Self-Adaptation and Global Convergence: A Counter-Example

Günter Rudolph*

January 14, 1999

Abstract

The self-adaptation of the mutation distribution is a distinguishing feature of evolutionary algorithms that optimize over continuous variables. It is widely recognized that self-adaptation accelerates the search for optima and enhances the ability to locate optima accurately, but it is generally unclear whether these optima are global ones or not. Here, it is proven that the probability of convergence to the global optimum is less than one in general—even if the objective function is continuous.

1 Introduction

The self-adaptation of the mutation distribution in evolution strategies (ES) was introduced by Rechenberg (1973). Here, self-adaptation means that the control parameters of the mutation distribution are evolved by the evolutionary algorithm internally, rather than being predetermined by some exogenously given schedule. A simple version of this mechanism was the so-called $1/5$-success rule of the $(1+1)$-ES which worked as follows: If the relative frequency of successful (i.e., improving) mutations within some prescribed period of time is larger than $1/5$, then the step size control parameter (mostly the variance of the mutation distribution) is increased by some factor, whereas it is decreased if the relative frequency of successful mutations is smaller than $1/5$. This mechanism was modified by Schwefel (1977) who replaced the prescribed factor by a lognormally distributed random variable and added the control parameter to the genome of each individual. As a consequence, the adjustment of the control parameter implicitly results from the competition among the individuals. Similar methods were independently proposed in evolutionary programming by Fogel (1992; 1995). Needless to say, self-adaptation is not limited to the control of mutation distributions. Further fields of application may be found in recent surveys (Hinterding et al. 1997; Bäck 1998).

Although it is widely recognized that self-adaptation of the mutation distribution accelerates the search for optima and enhances the ability to locate optima accurately, the theoretical underpinnings of this mechanism are essentially unexplored. For example, it is generally unclear whether the optima found are global ones or not. In the case of convex objective functions (to be minimized), Rappl (1984) has given a proof of exponentially fast convergence for a stochastic algorithm resembling a $(1+1)$-ES with $1/5$-success rule whereas Beyer (1995) examined also other evolutionary algorithms and self-adaptation rules.

As for non-convex objective functions, Rudolph (1997) has shown that every self-adaptation method leads to global convergence for objective functions with bounded lower level sets of non-zero measure provided that the selection method uses elitism and that the mutation distribution remains "strictly covering" under

*Department of Computer Science, University of Dortmund, Germany.

the self-adaptation rule (cf. Theorem 6.14, p. 204). A strictly covering mutation distribution was defined as follows:

> "Let $\hat{Z}$ be a random vector with support $\mathbb{R}^n$ and $S = \text{diag}(s_1, \ldots, s_n)$ be a diagonal matrix with $\det S = 1$ and $s_i \geq s_{min} > 0$. A mutation distribution F_Z will be termed *strictly covering* if Z can be generated via $Z = \sigma\, T\, S\, \hat{Z}$ for arbitrary orthogonal matrices T and where σ is allowed to vary in a fixed compact subset of the positive real numbers."

The main idea of the proof is sketched below: Since the range of the mutation control parameter σ is restricted to a fixed compact subset of the positive real numbers, there exists a positive lower as well as upper bound for legal values of σ. This implies a positive minimum probability for hitting some neighborhood of the global solution and therefore the event of an actual transition to this neighborhood will occur with probability 1 within a finite number of mutations. Since this result holds for arbitrary neighborhoods one immediately obtains global convergence in probability and eventually complete convergence to the global optimum because of the exponential increase of the probability for this event.
The crucial point of the proof is the assumption of the positive lower bound for the control parameter σ. Commonly used self-adaptation mechanisms do not obey this condition. This observation was the starting point of the work to be presented shortly. Here, it is examined whether global convergence can still be guaranteed if the assumption of a positive lower bound of σ is dropped. This is done by the construction of a scenario under which an evolutionary algorithm with self-adaptation *provably* fails. The scenario and some auxiliary results are presented in Section 2 whereas the analysis is given in Section 3. The extension of the scenario to related evolutionary algorithms can be found in Section 4. Section 5 finishes the paper by a discussion that is intended to clarify the implications of this negative result in regard to the practical use of evolutionary algorithms with self-adaptation rules.

