Self-Adaptive Mutations May Lead to Premature Convergence

Günter Rudolph

Universität Dortmund

Fachbereich Informatik, LS XI

D–44221 Dortmund, Germany

Rudolph@LS11.CS.Uni-Dortmund.de

July 28, 1999

Abstract

Self-adaptive mutations are known to endow evolutionary algorithms (EAs) with the ability of locating local optima quickly and accurately, whereas it was unknown whether these local optima are finally global ones provided that the EA runs long enough. In order to answer this question it is assumed that the $(1+1)$-EA with self-adaptation is located in the vicinity $\mathcal{P}$ of a local solution with objective function value ε. In order to exhibit convergence to the global optimum with probability 1 the EA must generate an offspring that is an element of the lower level set $\mathcal{S}$ containing all solutions (including a global one) with objective function value less than ε. In case of multimodal objective functions these sets $\mathcal{P}$ and $\mathcal{S}$ are generally not adjacent, i.e., $\min\{\|x-y\| : x \in \mathcal{P}, y \in \mathcal{S}\} > 0$, so that the EA has to surmount the barrier of solutions with objective function values larger than ε by a lucky mutation. It will be proven that the probability of this event is less than 1 even under an infinite time horizon. This result implies that the EA can get stuck at a non-global optimum with positive probability. Some ideas of how to avoid this problem are discussed as well.

Keywords: evolutionary algorithms, self-adaptation, premature convergence

1 Introduction

The self-adaptation of the mutation distribution in evolution strategies (ES) was introduced by Rechenberg [1]. 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 [2] 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 [3, 4]. Needless to say, self-adaptation is not limited to the control of mutation distributions. Further fields of application may be found in recent surveys [5, 6].

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 [7] has given a proof of exponentially fast convergence for a stochastic algorithm resembling a $(1+1)$-ES with $1/5$-success rule whereas Beyer [8] examined also other evolutionary algorithms and self-adaptation rules.

As for non-convex objective functions, Rudolph [9] has shown that every self-adaptation method leads to global convergence to the global optimum for objective functions with bounded lower level sets of non-zero measure provided that the selection method uses elitism and that the self-adaptation rule does not violate the property of the mutation distribution of ensuring a positive minimum probability for hitting arbitrary subsets of the search set (cf. Theorem 6.14, p. 204). If the latter condition is not valid it has been shown[1] in case of a one-dimensional continuous test problem and Rechenberg's

[1] I wish to thank Lin Dan, P. R. of China, who pointed out that the constants in equation (3) of my paper [10] were erroneously interchanged and then used throughout the entire paper. But apart from these wrong constants the results of the paper remain valid.

$1/5$-success rule that global convergence to the optimum will not happen with probability 1.

Here, this result is generalized to multivariate problems. Section 2 offers a detailed description of the scenario including the evolutionary algorithm under consideration, its presupposed current situation in its search process, and general criteria for deciding whether the EA may get stuck at a local optimum or not. In Section 3 these general criteria are then instantiated with the presupposed scenario and two specific mutation distributions. The analysis indicates that elitist evolutionary algorithms with a self-adaptation mechanism resembling the $1/5$-success rule may get caught by a local optimum with positive probability even under infinite time horizon. A subsequent reconsideration of this result and its proof finally reveals that this property remains valid for all mutation distributions with independent marginal distributions. A discussion of some ideas of how to circumvent this problem completes this section. The conclusions are drawn in Section 4.

2 Description of the Scenario

2.1 Algorithm

Let $g : \mathbb{R}^n \to \mathbb{R}$ be the objective function to be minimized and set $\mathbf{X}_0 = \mathrm{x}_0 \in \mathbb{R}^n$, $\ell_0 > 0$. Consider the Markovian process $(\mathbf{X}_k, \ell_k)_{k \geq 0}$ generated by the stochastic algorithm

$$\mathbf{X}_{k+1} = \begin{cases} \mathbf{X}_k + \ell_k\, \mathbf{Z}_k & \text{, if } g(\mathbf{X}_k + \ell_k\, \mathbf{Z}_k) < g(\mathbf{X}_k) \\ \mathbf{X}_k & \text{, otherwise} \end{cases} \tag{1}$$

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

where $\gamma_1 > 1$ and $\gamma_2 \in (0, 1)$. Each random vector $\mathbf{Z}_k$ of the sequence of independent and identically distributed random vectors $(\mathbf{Z}_k : k \geq 0)$ has a joint probability density function with independent marginal densities

$$f_{\mathbf{Z}}(z_1, \ldots, z_n) = \prod_{i=1}^{n} f_{Z_i}(z_i)$$

where $f_{Z_i}(\cdot)$ is unimodal with mode at zero and $f_{Z_i}(z_i) > 0$ for all $z_i \in \mathbb{R}$. 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 transfered easily to a broader class of evolutionary algorithms as shown in Section 2.5.

