness sets remain valid for the approach considered here.

1 Introduction

The Gaussian distribution is the predominant choice for modeling noise frequently observable in measurings of various kinds. Here, we hold the view that a noise distribution with unbounded support (like the Gaussian, Cauchy, Laplace, Logistic, and others) may be quite unrealistic. Actually it is at least equally plausible to assume that the noise cannot exceed certain limits due to technical characteristics of the involved measurement unit. Even if a distributional shape close to a Gaussian appears reasonable we can resort to a symmetrical Beta distribution which can converge weakly to a Gaussian distribution under continuously increasing but bounded support (see e.g. Evans et al. 1993, p. 36). This assumption will have significant theoretical and practical impacts on the evolutionary algorithms (EAs) considered here.

Traditional measures for coping with noisy fitness functions in evolutionary algorithms include the resampling of the random fitness value with averaging, the appropriate adjustment (i.e., enlargement) of the population size, and in case of continuous search spaces also the rescaling of inherited mutations; see Beyer (2000) for a summary of work on EAs for noisy fitness functions.

Here, we add yet another avenue for dealing with noisy fitness functions: Instead of using a selection procedure that is based on the totally ordered set of noisy fitness values we endow the probabilistic fitness set with an appropriate partial order and deploy EAs with those selection methods being explicitly designed for coping with arbitrary partially ordered fitness sets (Rudolph 1998, 2001a; Rudolph and Agapie 2000).

Section 2 offers a brief introduction to partially ordered sets in general and in particular to interval orders (Fishburn 1985) which constitute the first step towards the partial order to be used later on. Since the noise is supposed to have bounded support we can easily equip these intervals with a probability measure (representing a noise distribution). Thus, the interval order turns to a partial order on random variables. This subject is detailed in Section 3 which also contains the presentation of the EA along with a discussion of the inherited theoretical properties from the general case (Rudolph 2001a). Preliminary experimental results can be found in Section 4. Section 5 contains the conclusions.

2 Partially Ordered Sets

Let $\mathcal{F}$ be a set. A reflexive, antisymmetric, and transitive relation "$\preceq$" on $\mathcal{F}$ is termed a *partial order relation* whereas a *strict partial order relation* "$\prec$" must be antireflexive, asymmetric, and transitive. The latter relation may be obtained by the former relation by setting $x \prec y := (x \preceq y) \wedge (x \neq y)$. After these preparations one is in the position to turn to the actual objects of interest.

Definition 1 *Let $\mathcal{F}$ be some set. If the partial order relation "$\preceq$" is valid on $\mathcal{F}$ then the pair $(\mathcal{F}, \preceq)$ is called a* partially ordered set *(or short:* poset*). If $x \prec y$ for some $x, y \in \mathcal{F}$ then x is said to* dominate *y. Distinct points $x, y \in \mathcal{F}$ are said to be* comparable *when either $x \prec y$ or $y \prec x$. Otherwise, x and y are* incomparable *which is denoted by $x \parallel y$. If each pair of distinct points of a poset $(\mathcal{F}, \preceq)$ is comparable then $(\mathcal{F}, \preceq)$ is called a* totally ordered set *or a* chain. *Dually, if each pair of distinct points of a poset $(\mathcal{F}, \preceq)$ are incomparable then $(\mathcal{F}, \preceq)$ is termed an* antichain. □

For example, let $\mathbb{I} = \{[x_1, x_2] \subset \mathbb{R} : x_1 \leq x_2\}$ be the set of closed intervals of $\mathbb{R}$ and define

$$
\begin{aligned}
[x_1, x_2] \prec [y_1, y_2] \quad &\text{iff} \quad x_2 < y_1 \,, \\
[x_1, x_2] = [y_1, y_2] \quad &\text{iff} \quad x_1 = y_1 \wedge x_2 = y_2 \,, \\
[x_1, x_2] \preceq [y_1, y_2] \quad &\text{iff} \quad x \prec y \vee x = y \,.
\end{aligned}
$$

It is easily seen that $(\mathbb{I}, \preceq)$ is a partially ordered set in which distinct intervals with a nonvoid intersection are incomparable. Similarly, the infinitely large but countable set $(\mathbb{I}_\varepsilon, \preceq)$ with $\mathbb{I}_\varepsilon = \{[x - \varepsilon, x + \varepsilon] \subset \mathbb{R} : x \in \mathbb{N}_0\}$ with $\varepsilon > 1/2$ is a poset with incomparable elements whereas $(\mathbb{I}_\varepsilon, \preceq)$ with $\varepsilon < 1/2$ is totally ordered and therefore a chain. An example for an antichain is the set of "minimal elements" introduced next.

Definition 2 *An element $x^* \in \mathcal{F}$ is called a* minimal element *of the poset $(\mathcal{F}, \preceq)$ if there is no $x \in \mathcal{F}$ such that $x \prec x^*$. The set of all minimal elements, denoted $\mathcal{M}(\mathcal{F}, \preceq)$, is said to be* complete *if for each $x \in \mathcal{F}$ there is at least one $x^* \in \mathcal{M}(\mathcal{F}, \preceq)$ such that $x^* \preceq x$.* □

In contrast to infinitely large posets the completeness of $\mathcal{M}(\mathcal{F}, \preceq)$ is guaranteed for finitely large posets. Of course, completeness of infinitely large posets is not precluded. For example, the set $(\mathbb{I}_\varepsilon, \preceq)$ with $\varepsilon = 2/3$ is infinitely large and the set of minimal elements $\mathcal{M}(\mathbb{I}_\varepsilon, \preceq) = \{[-\frac{2}{3}, \frac{2}{3}], [\frac{1}{3}, \frac{5}{3}]\}$ is complete.