2 Description of the Constructed Scenario

2.1 Algorithm

Let $g : \mathbb{R}^n \to \mathbb{R}$ be the objective function to be minimized and set $X_0 = x_0 \in \mathbb{R}^n$, $\ell_0 > 0$. Consider the Markovian process $(X_k, \ell_k)_{k \geq 0}$ generated by the stochastic algorithm

$$X_{k+1} = \begin{cases} X_k + \ell_k\, Z_k & \text{, if } g(X_k + \ell_k\, Z_k) < g(X_k) \\ X_k & \text{, otherwise} \end{cases} \tag{1}$$

$$\ell_{k+1} = \begin{cases} \gamma_1\, \ell_k & \text{, if } g(X_k + \ell_k\, Z_k) < g(X_k) \\ \gamma_2\, \ell_k & \text{, otherwise} \end{cases} \tag{2}$$

where $\gamma_1 > 1$, $\gamma_2 \in (0, 1)$, and $(Z_k : k \geq 0)$ denotes a sequence of independent and identically distributed random vectors with spherically symmetric distribution around the zero vector and support $\mathbb{R}^n$. Thus, whenever there is a successful (i.e., improving) mutation, the step length control parameter ℓ_k is increased and decreased otherwise. This algorithm does not exactly match a $(1+1)$-ES with self-adapting step size control, but the analysis of this method can be modified easily to a broader class of evolutionary algorithms (see Section 4).

2.2 Test Problem

When a stochastic optimization algorithm is said to exhibit *global convergence* for a certain class of objective functions, then it is meant that the stochastic sequence of candidate solutions generated by the algorithm converges (in some stochastic mode) to the global optimum for *every* member of the problem class *regardless of the initial solution.* In order to falsify the property of global convergence for the entire problem class it is sufficient to prove a negative result for a single member of the problem class and a specific starting point. Here, the problem class is the set of continuous real-valued functions with search space $\mathbb{R}$. An instance of this problem class is the function $g : \mathbb{R} \to \mathbb{R}$ described next. Let $\epsilon > 0$, $\delta > 0$, $z \ge 6\,\delta > 0$ and

$$g(x) = \begin{cases} \frac{3}{4}\,\epsilon\, x^2/\delta^2 + \epsilon & \text{, if} \quad x \le 2\,\delta \\ 4\,\epsilon & \text{, if} \quad 2\,\delta < x < z - 2\,\delta \\ \epsilon\,(x-z)^2/\delta^2 & \text{, if} \quad x \ge z - 2\,\delta\ . \end{cases}$$

Figure 1 shows a qualitative plot of the objective function $g(\cdot)$. Its global optimum is located at $x^* = z$ with value $g(z) = 0$. The starting point of the algorithm described in equations (1) and (2) is $X_0 = 0$ whereas $\ell_0 > 0$ may be chosen arbitrarily.

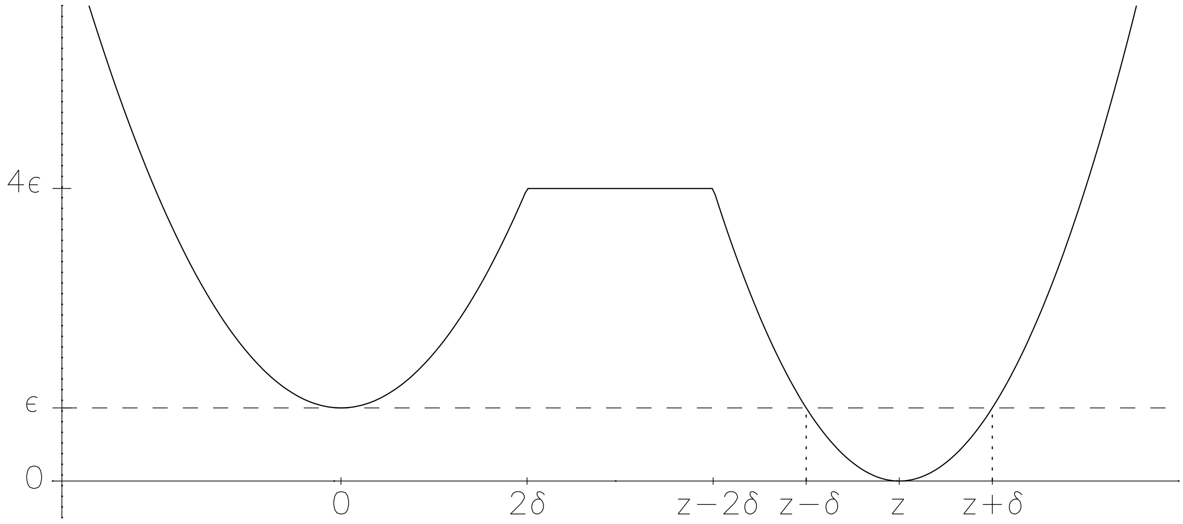

Figure 1: A qualitative plot of the objective function $g(\cdot)$.