2.2 Abstract Test Problem

Let there be two disjoint compact sets $\mathcal{P} \subset \mathbb{R}^n$ and $\mathcal{S} \subset \mathbb{R}^n$. For each $\mathbf{x} \in \mathcal{P}$ the objective function value is $g(\mathbf{x}) = \varepsilon > 0$ whereas $g(\mathbf{x}) < \varepsilon$ for all $\mathbf{x} \in \mathcal{S}$. If $\mathbf{x} \notin \mathcal{P} \cup \mathcal{S}$ then $g(\mathbf{x}) > \varepsilon$. Without loss of generality let $\mathcal{P}$ be a hypercube, whose vertices have nonpositive vector components (inclusive the zero vector), whereas $\mathcal{S}$ is a hyberball whose center $\mathbf{c} \in \mathbb{R}^n$ has identical positive vector components such that the radius of the hyperball is less than $\|\mathbf{c}\|/2$. Figure 1 offers a sketch of the test problem if the dimension is $n = 2$.

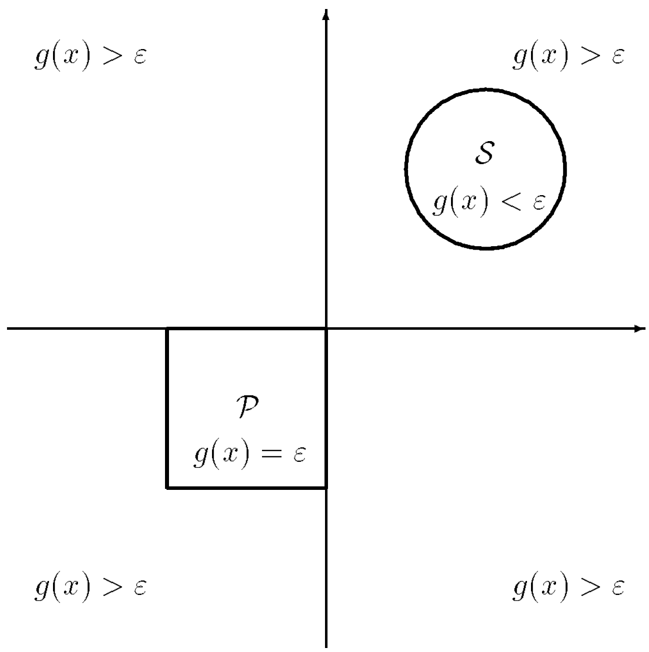

Figure 1: Sketch of the test problem $g(\cdot)$ if the dimension is $n = 2$.

Suppose that the initial individual $\mathbf{X}_0$ is located in the hypercube $\mathcal{P} \subset \mathbb{R}^n$. The EA specified in equations (1) and (2) will accept a mutation if and only if the new point hits the set $\mathcal{S}$. As long as this event has not happened the step length control parameter ℓ_k is steadily decreased. If this decrease is driven too fast then the event of a transition to the set $\mathcal{S}$ will not occur with probability one. As a consequence, the convergence with probability one to the global optimum would be precluded in this case.

In order to provide a tool for deciding whether a “decrease is driven too fast” or not two simple criteria are developed next. Without loss of generality it will be assumed that $\mathbf{X}_0 = \mathbf{0} \in \mathcal{P}$; other starting points inside $\mathcal{P}$ will only change some constants without affecting the results qualitatively.