3 Coping with Noisy Fitness Functions

3.1 Assumptions

Let S be the finite search set and assume that the deterministic fitness function $f : S \to \mathbb{R}$ is perturbed by additive noise Z, i.e., $\tilde{f}(x) = f(x) + Z$ for $x \in S$. As mentioned earlier, here we insist that random variable Z has bounded and *known* support in form of a closed interval of $\mathbb{R}$. For example, Z may have a uniform or symmetric Beta distribution on its support $[-a, a]$ with $a > 0$.

At first it is assumed that every point/individual $x \in S$ is evaluated only once. Later on this assumption is dropped.

3.2 Partial Order Approach

When an individual $x \in S$ is evaluated via $\tilde{f}(x) = f(x) + Z$ then the noisy fitness value is an element of the interval $[f(x) - a, f(x) + a]$. Since the EA only has knowledge of the support bound $a > 0$ and in no case of the true fitness value $f(x)$, the noisy evaluation of $x \in S$ only leads to the information that the true fitness value $f(x)$ must be in the interval $[\tilde{f}(x) - a, \tilde{f}(x) + a]$. Thus, each point or individual is associated with a realization of a random interval.

Next we declare a strict partial order on these intervals and thereby also a strict partial order on the individuals. Let $x, y \in S$ and w.l.o.g. $\tilde{f}(x) < \tilde{f}(y)$. If

$$\tilde{f}(x) + a < \tilde{f}(y) - a \tag{1}$$

then we define $\tilde{f}(x) \prec \tilde{f}(y)$ and thereby $x \prec y$. This choice is reasonable because we can immediately infer from $x \prec y$ that $f(x) < f(y)$ with probability 1. One should mention that this partial order is a special case of a partial order introduced in Guddat et al. (1985), p. 29. Moreover, notice that the connection to interval orders gets evident by the equivalence between equation (1) and

$$[\tilde{f}(x) - a, \tilde{f}(x) + a] \cap [\tilde{f}(y) - a, \tilde{f}(y) + a] = \emptyset\,. \tag{2}$$

Thus, whenever two intervals as those above have a nonvoid intersection then the noisy fitness values and therefore also the individuals are incomparable, in symbols: $\tilde{f}(x) \parallel \tilde{f}(y)$ resp. $x \parallel y$.

It remains to examine whether the set of minimal elements of such posets represents a reasonable and useful set of candidate solutions. For this purpose define

$$\begin{array}{lll} f^* = \min\{f \in \mathcal{F}\} & \text{with} & \mathcal{F} = \{f(x) : x \in S\} \text{ and} \\ \tilde{f}^* = \min\{\tilde{f} \in \tilde{\mathcal{F}}\} & \text{with} & \tilde{\mathcal{F}} = \{\tilde{f}(x) : x \in S\}\,. \end{array}$$

In general, $\tilde{f}^*$ and $\tilde{\mathcal{F}}$ are random objects. But since it is assumed that each element $x \in S$ is evaluated only once, one can hold the view that *each* element of S has been evaluated already *before* the EA is run such that the set $\tilde{\mathcal{F}}$ and the quantity $\tilde{f}^*$ are deterministic during the run of the EA. In this manner one obtains a unique partial order on $\tilde{\mathcal{F}}$ and on S for each run. The set of minimal elements is then given by

$$\begin{aligned} \tilde{\mathcal{F}}^* &= \{\tilde{f} \in \tilde{\mathcal{F}} \mid \nexists \tilde{f}' \in \tilde{\mathcal{F}} : \tilde{f}' \prec \tilde{f}\} \\ &= \{\tilde{f} \in \tilde{\mathcal{F}} \mid \tilde{f} \le \tilde{f}^* + 2\,a\}\,. \end{aligned} \tag{3}$$

Needless to say, it is reasonable to postulate that the noisy image $\tilde{f}(x^*)$ of an unperturbed optimal point $x^* \in S$ is contained in the set of minimal elements. As shown below, this requirement is fulfilled.

Theorem 1
For all $x^* \in S$ with $f(x^*) = f^*$ holds $\tilde{f}(x^*) \in \tilde{\mathcal{F}}^*$ regardless of the value of $a > 0$.
Proof: First notice the equivalence

$$\tilde{f}(x^*) \in \tilde{\mathcal{F}}^* \;\Leftrightarrow\; \tilde{f}(x^*) \le \tilde{f}^* + 2\,a$$

which is easily deduced from equation (3). Since the support of random variable Z is $[-a, a]$ the relation

$$\tilde{f}(x^*) = f(x^*) + Z = f^* + Z \le f^* + a \tag{4}$$

holds with probability 1. For the same reason one obtains $\tilde{f}^* > f^* - a \Leftrightarrow f^* < \tilde{f}^* + a$. Insertion in equation (4) leads to $\tilde{f}(x^*) \le f^* + a < (\tilde{f}^* + a) + a = \tilde{f}^* + 2\,a$. □

The next result offers an assessment of the 'solutions' contained in the set $\tilde{\mathcal{F}}^*$.