With these initial values the algorithm only accepts a solution different from 0 if the outcome of the random variable $\ell_k\, Z_k$ is within the range $(z - \delta, z + \delta)$. It is clear that the occurrence of such an event in the course of an infinite sequence of trials with probability 1 is a *necessary condition* for global convergence. If the probability of this event is less than 1, then the algorithm will never enter the set $(z - \delta, z + \delta)$ with positive probability. As a consequence, stochastic convergence to z is precluded in this case. In order to prove when the necessary condition is violated, sufficient criteria for success and failure are developed next.

2.3 Criteria for Success or Failure

Suppose that the cumulative distribution function F_Z of the symmetrical random variable Z with zero median is differentiable. Owing to the mean value theorem the probability of a transition from point 0 to the set $(z - \delta, z + \delta) \subset \mathbb{R}$ can be derived via

$$\mathsf{P}\{\, 0 \to (z - \delta, z + \delta)\,\} = \mathsf{P}\{\, z - \delta < Z < z + \delta\,\} = F_Z(z + \delta) - F_Z(z - \delta) = 2\,\delta\, f_Z(z - \delta + \theta \cdot 2\,\delta)$$

where $0 < \theta < 1$ and $f_Z(x) = \frac{d}{dx} F_Z(x)$. Since $z > \delta > 0$ the transition probability can be bracketed as follows:

$$2\,\delta\, f_Z(z - \delta) \;\leq\; \mathsf{P}\{\, 0 \to (z - \delta, z + \delta)\,\} \;\leq\; 2\,\delta\, f_Z(z + \delta)\,. \tag{3}$$

Inequality (3) is valid for arbitrary distributions of random variable Z. In the course of the evolution, however, the distribution of Z will not remain constant in general. This fact will be taken into account by adding the subscript k (the iteration counter) to Z hereinafter.

2.3.1 Criterion for Successful Crossing

The probability of crossing the hill between the local and global optimum at step $k \geq 1$ is

$$p_k = \mathsf{P}\{\, 0 \to (z - \delta, z + \delta) \text{ at step } k\,\} \geq q_k = 2\,\delta\, f_{Z_k}(z - \delta) \tag{4}$$

where the lower bound q_k results from inequality (3). It follows that the probability of a transition to $(z - \delta, z + \delta)$ within $t \geq 0$ trials is

$$1 - \prod_{k=1}^{t} (1 - p_k) \geq 1 - \prod_{k=1}^{t} (1 - q_k)\,.$$

Therefore the sufficient criterion for a successful crossing is simply

$$\prod_{k=1}^{t} (1 - q_k) \to 0 \quad \text{or, equivalently,} \quad \sum_{k=1}^{t} \log\left(\frac{1}{1 - q_k}\right) \to \infty \tag{5}$$

as $t \to \infty$. If criterion (5) is fulfilled, then the evolutionary algorithm will jump to $(z - \delta, z + \delta)$ within a finite number of steps with probability one. This does not imply that the EA will converge to the global optimum at $x^* = z$, but the necessary condition for this property would be fulfilled.

2.3.2 Criterion for Unsuccessful Crossing

In contrast to the pessimistic point of view in the preceding criterion, one has to assume an optimistic point of view now. Thanks to inequality (3) one obtains

$$p_k = \mathsf{P}\{\, 0 \to (z - \delta, z + \delta) \text{ at step } k\,\} \leq q_k = 2\,\delta\, f_{Z_k}(z + \delta)\,. \tag{6}$$

As a consequence, the upper bound for the probability of a transition to $(z - \delta, z + \delta)$ within $t \geq 0$ trials is given by

$$1 - \prod_{k=1}^{t} (1 - p_k) \leq 1 - \prod_{k=1}^{t} (1 - q_k)\,.$$

If the probability on the right hand side above is smaller than 1 in the limit, then the transition to $(z - \delta, z + \delta)$ is not guaranteed. In other words, it may happen with positive probability that the EA never enters the set $(z - \delta, z + \delta)$. It is clear that such an event precludes the convergence to the global optimum at $x^* = z \in (z - \delta, z + \delta)$. Thus, the sufficient criterion of failure is simply

$$\prod_{k=1}^{\infty} (1 - q_k) > 0 \quad \text{or, equivalently,} \quad \sum_{k=1}^{\infty} \log\left(\frac{1}{1 - q_k}\right) < \infty\,. \tag{7}$$