2.3 Criterion for Unsecured Escape from Local Optima

Suppose the existence of an easily determinable *upper* bound q_k for the probability p_k of a transition from the zero vector ($\in \mathcal{P}$) to the set $\mathcal{S}$, i.e.,

$$p_k = \mathsf{P}\{\,\mathbf{0} \to \mathcal{S} \text{ at step } k\,\} \le q_k\,.$$

As a consequence, the upper bound for the probability of a transition to the set $\mathcal{S}$ within $t \ge 1$ trials is given by

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

If the probability on the right hand side above is smaller than 1 in the limit, then the transition to $\mathcal{S}$ is not guaranteed. In other words, it may happen with positive probability that the EA never enters the set $\mathcal{S}$. It is clear that such an event precludes the convergence to the global optimum located in $\mathcal{S}$. Thus, the sufficient criterion of a potential failure is simply

$$\prod_{k=1}^{\infty}(1-q_k) > 0 \quad \Leftrightarrow \quad \sum_{k=1}^{\infty}\log\left(\frac{1}{1-q_k}\right) < \infty\,. \tag{3}$$

2.4 Criterion for Secured Escape from Local Optima

Now assume there exists an easily determinable *lower* bound q_k for the probability p_k of a transition from the zero vector ($\in \mathcal{P}$) to the set $\mathcal{S}$, i.e.,

$$p_k = \mathsf{P}\{\,\mathbf{0} \to \mathcal{S} \text{ at step } k\,\} \ge q_k\,.$$

It follows that the probability of a transition to set $\mathcal{S}$ within $t \ge 1$ trials is

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

Therefore the sufficient criterion for a secured escape from the local optimum is simply

$$\prod_{k=1}^{t}(1-q_k) \to 0 \quad \Leftrightarrow \quad \sum_{k=1}^{t}\log\left(\frac{1}{1-q_k}\right) \to \infty \tag{4}$$

as $t \to \infty$. If criterion (4) is fulfilled, then the evolutionary algorithm will jump to the set $\mathcal{S}$ within a finite number of steps with probability one. This does not imply that the EA will converge to the global optimum located in $\mathcal{S}$, but the necessary condition for this property would be fulfilled.

2.5 Extensions of the Scenario

2.5.1 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. Now it will be shown that the results to be obtained for the EA described in Section 2.1 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 a mutation hits the set $\mathcal{S}$ the step size control parameter is decreased by factor $\gamma \in (0,1)$ after every mth trial. This behavior can be squeezed into the original scenario by considering these m trials as an elementary event of stage k. The probability of observing a transition to the set $\mathcal{S}$ within m trials at stage k is $\hat{p}_k = 1-(1-p_k)^m$ where p_k is the probability of a transition to $\mathcal{S}$ for a single trial. 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 potential premature convergence (eqn. 3) remains valid if q_k is replaced by $\hat{q}_k$. Since m is finite and

$$\prod_{k=1}^{\infty}(1-\hat{q}_k) = \prod_{k=1}^{\infty}(1-q_k)^m = \left[\prod_{k=1}^{\infty}(1-q_k)\right]^m \;>\; 0$$

$$\Longleftrightarrow \quad m \cdot \sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) \;<\; \infty$$

one may conclude that it is sufficient to examine the properties of the EA described in Section 2.1 in order to obtain results for the $(1+1)$-EA with usual self-adaptation mechanism.

2.5.2 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 2.5.1 is directly transferable to this scenario: Again it is sufficient to examine the properties of the EA described in Section 2.1 in order to obtain results for the $(1+\lambda)$-EA with usual self-adaptation mechanism.

3 Analysis

3.1 Determination of Simple Probability Bounds

Recall from Section 2.1 that $(\mathbf{Z}_k)_{k\geq 0}$ is a sequence of independent and identically distributed random vectors, each of them with independent, identical, and unimodal marginal densities $f_{Z_i^{(k)}}(\cdot)$ where the mode is zero and $f_{Z_1^{(k)}}(z_1) > 0$ for all $z_1 \in \mathbb{R}$. Let $\tilde{\mathbf{Z}}_k = \ell_k\,\mathbf{Z}_k$ be the scaled mutation vector at step $k \geq 0$. Since $\mathbf{X}_0 = \mathbf{0} \in \mathcal{P}$ one obtains

$$p_k = \mathsf{P}\{\,\mathbf{0} \to \mathcal{S} \text{ at step } k \,|\, \mathbf{X}_k = \mathbf{0}\,\} = \mathsf{P}\{\,\tilde{\mathbf{Z}}_k \in \mathcal{S} \,|\, \mathbf{X}_k = \mathbf{0}\,\}\,.$$