Theorem 2 $\max\{\tilde{f} \in \tilde{\mathcal{F}}^*\} \le f^* + 3\,a$.
Proof: Owing to equation (3) each element of $\tilde{\mathcal{F}}^*$ is upper bounded by $\tilde{f}^* + 2\,a$. Since the support of Z is $[-a, a]$ one obtains $\tilde{f}^* \le f^* + a$. Putting all together yields the desired inequality $\max\{\tilde{f} \in \tilde{\mathcal{F}}^*\} \le \tilde{f}^* + 2\,a \le f^* + 3\,a$. □

Notice that under this partial order the set of minimal elements of a given finite population P is determinable in linear time: Find the individual with the smallest perturbed fitness value $\tilde{f}^*$ in the population. This takes $\Theta(|P|)$ time. Each individual with fitness $\tilde{f} < \tilde{f}^* + 2\,a$ moves to the set of minimal elements. Since this takes $\Theta(|P|)$ time the entire run time is $\Theta(|P|)$.

3.3 The Base Algorithm

The pseudo code given in Figure 1 is taken from Rudolph (2001a). Notice that an individual $p \in P_t$ of a population at generation $t \geq 0$ gathers all quantities of interest, i.e., $p \in S \times \widetilde{\mathcal{F}} \times \ldots$. The partial ordering of the individuals is based on their noisy fitness values $\tilde{f} \in \widetilde{\mathcal{F}}$ of course. The expression $Q = \mathtt{offspring}(P_t)$ in phase 1 encapsulates the operation of generating λ new individuals from the current population P_t of size μ with $\mu \leq \lambda$ at generation $t \geq 0$.

```
initialize P_0; set t = 0
repeat
  (* PHASE 1 *)
  Q = offspring(P_t)
  Q* = M(Q, ⪯)
  Q = Q \ Q*
  P' = Q' = ∅
  (* PHASE 2 *)
  for each q ∈ Q*:
    D(q) = {p ∈ P_t : q ≺ p}
    if D(q) ≠ ∅ then
      P_t = P_t \ D(q)
      P' = P' ∪ {q}
      Q* = Q* \ {q}
    endif
    if D(q) = ∅ ∧ q ∥ p for all p ∈ P_t then
      Q' = Q' ∪ {q}
      Q* = Q* \ {q}
    endif
  endfor
  (* PHASE 3 *)
  P_{t+1} = P_t ∪ P'
  if |P_{t+1}| < μ then
    fill P_{t+1} with elements from:
      1. Q'
      2. Q*
      3. Q
    until P_{t+1} = μ
  endif
  t = t + 1
until stopping criterion fulfilled
```

Figure 1: Pseudo code of the evolutionary algorithm with partially ordered fitness.

3.4 The Theoretical Property Inherited

Since the base algorithms' properties are valid for arbitrary partially ordered fitness sets, any instantiation inherits the theoretical properties from the general case (Rudolph 2001a).

Theorem 3 *Let the search space of the base algorithm in Figure 1 be finite and the partial order of the fitness set be as described in Section 3.2. If every collection of offspring can be generated from an arbitrary collection of parents with some positive minimum probability, then the entire population will enter the set $\widetilde{\mathcal{F}}^*$ after a finite number of generations with probability* 1 *and stays there forever.* □

As we know from Theorems 1 and 2 the set of minimal elements $\widetilde{\mathcal{F}}^*$ contains the noisy version of the global minimum and each member of $\widetilde{\mathcal{F}}^*$ is at most $3\,a$ away from the true minimum f^*. In this sense, we may call $\widetilde{\mathcal{F}}^*$ also the set of ε-optimal solutions with $\varepsilon = 3\,a$.

3.5 The Virtue of Resampling Revisited

A $(3\,a)$-optimal solution may be sufficient or may not. For the latter case one should look for a method of decreasing this bound. Here we use the technique of resampling that is common practice in EAs with noisy fitness values. Usually each point/individual is sampled several times and the thereby obtained noisy fitness values are averaged. This makes the estimator of the true fitness value more and more reliable by reducing its variance by a factor of $1/n$ in case of n samples. Nevertheless, the estimator's value is afflicted with some uncertainty and every statement concerning its deviation from the true value may be false with some probability. For example, in case of Gaussian noise with known variance σ^2 we only can say that the true value f is in the interval

$$\left[\bar{f} - \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1+\gamma}{2}\right), \bar{f} + \frac{\sigma}{\sqrt{n}}\,\Phi^{-1}\left(\frac{1+\gamma}{2}\right)\right]$$

with probability $\gamma > 0$. Here, $\bar{f}$ denotes the average of n fitness samples and $\Phi^{-1}(\cdot)$ is the inverse of the cumulative distribution function (c.d.f.) of the standard normal distribution.

A moment reflection reveals that the partial order introduced in Section 3.2 is actually a partial ordering of confidence intervals. Since the support of the noise is bounded the confidence intervals for the true value have a confidence level of $\gamma = 1$ in contrast to $\gamma < 1$ for Gaussian noise. But here we don't use resampling for averaging; rather, the information gained from the additional samples is used to narrow the uncertainty interval containing the true fitness value.