3 Analysis

Suppose that the algorithm described in equations (1) and (2) is used to minimize the objective function $g(\cdot)$ of the preceding section. If the starting point is $X_0 = 0$ then equations (1) and (2) reduce to $X_k = 0$ and $\ell_k = \ell_0 \gamma^k$ with $\gamma \equiv \gamma_2 \in (0,1)$ as long as a transition to the set $(z-\delta, z+\delta)$ has not been occurred. As a consequence, the transition probabilities of the associated Markov process prior to a successful crossing of the hill can be bounded via inequality (3). Before explicit mutation distributions, namely Gaussian as well as Cauchy distributions, are inserted in this inequality, two simple inequalities are proven for later use.

Lemma 1 *If* $x \in (0,1)$ *then*

$$x \;<\; \log\left(\frac{1}{1-x}\right) \;<\; \frac{x}{1-x}\,.$$

Proof:
Consider the series expansion

$$\log\left(\frac{1}{1-x}\right) \;=\; -\log(1-x) \;=\; \sum_{i=1}^{\infty} \frac{x^i}{i}$$

which is valid for $|x| < 1$. This leads immediately to the inequalities

$$0 \;<\; x \;<\; \sum_{i=1}^{\infty} \frac{x^i}{i} \;<\; \sum_{i=1}^{\infty} x^i \;=\; \sum_{i=0}^{\infty} x^i - 1 \;=\; \frac{1}{1-x} - 1 \;=\; \frac{x}{1-x}$$

and the proof is completed. □

3.1 Gaussian Mutations

Suppose that random variable Z possesses Gaussian distribution with probability density function

$$f_Z(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$$

where σ will play the role of the step length control parameter ℓ. It will now be shown that a successful crossing of the hill between the local and global optimum is not guaranteed. For this purpose, the probability density function $f_Z(\cdot)$ is inserted into inequality (6) yielding

$$p_k \le q_k = \delta \sqrt{\frac{2}{\pi}}\, \frac{1}{\sigma_k} \exp\left(-\frac{(z+\delta)^2}{2\sigma_k^2}\right) = A\,\eta_k \exp(-B\,\eta_k^2) \tag{8}$$

where $A = \delta\,(2/\pi)^{1/2}$, $B = (z+\delta)^2/2$, and $\eta_k = 1/\sigma_k$. Lemma 1 leads to

$$\sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) \;<\; \sum_{k=1}^{\infty} \frac{q_k}{1-q_k} \;\le\; \frac{1}{1-q_1} \sum_{k=1}^{\infty} q_k\,. \tag{9}$$

Let $\eta_k = \eta_0\,\beta^k$ with $\beta = 1/\gamma > 1$. Insertion of the rightmost expression of equation (8) into the rightmost expression of equation (9) yields

$$\frac{1}{1-q_1} \sum_{k=1}^{\infty} q_k = \frac{A}{1-q_1} \sum_{k=1}^{\infty} \frac{\eta_k}{\exp(B\,\eta_k^2)} = \frac{A\,\eta_0}{1-q_1} \sum_{k=1}^{\infty} \frac{\beta^k}{\exp(B\,\eta_0^2\,\beta^{2k})} = \frac{A\,\eta_0}{1-q_1} \sum_{k=1}^{\infty} a_k\,.$$

The series above converges to a finite limit if the sequence $|a_k|^{1/k}$ converges to some limit $\alpha < 1$. Since

$$|a_k|^{1/k} = \frac{\beta}{\exp(B\,\eta_0^2\,\beta^{2\,k}/k)} \to \alpha = 0$$

as $k \to \infty$ it was shown that

$$\sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) \;<\; \infty\,.$$

Owing to criterion (7) one may conclude that the algorithm with self-adaptation gets stuck at the local optimum with positive probability precluding the property of global convergence.