Let $\mathcal{A} = [a_1, a_2]^n$ and $\mathcal{B} = [b_1, b_2]^n$ with $b_1 > 0$ be two hypercubes such that $\mathcal{A} \subset \mathcal{S} \subset \mathcal{B}$. This leads to the upper bound

$$\begin{aligned}
\mathsf{P}\{\,\tilde{\mathbf{Z}}_k \in \mathcal{S} \,|\, \mathbf{X}_k = \mathbf{0}\,\} &= \int_{\mathcal{S}} f_{\tilde{\mathbf{Z}}_k}(\mathbf{z})\,d\mathbf{z} \\
&< \int_{\mathcal{B}} f_{\tilde{\mathbf{Z}}_k}(\mathbf{z})\,d\mathbf{z} \\
&\quad \text{(since } \mathcal{S} \subset \mathcal{B}) \\
&= \int_{b_1}^{b_2} \cdots \int_{b_1}^{b_2} \left(\prod_{i=1}^{n} f_{\tilde{Z}_i^{(k)}}(z_i)\,dz_i \right) \\
&\quad \text{(by independence)} \\
&= \prod_{i=1}^{n} \left(\int_{b_1}^{b_2} f_{\tilde{Z}_i^{(k)}}(z_i)\,dz_i \right) \\
&\quad \text{(by Fubini's Theorem)} \\
&= \left(\int_{b_1}^{b_2} f_{\tilde{Z}_1^{(k)}}(z_1)\,dz_1 \right)^n \\
&\quad \text{(by identical marginal distributions)} \\
&= \left(\int_{b_1}^{b_2} \frac{1}{\ell_k}\, f_{Z_1^{(k)}}\left(\frac{z_1}{\ell_k}\right)\,dz_1 \right)^n \\
&\quad \text{(by density transformation)} \\
&= \left[F_{Z_1^{(k)}}\left(\frac{b_2}{\ell_k}\right) - F_{Z_1^{(k)}}\left(\frac{b_1}{\ell_k}\right) \right]^n \\
&= \left[f_{Z_1^{(k)}}\left(\frac{b_1 + \delta\,(b_2 - b_1)}{\ell_k}\right) \cdot \frac{b_2 - b_1}{\ell_k} \right]^n \qquad (0 \leq \delta \leq 1) \\
&\quad \text{(by mean value theorem)} \\
&\leq \left(\frac{b_2 - b_1}{\ell_k}\right)^n \cdot \left[f_{Z_1^{(k)}}\left(\frac{b_1}{\ell_k}\right) \right]^n \qquad (5) \\
&\quad \text{(by unimodality of densities with mode zero)}
\end{aligned}$$

where $F_Z(\cdot)$ denotes the cumulative distribution function of random variable Z. In analogous manner one obtains the lower bound via

$$\begin{aligned} \mathsf{P}\{\, \tilde{\mathbf{Z}}^{(k)} \in \mathcal{S} \,|\, \mathbf{X}_k = \mathbf{0} \,\} &= \int_{\mathcal{S}} f_{\tilde{\mathbf{Z}}^{(k)}}(\mathbf{z})\, d\mathbf{z} \\ &> \int_{\mathcal{A}} f_{\tilde{\mathbf{Z}}^{(k)}}(\mathbf{z})\, d\mathbf{z} \\ &= \left(\frac{a_2 - a_1}{\ell_k}\right)^n \left[f_{Z_1^{(k)}}\left(\frac{a_1 + \delta\,(a_2 - a_1)}{\ell_k}\right)\right]^n \\ &\ge \left(\frac{a_2 - a_1}{\ell_k}\right)^n \left[f_{Z_1^{(k)}}\left(\frac{a_2}{\ell_k}\right)\right]^n . \end{aligned} \tag{6}$$

For the sake of notational convenience the sub- and superscript appearing in the marginal density $f_{Z_1^{(k)}}(\cdot)$ will be omitted hereinafter. This does not cause any problem since the sequence of random vectors $(\mathbf{Z}_k)_{k\ge 0}$ are independent and identically distributed and since the marginal densities are identical for each vector.