Let $\tilde{f}_n = f + Z_n$ denote the nth sample of the noisy fitness function at a certain point in the search space. The first sample $\tilde{f}_1 = f + Z_1$ leads to the initial confidence interval $[\tilde{f}_1 - a, \tilde{f}_1 + a]$ for the true value f. Since each sample leads to a different confidence interval in general and f must be contained in each of these intervals we immediately obtain

$$\begin{aligned} f &\in \bigcap_{k=1}^{n} [\,\tilde{f}_k - a, \tilde{f}_k + a\,] \\ &= \left[\max_{k\leq n}\{\tilde{f}_k\} - a, \min_{k\leq n}\{\tilde{f}_k\} + a\right] \\ &= \left[f + \max_{k\leq n}\{Z_k\} - a, f + \min_{k\leq n}\{Z_k\} + a\right] \\ &= [f, f] + [\,Z_{n:n} - a, Z_{1:n} + a\,] \end{aligned} \tag{5}$$

where $Z_{k:n}$ denotes the kth smallest outcome of n samples in total. Thus, after n samples one knows for sure that the true value f is somewhere in the interval given in equation (5). The uncertainty interval $[\, Z_{n:n} - a, Z_{1:n} + a\,]$ shrinks to $[0, 0]$ for $n \to \infty$. The speed of narrowing can be determined as follows: Let $L_n = Z_{n:n} - a$ and $R_n = Z_{1:n} + a$. Then $|[L_n, R_n]|/(2\,a)$ is the relative size of the uncertainty or incomparability interval $[L_n, R_n]$ after n samples and the probability that it is then still larger than $100\,\varepsilon$ percent of its initial size is given by

$$\begin{aligned} \mathsf{P}\left\{ \frac{|[L_n, R_n]|}{2\,a} > \varepsilon \right\} &= \mathsf{P}\left\{ \frac{R_n - L_n}{2\,a} > \varepsilon \right\} \\ &= \mathsf{P}\left\{ 1 - \frac{Z_{n:n} - Z_{1:n}}{2\,a} > \varepsilon \right\} \\ &= \mathsf{P}\{\, 1 - W_n > \varepsilon \,\} \\ &= \mathsf{P}\{\, W_n < 1 - \varepsilon \,\} \end{aligned}$$

where $W_n = B_{n:n} - B_{1:n}$ is known as the *sample range* of n samples of random variable B. Here it is assumed that B has support $[0, 1]$ such that the noise $Z = a\,(2\,B - 1)$ has support $[-a, a]$. The distribution of the sample range can be determined via

$$\mathsf{P}\{\, W_n \le w \,\} = n \int_{-\infty}^{\infty} [F_B(x + w) - F_B(x)]^{n-1}\, f(x)\, dx$$

(see e.g. Arnold et al. 1992, p. 31) such that

$$\begin{aligned} \mathsf{P}\{\, W_n < 1 - \varepsilon \,\} &= n\,(1 - \varepsilon)^{n-1} - (n-1)\,(1-\varepsilon)^n \\ &\le n\,(1-\varepsilon)^{n-1} \end{aligned}$$

if B is the uniform distribution on $[0, 1]$. In this case the relative size of the uncertainty interval after n samples is given by $1 - W_n$ with $\mathsf{E}[\, 1 - W_n \,] = 2/(n+1)$ and $\mathsf{V}[\, 1 - W_n \,] \le 2/n^2$. If B is a symmetric Beta random variable with probability density function

$$f_B(x) = \frac{\Gamma(2\,b)}{\Gamma(b)^2}\, x^{b-1}\,(1-x)^{b-1} \cdot 1_{[0,1]}(x) \tag{6}$$

then $\mathsf{E}[\, 1 - W_n \,] \approx 2\,n^{-1/b}$. Thus, the closer B should resemble Gaussian noise ($b \to \infty$) the slower is the narrowing of the uncertainty or incomparability interval.

4 First Numerical Experiments

4.1 Instantiation of the Base Algorithm

The search space S is a finite subset of the d-dimensional set of integers with box constraints, i.e., $S = \{x \in \mathbb{Z}^d : \hat{a} \le x \le \hat{b}\}$ with $\hat{a}, \hat{b} \in \mathbb{Z}^d$. An individual $(x, e, \ell, r) \in \mathbb{Z}^d \times \mathbb{R} \times \mathbb{R} \times \mathbb{R}$ is represented by a d-tuple of integers x (the chromosomes), an 'expected step length' e and the bounds ℓ, r of the confidence interval that is needed for the comparison of the individuals according to equation (2). Next we detail the mutation and recombination operators.

4.1.1 Mutation

Mutations of the chromosomes obey a bilateral geometrical distribution on the integers (Rudolph 1994) that is truncated at the box constraints. It suffices to consider the one-dimensional case since each random entry of the mutation vector is drawn independently. The probability density function of the bilateral geometrically distributed (not truncated) random variable M_i is

$$\mathsf{P}\{\, M_i = k \,\} = \frac{p}{2-p}\,(1-p)^{|k|} = \frac{1-q}{1+q}\, q^{|k|}$$

for $k \in \mathbb{Z}$ and $p = 1 - q \in (0, 1) \subset \mathbb{R}$. The distribution is unimodal with mode at zero. Needless to say, this remains true if the distribution is truncated at $\hat{a}_i$ to the left and $\hat{b}_i$ to the right.