Evidently, this failure is caused by a too fast decrease of the step size control parameter in the case of unsuccessful mutations. Therefore one may inquire under which step size control schedule a transition to $(z-\delta, z+\delta)$ is guaranteed. According to equation (4) we obtain

$$p_k \geq q_k = \delta\,\sqrt{\frac{2}{\pi}}\,\frac{1}{\sigma_k}\,\exp\left(-\,\frac{(z-\delta)^2}{2\,\sigma_k^2}\right) = A\,\eta_k\,\exp(-C\,\eta_k^2) \tag{10}$$

where $A = \delta\,(2/\pi)^{1/2}$, $C = (z-\delta)^2/2$, and $\eta_k = 1/\sigma_k$. Choose

$$\eta_k = \sqrt{\frac{1}{C}\,\log\left(\frac{A}{C^{1/2}}\,k\right)}$$

for $k > C^{1/2}/A$. Insertion into eqn. (10) leads to

$$q_k = \sqrt{\log\left(\frac{A}{C^{1/2}}\,k\right)}\cdot\frac{1}{k} \;\geq\; \frac{1}{k}$$

where the inequality is valid for sufficiently large k, i.e., for all $k \geq k_0$, say. Finally, Lemma 1 yields

$$\sum_{k=k_0}^{t} \log\left(\frac{1}{1-q_k}\right) \geq \sum_{k=k_0}^{t} q_k \geq \sum_{k=k_0}^{t} \frac{1}{k} \to \infty$$

as $t \to \infty$. This result is equivalent to the statement

$$1 - \prod_{k=k_0}^{t} (1-q_k) \to 1 \quad \text{ as } \; t \to \infty$$

which implies that a transition from 0 to the set $(z-\delta, z+\delta)$ happens in finite time with probability 1. This proves the existence of a schedule to escape from the local optimum under self-adaptation, but the mean step sizes have to be changed extremely slowly. Since such a schedule is also in action after entering the set $(z-\delta, z+\delta)$ the mean convergence velocity to the global optimum would be extremely slow as well (provided that convergence occurs at all). One is therefore tempted to inquire for another mutation distribution that might work with a faster self-adapting schedule than the inversely logarithmic one or even with the original one. Few years ago, some authors have suggested to replace the Gaussian mutation distribution by a Cauchy mutation (Kappler 1996; Yao and Liu 1996). Since the tails of the Cauchy distribution decrease much slower than the tails of the Gaussian distribution, it is hoped that such evolutionary algorithms exhibit an improved ability to escape from local optima. This hypothesis is examined next.

3.2 Cauchy Mutations

Suppose that the random variable Z possesses a Cauchy distribution with probability density function

$$f_Z(x) = \frac{1}{\pi} \, \frac{s}{s^2 + x^2} \tag{11}$$

where the scale parameter s plays the role of the step size control parameter ℓ. At first, consider the original schedule as given in equation (2). Unless one gets a successful mutation the scale parameter s_k decreases as $s_k = s_0 \, \gamma^k$ with $0 < \gamma < 1$. According to inequality (6) one obtains

$$p_k \le q_k = 2\,\delta\, f_{Z_k}(z+\delta) = \frac{2\,\delta}{\pi} \, \frac{s_k}{s_k^2 + (z+\delta)^2} = D \, \frac{s_k}{s_k^2 + E} \tag{12}$$

where $D = 2\,\delta/\pi$ and $E = (z+\delta)^2$. Inequalities (9) and (12) lead to

$$\sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) < \frac{1}{1-q_1} \sum_{k=1}^{\infty} q_k = \frac{D}{1-q_1} \sum_{k=1}^{\infty} \frac{s_k}{s_k^2 + E} = \frac{D\, s_0}{1-q_1} \sum_{k=1}^{\infty} \frac{\gamma^k}{s_0^2\, \gamma^{2\,k} + B} = \frac{D\, s_0}{1-q_1} \sum_{k=1}^{\infty} a_k \, .$$

The series above converges to a finite limit if the sequence $|a_k|^{1/k}$ converges to a limit $\alpha < 1$. Since

$$|a_k|^{1/k} = \frac{\gamma}{(s_0^2\, \gamma^{2\,k} + E)^{1/k}} \to \gamma < 1$$

as $k \to \infty$ it was shown that Cauchy mutations with the original step size control schedule are not a remedy: The hill between the local and global optimum will not be crossed with positive probability. Consequently, the property of global convergence is not present.