3.2 Proof of Potential Premature Convergence

In order to check for potential premature convergence one has to insert the upper bound in eqn. (5) into the criterion given in eqn. (3). The calculations required for this purpose can be facilitated by exploiting the following inequalities which follow immediately from the series expansion of the logarithm (also see [10]). Since $q_k \in (0,1)$ and $q_{k+1} \le q_k$ for all $k \ge 1$ one obtains

$$\sum_{k=1}^{\infty} q_k < \sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) < \sum_{k=1}^{\infty} \frac{q_k}{1-q_k} \le \frac{1}{1-q_1} \sum_{k=1}^{\infty} q_k\,. \tag{7}$$

Consequently, instead of inserting eqn. (5) into eqn. (3) directly it suffices to insert eqn. (5) into the rightmost expression of eqn. (7). Moreover, since the constant factor $1/(1-q_1)$ does not affect the convergence behavior of the series the sufficient criterion for potential premature convergence reduces to

$$\sum_{k=1}^{\infty} q_k = (b_2 - b_1)^n \sum_{k=1}^{\infty} \ell_k^{-n} \left[f_Z\left(\frac{b_1}{\ell_k}\right)\right]^n < \infty$$

and eventually, after ignoring the constant factor $(b_2 - b_1)^n$, to

$$\sum_{k=1}^{\infty} \ell_k^{-n} \left[f_Z\left(\frac{b_1}{\ell_k}\right)\right]^n = \sum_{k=1}^{\infty} r_k < \infty$$

where $r_k = (f_Z(b_1/\ell_k)/\ell_k)^n \ge 0$. Notice that the series above converges if, for example, $r_k^{1/k} \to \alpha < 1$ as $k \to \infty$. In the scenario considered here one has $\ell_k = \ell_0\,\gamma^k$ with $\gamma \in (0,1)$ such that

$$r_k^{1/k} = \left(\frac{f_Z(b_1\,\ell_0^{-1}\,\gamma^{-k})}{\ell_0\,\gamma^k}\right)^{n/k} \to \gamma^{-n} \lim_{k\to\infty} \left[f_Z(B/\gamma^k)\right]^{n/k}$$

where $B = b_1/\ell_0 > 0$. As a consequence, the final criterion for potential premature convergence reads

$$\lim_{k\to\infty} \left[f_Z(B/\gamma^k) \right]^{1/k} < \gamma \,. \tag{8}$$

Popular choices for mutation distributions are Gaussian and Cauchy distributions which will serve as explicit examples for using the criterion given in eqn. (8). The marginal density of the standard multivariate Gaussian distribution is $f_Z(z) = (2\,\pi)^{-1/2}\,\exp(-z^2/2)$. Insertion into eqn. (8) leads to

$$\left[f_Z(B/\gamma^k) \right]^{1/k} = \frac{1}{(2\,\pi)^{1/(2k)}}\,\exp\left(-\frac{B^2}{2\,k\,\gamma^{2\,k}}\right) \to 0 < \gamma$$

as $k \to \infty$. As for the Cauchy distribution $f_Z(z) = 1/[\pi\,(1+z^2)]$, insertion into eqn. (8) yields

$$\left[f_Z(B/\gamma^k) \right]^{1/k} = \frac{1}{\pi^{1/k}}\,\frac{\gamma^2}{(\gamma^{2\,k} + B^2)^{1/k}} \to \gamma^2 < \gamma$$

as $k \to \infty$. Thus, neither Gaussian nor Cauchy mutations can ensure the escape from local optima with probability one. This observation raises the question which distribution might lead to secured escape from local optima under the self-adaptation mechanism considered here. A closer look at equation (8) reveals that there is no marginal density not fulfilling criterion (8). Suppose that the tail of the marginal density decreases like $f_Z(z) \sim 1/z^{1+\delta}$ as $z \to \infty$ for some $\delta > 0$. Owing to criterion (8) one obtains

$$\left[f_Z(B/\gamma^k) \right]^{1/k} \sim \frac{\gamma^{1+\delta}}{B^{(1+\delta)/k}} \to \gamma^{1+\delta} < \gamma$$

as $k \to \infty$. Notice that a function with $\delta \le 0$ is not a marginal density function since its integral over $\mathbb{R}$ diverges.

3.3 Schedules with Secured Escape from Local Optima

The preceding section has shown that the too quickly decreasing step sizes produced by the step size rule given in eqn. (2) are the reason for potential premature convergence. Therefore the focus of interest is now shifted towards appropriate modifications of the step size rule such that the decrease in the step sizes is sufficiently decelerated for ensuring the escape from local optima.