In order to avoid the generation of lethal (infeasible) mutations it is necessary to truncate the distribution of M_i such that for each feasible individual $X \in [\hat{a}, \hat{b}] \cap \mathbb{Z}^d$ the mutation $\tilde{M}_i$ satisfies $X_i + \tilde{M}_i \in [\hat{a}_i, \hat{b}_i] \cap \mathbb{Z}$. Therefore we must recalculate the distribution of every $\tilde{M}_i$ prior to each mutation. Since we like to have the mode of the distribution of $X_i + \tilde{M}_i$ at X_i (such that $\tilde{M}_i$ has the mode at zero) we set the truncation values to

$$a_i = \hat{a}_i - X_i \quad \text{and} \quad b_i = \hat{b}_i - X_i$$

which leads to the relation $a_i \le 0 \le b_i$ for all $i = 1, \ldots, d$. Moreover, we postulate that $a_i < b_i$. Hereinafter, we omit the index i while elaborating the one-dimensional case.

If random variable M with cumulative distribution function (c.d.f.) $F(\cdot)$ is truncated at a and b then the c.d.f. $\tilde{F}(\cdot)$ of the truncated random variable $\tilde{M}$ is given by

$$\tilde{F}(k) = \frac{F(k) - F(a-1)}{F(b) - F(a-1)} \tag{7}$$

for $a \le k \le b$, $\tilde{F}(k) = 0$ for $k < a$ and $\tilde{F}(k) = 1$ for $k > b$. We distinguish three cases:

I. $0 = a \le k \le b$ and $b > 0$
Since in this case $F(k) = F(-1) + (1 - q^{k+1})/(1+q)$ we immediately obtain

$$\tilde{F}(k) = \frac{1 - q^{k+1}}{1 - q^{b+1}}$$

by insertion in equation (7) and the rearrangement of the equation $\tilde{F}(K) = U$ leads to

$$K = \frac{\log(1 - U\,[1 - q^{b+1}])}{\log\, q} - 1 \tag{8}$$

where U is a uniformly distributed random number in $(0, 1]$.

II. $a \le k \le b = 0$ and $a < 0$
Similar to the previous case we obtain

$$\tilde{F}(k) = \frac{q^{-k} - q^{1-a}}{1 - q^{1-a}}$$

and hence

$$K = -\frac{\log(q^{1-a} + U\,[1 - q^{1-a}])}{\log\, q} \qquad (9)$$

III. $a < 0 < b$
Since

$$F(k) = \begin{cases} \dfrac{q^{|k|}}{1+q} & \text{if } k < 0, \\[2ex] 1 - \dfrac{q^{k+1}}{1+q} & \text{if } k \geq 0 \end{cases}$$

the c.d.f. of $\tilde{M}$ is

$$\tilde{F}(k) = \begin{cases} \dfrac{q^{-k} - q^{1-a}}{1 + q\,(1 - q^b - q^{-a})} & \text{if } a \leq k < 0 \\[2ex] \dfrac{1 + q\,(1 - q^k - q^{-a})}{1 + q\,(1 - q^b - q^{-a})} & \text{if } 0 \leq k \leq b. \end{cases}$$

Let $c = 1 + q\,(1 - q^b - q^{-a})$. If $U < \tilde{F}(0) = (1 - q^{1-a})/c$ then

$$K = -\frac{\log(c\,U + q^{1-a})}{\log\, q} \qquad (10)$$

else

$$K = \frac{\log(1 - q^{-a} + [1 - c\,U]/q)}{\log\, q} \qquad (11)$$

Thus, for each entry $\tilde{M}_i$ we draw a standard uniform random number U and use one of the equations (8)-(11) depending on the values of the bounds a and b. This yields the random number K and we finally set $\tilde{M}_i = \lfloor K \rfloor + 1$.

But before we mutate the decision vector x according to the method explained above, we have to mutate the 'expected step length' e which leads to the parameter q used in equations (8)-(11). This mechanism is also borrowed from the not truncated case (Rudolph 1994). If M is the d-dimensional (not truncated) mutation vector then the expected length in ℓ_1-norm is

$$e = \frac{2\,d\,q}{(1-q)\,(1+q)} \Leftrightarrow q = \frac{e/d}{(1 + (e/d)^2)^{1/2} + 1} \qquad (12)$$

with a unique expression for parameter q. In principle, we would like to control (or self-adapt) parameter q in the interval $(0, 1)$. But a satisfactory method for doing this is not known yet. The remedy is as follows: We simply self-adapt the expected step length $e \in \mathbb{R}^+$ according to the same method known for evolutionary strategies and programming: Let $\tau = 1/\sqrt{9\,d}$ and set the initial value of the step length to

$$e = \min\{\hat{b}_i - \hat{a}_i : i = 1, \ldots, d\}$$

for each individual. When an individual is mutated then at first we mutate e via

$$e_{\text{new}} = e_{\text{old}} \cdot L$$

where L is a log-logistic random variable that can be obtained from a standard uniform random variable U via

$$L = \left(\frac{U}{1-U}\right)^{\tau}.$$

If e gets smaller than $1/d$ then we set $e = 1/d$. After the mutation of e is finished we determine the associated value of q from equation (12) before we draw the mutation vector $\tilde{M}$.

It is clear that e is only a caricature of the true expected step length of the truncated mutation vector $\tilde{M}$, but this mechanism seems to work despite this theoretical deficiency.