Again, one might seek for a slightly slower schedule than the original one that guarantees the crossing of the hill with probability 1. For this purpose choose

$$s_k = \frac{1}{k}$$

for $k \ge 1$ and insert this expression into the inequality (4). This leads to

$$p_k \ge q_k = 2\,\delta\, f_{Z_k}(z-\delta) = D \, \frac{s_k}{s_k^2 + G} = D \, \frac{1/k}{1/k^2 + G}$$

where $D = 2\,\delta/\pi$ and $G = (z-\delta)^2$. Owing to Lemma 1 one obtains

$$\sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) > \sum_{k=1}^{\infty} q_k = D \sum_{k=1}^{\infty} \frac{1/k}{1/k^2 + G} \ge \frac{D}{1+G} \sum_{k=1}^{\infty} \frac{1}{k} = \infty$$

which fulfills the criterion for successful crossing of the hill between local and global optimum. At a first glance, self-adaptation with Cauchy mutations seems to work with a faster schedule than in the case of Gaussian mutations. But this impression is misleading since the choice of a sufficiently slowly schedule depends on the representation of the step size control parameter in the definition of the probability density function $f_Z(\cdot)$. For example, if parameter s in equation (11) is replaced by $h(s)$ with $h(s) = \log(\gamma)/\log(s/s_0)$, then the original schedule would lead to a crossing of the hill with probability 1. Thus, some caution is appropriate here.

4 Extensions of the Scenario

4.1 Original (1+1)-Strategy with Self-Adaptation

As mentioned in Section 2.1 the stochastic algorithm considered so far does not match the original $(1+1)$-strategy with self-adaptation exactly. It will be shown that the results obtained by now remain valid for the original $(1+1)$-EA which changes the step size control parameter if the relative frequency of improving mutations is below or above some threshold within m, say, trials. Unless there is an outcome of Z_k in the set $(z-\delta, z+\delta)$ the step size control parameter is decreased by factor $\gamma \in (0,1)$ after every mth trial. This behavior can be squeezed into our original scenario by considering these m trials as an elementary event of stage k. The probability of observing a transition to $(z-\delta, z+\delta)$ within m trials at stage k is $\hat{p}_k = 1-(1-p_k)^m$ where p_k is the probability of successful crossing for a single trial (as known from the previous section). Notice that $p_k \le q_k$ if and only if $\hat{p}_k = 1-(1-p_k)^m \le \hat{q}_k = 1-(1-q_k)^m$ and analogous for the reversed inequality. Therefore the sufficient criterion for unsuccessful crossing of the hill (eqn. 7) remains valid if q_k is replaced by $\hat{q}_k$. Since m is finite and

$$\prod_{k=1}^{\infty}(1-\hat{q}_k) = \prod_{k=1}^{\infty}(1-q_k)^m = \left[\prod_{k=1}^{\infty}(1-q_k)\right]^m < 1 \quad\Longleftrightarrow\quad m \cdot \sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) < \infty$$

one may conclude from the preceding analysis with $m=1$ that the $(1+1)$-EA with usual self-adaptation and Gaussian or Cauchy mutations does not globally converge for the class of continuous functions in general.

4.2 Multiple Offspring

Now let the parent produce $\lambda \ge 2$ offspring with the same mutation distribution. The offspring with the least objective function value replaces the parent if and only if its objective function value is less than that of the parent. This EA is known as the $(1+\lambda)$-EA. Thus, at every stage there are now $\lambda \cdot m$ in lieu of m trials. Since λ is finite the argumentation of Section 4.1 is directly transferable to this scenario: The $(1+\lambda)$-EA with self-adaptation and Gaussian or Cauchy mutations does not globally converge for the class of continuous functions in general.

4.3 Different Starting Point

It might be argued that the entire analysis is irrelevant since the event of starting the EA at $X_0 = 0$ has zero measure under random initialization. But this hope is in vain. Suppose that the EA is started in the set $[-\delta, \delta]$ which has non-zero measure under random initialization.

With $X_0 \in [-\delta, \delta]$ the EA only accepts offspring with objective function value less than $0.75\,\epsilon\, X_0^2/\delta^2 + \epsilon$. Suppose for the moment that the EA does not accept improvements within the left valley. In the best case the starting point is at $X_0 = \delta$ because the set of improving solutions on the other side of the hill as well as the assigned probability mass becomes maximal. Now one is back at the preceding scenarios—if the EA is unable to enter the set $\mathcal{H}_\delta = (z - \delta\sqrt{7/4}, z + \delta\sqrt{7/4})$ with probability 1 then global convergence is precluded. The original analysis remains the same for this situation—only the values of the constants $A, B, \ldots$ in the original analysis are affected. Since they do not influence the argumentation of the analysis the main result is also valid under this scenario.