According to the criterion given in eqn. (4) together with inequality (7) it suffices to show that

$$\sum_{k=1}^{\infty} \log\left(\frac{1}{1-q_k}\right) > \sum_{k=1}^{\infty} q_k = \sum_{k=1}^{\infty} \left(\frac{a_2 - a_1}{\ell_k}\right)^n \left[f_Z\left(\frac{a_2}{\ell_k}\right) \right]^n = \infty$$

where q_k is the lower bound given in eqn. (6). Assume that the tails of the marginal densities behave like $f_Z(z) \sim 1/z^{1+\delta}$ as $z \to \infty$ for some $\delta > 0$ (this includes the Cauchy distribution with $\delta = 1$).

Then it is easily seen (after ignoring constant factors) that the series above diverges if, for example,

$$\ell_k^{-n}\,\ell_k^{(1+\delta)\,n} \geq \frac{1}{k}$$

for all $k \geq 1$. Thus, a step size rule producing the schedule $\ell_k \sim 1/k^{1/(\delta\,n)}$ in case of continued unsuccessful mutations would finally lead to a guaranteed escape from a local optimum (after a *finite* number of trials). An imaginable realization of such a step rule might be as follows:

$$\ell_{k+1} = \begin{cases} \ell_k \left(1 - \dfrac{1}{k+1}\right)^{1/(\delta\,n)} & \text{, if } g(\mathbf{X}_k + \ell_k\,\mathbf{Z}_k) < g(\mathbf{X}_k) \\[2ex] \ell_k \left(1 + \dfrac{1}{k}\right)^{1/(\delta\,n)} & \text{, otherwise} \end{cases}$$

where $k \geq 1$. A step size rule of this kind, however, will lead to very slow local convergence velocity—and the situation gets dramatically worse in case of similar step size rules for mutation distributions with exponentially decreasing tails (e.g. Gaussian mutations). If an improvement is not found quickly then k gets large and the adapting factors are practically equal to 1. As a consequence, the step sizes are hardly altered then. It might therefore be a more favorable alternative to apply the usual step size rule (guarantees fast local convergence speed) accompanied by "occasional" mutations with a fixed distribution (guarantees global convergence to global optimum). For example, in addition to the original λ Gaussian mutations with adapted step sizes one could simply generate an offspring from a fixed Cauchy distribution. This would yield a $(1 + (\lambda + 1))$-ES with fast local convergence speed *and* the property of global convergence to the optimum at the expense of only little additional computing effort.

4 Conclusions

It was proven that elitist evolutionary algorithms with a self-adaptation mechanism resembling Rechenberg's $1/5$-success rule will get caught by non-global optima with positive probability even under an infinite time horizon. The proof is specialized to mutation distributions with independent marginal distributions. The conjecture that the result remains valid for mutations with dependent marginal distributions is certainly not too risky—but the proof will be technically different to the one presented here. Schwefel's version of self-adaptation with lognormally distributed adaptation factors will pose additional technical difficulties. Moreover, there is no obvious reason why self-adaptive evolutionary algorithms *without* elitist selection should behave differently in this context. As can be seen from this list of open problems, the theoretical foundations of self-adaptation are, at best, in statu nascendi. This list may be continued by inquiring after self-adaptation rules that won't get stuck at non-global optima. A simple example of such a rule has been given here. This rule does not only offer convergence to the global optimum but also fast convergence towards local optima. Although these properties are theoretically appealing it is currently unknown whether this modified self-adaptation rule will be beneficial in practical applications or not. Analytical as well as empirical studies of this kind, however, would go beyond the scope of this paper and must therefore remain open for future work.

Acknowledgment

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

References

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

[2] H.-P. Schwefel. *Numerische Optimierung von Computer-Modellen mittels der Evolutionsstrategie*. Birkhäuser, Basel, 1977.

[3] D. B. Fogel. *Evolving Artificial Intelligence*. PhD thesis, University of California, San Diego, 1992.

[4] D. B. Fogel. *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*. IEEE Press, New York, 1995.

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

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

[7] G. Rappl. *Konvergenzraten von Random Search Verfahren zur globalen Optimierung*. Doctoral Dissertation, University of the Bundeswehr, Munich, Germany, 1984.

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

[9] G. Rudolph. *Convergence Properties of Evolutionary Algorithms*. Kovač, Hamburg, 1997.

[10] G. Rudolph. Global convergence and self-adaptation: A counter-example. In *Proceedings of the 1999 Congress of Evolutionary Computation (CEC '99), Vol. 1*, pages 646–651. IEEE Press, Piscataway (NJ), 1999.