4.1.2 Recombination

The recombination of two chromosomes $x, y \in \mathbb{Z}^d$ is realized by uniform crossover, i.e., offspring z is assembled via $z_i = \xi_i\, x_i + (1 - \xi_i)\, y_i$ for $i = 1, \ldots, d$ where the ξ_i are independent and identically distributed Bernoulli random numbers with $\mathsf{P}\{\,\xi_i = 0\,\} = \mathsf{P}\{\,\xi_i = 1\,\} = 1/2$. The real-valued chromosomes e_x and e_y for self-adaptation are averaged, i.e., the offspring inherits the value $(e_x + e_y)/2$.

4.2 Realization of the Noise Generator

The noise used here is represented by a Beta random $Z = a\,(2\,B - 1)$. A Beta random variable B on $(0, 1)$ with parameter $b \in \mathbb{N}$ and probability density function as given in equation (6) can be generated via $B = G_1/(G_1 + G_2)$ where

$$G_1 = -\log\left(\prod_{i=1}^{b} U_i\right) \quad \text{and} \quad G_2 = -\log\left(\prod_{j=1}^{b} U_j\right)$$

with uniformly distributed random numbers $U_i, U_j \in (0, 1) \subset \mathbb{R}$.

4.3 Preliminary Results

The search space is $S = \mathbb{Z}^{30} \cap [-1000, 1000]^{30}$ for the test function

$$f(x) = \sum_{i=1}^{30} \left(\frac{x_i}{10}\right)^2$$

with optimal solution $x^* = (0, \ldots, 0)$ and $f(x^*) = 0$.

Figure 2 shows a typical run for a population size of $\mu = \lambda = 50$ and a support bound $a = 1$ for the uniformly distributed noise ($b = 1$), whereas the support bound was increased to $a = 5$ in Figure 3 and $a = 10$ in Figure 4.

As can be seen in all three cases the number of surviving parents rapidly increases at about the maximum size as soon as the population enters the set of $(3\,a)$-optimal solutions. It appears plausible that this happens the earlier the larger is the noise bound a, but it must be noted that there is no statistical support for this conjecture at the moment.

Nevertheless, we dare to use the number of surviving parents as an indicator for the event of entering the $(3\,a)$-optimal

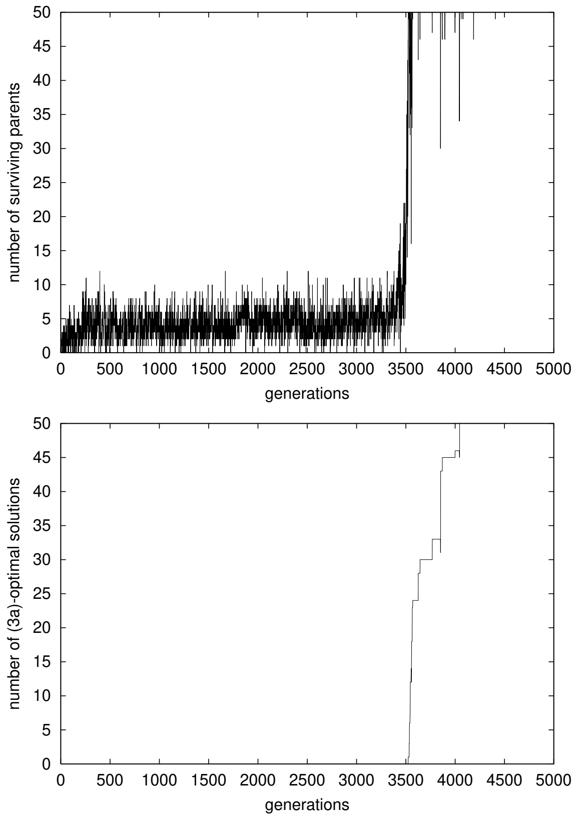

Figure 2: Number of surviving parents and $(3a)$-optimal solutions for noise bound $a = 1$.

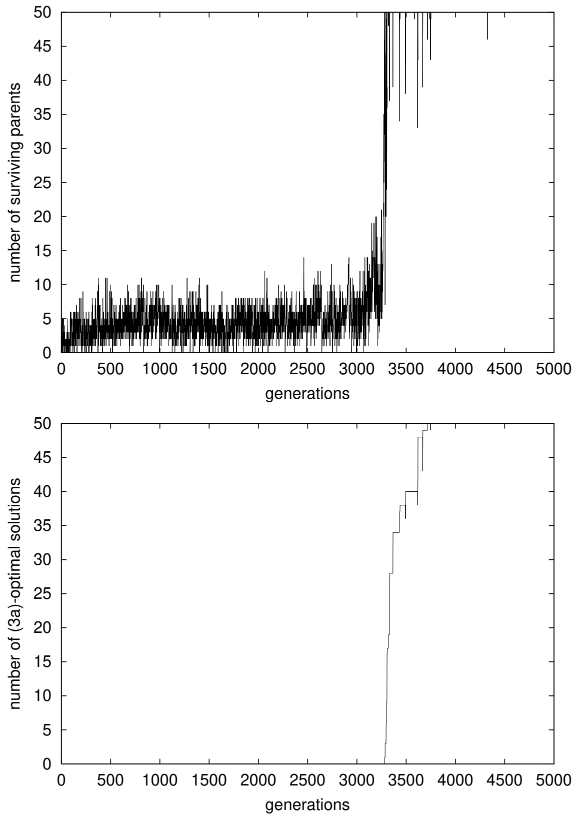

Figure 3: Number of surviving parents and $(3a)$-optimal solutions for noise bound $a = 5$.

set. When this happens then the individuals cannot be compared due to the size of the noise interval. Therefore, the resampling of the fitness values should begin right now. Many rules for the indicator mechanism are possible.