Finally, let the EA also accept better solutions in the left valley. To preclude the property of global convergence it is sufficient to show that the EA is unable to enter the set $\mathcal{H}_\delta$ with probability one. This leads to the following model: The EA minimizes the convex function with domain $[-\delta, \delta]$ exponentially fast (as proven in Rappl 1984) or the EA enters the set $\mathcal{H}_\delta$ and halts. At each stage the probability of a transition to $\mathcal{H}_\delta$ is less than $1 - (1 - \tilde{q}_k)^m$ where $\tilde{q}_k$ plays the role of q_k with altered constants. While the EA minimizes the convex function the step size control parameter may temporarily increase—in the limit, however, the step size control parameter decreases exponentially fast. Therefore, the probability $\tilde{q}_k$ may increase temporarily as well—but in the limit it decreases too fast to make the event of entering the set $\mathcal{H}_\delta$ a sure one. Summing up:

Theorem 1 *A randomly initialized* $(1 + \lambda)$*-EA with a self-adaptation method resembling the* $1/5$*-success rule does not converge with probability* 1 *to the global optimum of a continuous objective function in general.* □

5 Discussion

The negative result concerning self-adaptation just presented should be accompanied by some comments. From a theoretical point of view, this result is remarkable because many EAs without self-adaptation do globally converge to the optimum. But this does not automatically imply that self-adaptation is an obstructive feature of evolutionary algorithms. Proofs of global convergence for EAs without self-adaptation ensure that the global optimum will be found in finite time with probability 1 regardless of the starting point. Although finite, the time actually necessary to enter some vicinity of the global optimum (in which the objective function is locally convex) may be transcomputational long. Moreover, even if this event has been occurred, these EAs require much time to locate the optimum with more accuracy. Evolutionary algorithms with self-adaptation do not always enter the vicinity of the global optimum in finite time, but as soon as they have reached this set the global optimum is approached exponentially fast. If the union of vicinities with locally convex objective functions and sufficiently good local optima is large enough, then repeated runs of the EA will ensure that at least one run enters this union of vicinities with high probability (the failure probability decreases exponentially in the number of runs). After this has happened a sufficiently good local optimum is approached exponentially fast. From a practical point of view, such a behavior is certainly satisfying. Therefore, future research should be engaged in replacing the vague phrase "... *sufficiently good* local optima is *large enough* ..." by rigorous statements.

Acknowledgments

— will be placed here —

References

T. Bäck (1998). An overview of parameter control methods by self-adaptation in evolutionary algorithms. *Fundamenta Informaticae 45*(1-4), 51–66.

H.-G. Beyer (1995). Toward a theory of evolution strategies: Self–adaptation. *Evolutionary Computation 3*(3), 311–347.

D. B. Fogel (1992). *Evolving Artificial Intelligence*. Ph. D. thesis, University of California, San Diego.

D. B. Fogel (1995). *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. New York: IEEE Press.

R. Hinterding, Z. Michalewicz, and A. E. Eiben (1997). Adaptation in evolutionary computation: A survey. In *Proceedings of the Fourth International Conference on Evolutionary Computation (ICEC 97)*, pp. 65–69. Piscataway (NJ): IEEE Press.

C. Kappler (1996). Are evolutionary algorithms improved by large mutations? In H.-M. Voigt, W. Ebeling, I. Rechenberg, and H.-P. Schwefel (Eds.), *Parallel Problem Solving From Nature—PPSN IV*, pp. 346–355. Berlin: Springer.

G. Rappl (1984). *Konvergenzraten von Random Search Verfahren zur globalen Optimierung*. Doctoral Dissertation, Hochschule der Bundeswehr München, Germany.

I. Rechenberg (1973). *Evolutionsstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*. Stuttgart: Frommann–Holzboog Verlag.

G. Rudolph (1997). *Convergence Properties of Evolutionary Algorithms*. Hamburg: Kovač.

H.-P. Schwefel (1977). *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Basel: Birkhäuser.

X. Yao and Y. Liu (1996). Fast evolutionary programming. In L. J. Fogel, P. J. Angeline, and T. Bäck (Eds.), *Proceedings of the Fifth Annual Conference on Evolutionary Programming*, pp. 451–460. Cambridge (MA): MIT Press.