Here, we introduce a width parameter ω that is initially set to $\omega = 2\,a$. If two individuals are incomparable and at least one of them has a confidence interval larger than ω, then the individual with the largest interval is re-evaluated and its interval bounds are updated. This is repeated as long as the individuals are incomparable or both confidence intervals are smaller than ω. The decrease of ω by 10 % is triggered as soon as 90 % of the parents have survived, provided that $\omega \geq 1/100$. The EA is stopped as soon as $\omega < 1/100$ and no parent has been replaced in the last selection process.

Of course, these parameters are chosen arbitrarily for the next experiment—other values may yield far better results. The identification of "good" parameter settings is not the goal here; rather, we like to gain first insights concerning the behavior of this EA.

For this purpose the noise bound is set to $a = 10$ and the EA is run. Figure 5 shows some characteristic quantities recorded during a typical run of the EA.

At about generation 2000 the population begins to enter the set of 30-optimal solutions. No re-evaluation of the individuals was triggered until now. The number of surviving parents oscillates considerably while the width parameter is continuously decreased. This leads to a rapid increase of the number of re-evaluations along with a significant improvement of the solutions as can be seen from the number of parents below certain (normally unavailable) true fitness values.

The number of re-evaluations, however, is much too large for practical use. This might be caused by an unlucky choice of the EA parameters. In the next experiment the width threshold was chosen a magnitude larger, namely, $\omega = 1/10$. Table 1 summarizes the results obtained from 50 independent runs. The number of re-evaluations is considerably smaller now, but the optimal true solution was never contained in the final population, in contrast to some other tests with $\omega = 1/100$. Needless to say, some parameter studies are necessary to find a useful parametrization of this EA.

	mean	std. dev.	skew.
generations	3231	740.6	0.46
re-evaluations	2.2×10^6	6.5×10^5	0.85
best f of last pop.	0.13	0.031	0.17
worst f of last pop.	0.25	0.039	0.46

Table 1: Summary of results for 50 runs with $\omega = 1/10$.

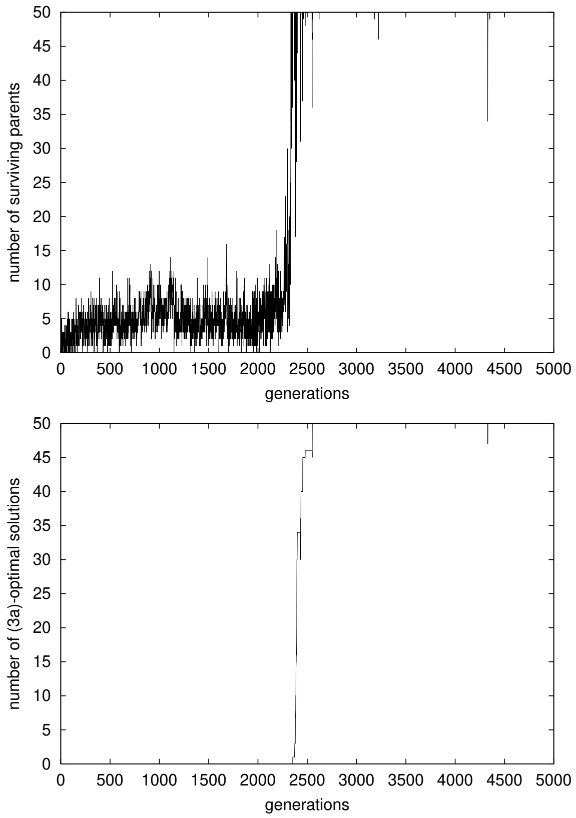

Figure 4: Number of surviving parents and $(3a)$-optimal solutions for noise bound $a = 10$.

5 Conclusions

The limit theory for EAs under partially ordered fitness sets immediately delivers a limit theory for EAs tackling problems with multiple objectives, noisy fitness functions, and others for free. See Rudolph (2001b) for a collection of fields where this approach can come into operation. Interval-valued fitness functions (Teich 2001) or even fuzzy fitness functions are obvious candidates.

In spite of these good news we must not conceal that there are problems, too. First, the proof of Theorem 3 is not designed for covering the version of the EA with resampling. In Theorem 3 it is assumed that the partial order remains fixed during the search. In case of resampling, however, the partial order gets more and more total in the course of the search.

Second, the rules for triggering width reduction and resampling are completely unexplored currently. The rules used here turned out to be computationally expensive. Thus, many parameter studies are necessary to come up with a reasonable heuristic finally.

Acknowledgments

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

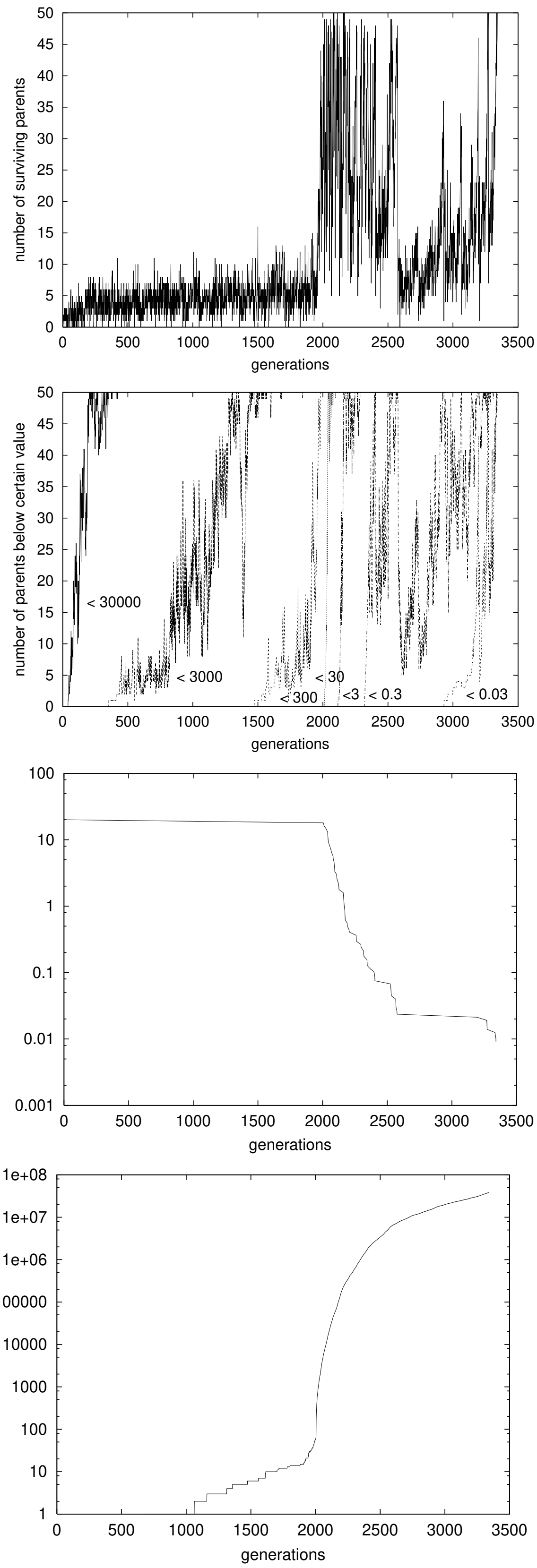

Figure 5: From top to bottom: Number of surviving parents, number of parents below certain fitness thresholds, decay of width threshold and cumulated number of re-evaluations.

Bibliography

B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja (1992). *A First Course in Order Statistics*. New York: Wiley.

H.-G. Beyer (2000). Evolutionary algorithms in noisy environments: Theoretical issues and guidelines for practice. *Computer Methods in Applied Mechanics and Engineering 186*(2-4), 239–267.

M. Evans, N. Hastings, and B. Peacock (1993). *Statistical Distributions* (2nd ed.). New York: Wiley.

P. C. Fishburn (1985). *Interval Orders and Interval Graphs: A Study of Partially Ordered Sets*. New York: Wiley.

J. Guddat, F. Guerra Vasquez, K. Tammer, and K. Wendler (1985). *Multiobjective and Stochastic Optimization Based on Parametric Optimization*. Berlin: Akademie-Verlag.

G. Rudolph (1994). An evolutionary algorithm for integer programming. In Y. Davidor, H.-P. Schwefel, and R. Männer (Eds.), *Parallel Problem Solving From Nature, 3*, pp. 139–148. Berlin and Heidelberg: Springer.

G. Rudolph (1998). Evolutionary search for minimal elements in partially ordered finite sets. In V. W. Porto, N. Saravanan, D. Waagen, and A. E. Eiben (Eds.), *Evolutionary Programming VII, Proceedings of the 7th Annual Conference on Evolutionary Programming*, pp. 345–353. Berlin: Springer.

G. Rudolph (2001a). Evolutionary search under partially ordered fitness sets. In *Proceedings of the International Symposium on Information Science Innovations in Engineering of Natural and Artificial Intelligent Systems (ENAIS 2001)*. ICSC Academic Press. In print.

G. Rudolph (2001b). Some theoretical properties of evolutionary algorithms under partially ordered fitness values. In C. Fabian (Ed.), *Proceedings of the Evolutionary Algorithms Workshop (EAW 2001)*. Bucharest, Romania: INFOREC Printing House. In print.

G. Rudolph and A. Agapie (2000). Convergence properties of some multi-objective evolutionary algorithms. In A. Zalzala, C. Fonseca, J.-H. Kim, A. Smith, and X. Yao (Eds.), *Proceedings of the 2000 Congress on Evolutionary Computation (CEC 2000), Vol. 2*, pp. 1010–1016. Piscataway (NJ): IEEE Press.

J. Teich (2001). Pareto-front exploration with uncertain objectives. In E. Zitzler, K. Deb, L. Thiele, C.A. Coello Coello, and D. Corne (Eds.), *Proceedings of the 1st International Conference on Evolutionary Multi-Criterion Optimization (EMO 2001)*, pp. 314–328. Berlin and Heidelberg: Springer.